



Print ISSN: 2152-4157
Online ISSN: 2152-4165

FALL/WINTER 2016
VOLUME 8, NUMBER 2

WWW.IJERI.ORG

International Journal of Engineering Research & Innovation

Editor-in-Chief: Mark Rajai, Ph.D.
California State University Northridge



Published by the
International Association of Journals & Conferences



www.ijeri.org

Print ISSN: 2152-4157
Online ISSN: 2152-4165



www.iajc.org

INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH AND INNOVATION

ABOUT IJERI:

- IJERI is the second official journal of the International Association of Journals and Conferences (IAJC).
- IJERI is a high-quality, independent journal steered by a distinguished board of directors and supported by an international review board representing many well-known universities, colleges, and corporations in the U.S. and abroad.
- IJERI has an impact factor of **1.58**, placing it among an elite group of most-cited engineering journals worldwide.

OTHER IAJC JOURNALS:

- The International Journal of Modern Engineering (IJME)
For more information visit www.ijme.us
- The Technology Interface International Journal (TIIJ)
For more information visit www.tiij.org

IJERI SUBMISSIONS:

- Manuscripts should be sent electronically to the manuscript editor, Dr. Philip Weinsier, at philipw@bgsu.edu.

For submission guidelines visit
www.ijeri.org/submissions

TO JOIN THE REVIEW BOARD:

- Contact the chair of the International Review Board, Dr. Philip Weinsier, at philipw@bgsu.edu.

For more information visit
www.ijeri.org/editorial

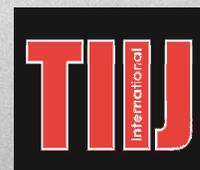
INDEXING ORGANIZATIONS:

- IJERI is indexed by numerous agencies. For a complete listing, please visit us at www.ijeri.org.

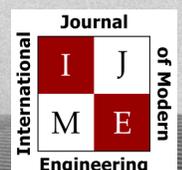
Contact us:

Mark Rajai, Ph.D.

Editor-in-Chief
California State University-Northridge
College of Engineering and Computer Science
Room: JD 4510
Northridge, CA 91330
Office: (818) 677-5003
Email: mrajai@csun.edu



www.tiij.org



www.ijme.us

INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH AND INNOVATION

The INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH AND INNOVATION (IJERI) is an independent and not-for-profit publication, which aims to provide the engineering community with a resource and forum for scholarly expression and reflection.

IJERI is published twice annually (fall and spring issues) and includes peer-reviewed research articles, editorials, and commentary that contribute to our understanding of the issues, problems, and research associated with engineering and related fields. The journal encourages the submission of manuscripts from private, public, and academic sectors. The views expressed are those of the authors and do not necessarily reflect the opinions of the IJERI editors.

EDITORIAL OFFICE:

Mark Rajai, Ph.D.
Editor-in-Chief
Office: (818) 677-2167
Email: ijmeeditor@iajc.org
Dept. of Manufacturing Systems
Engineering & Management
California State University-
Northridge
18111 Nordhoff Street
Northridge, CA 91330-8332

THE INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH AND INNOVATION EDITORS

Editor-in-Chief:

Mark Rajai

California State University-Northridge

Associate Editors:

Paul Wilder

Vincennes University

Li Tan

Purdue University Northwest

Production Editor:

Philip Weinsier

Bowling Green State University-Firelands

Subscription Editor:

Morteza Sadat-Hossieny

Northern Kentucky University

Web Administrator:

Saeed Namyar

Advanced Information Systems

Manuscript Editor:

Philip Weinsier

Bowling Green State University-Firelands

Copy Editors:

Li Tan

Purdue University Northwest

Ahmad Sarfaraz

California State University-Northridge

Technical Editors:

Marilyn Dyrud

Oregon Institute of Technology

Publisher:

Bowling Green State University Firelands

TABLE OF CONTENTS

<i>Editor's Note (In This Issue): LIDAR: <u>L</u>ight <u>D</u>etection <u>A</u>nd <u>R</u>anging.....</i>	<i>3</i>
Philip Weinsier, IJERI Manuscript Editor	
<i>Evaluation of Semi-Automated Target Extraction Methods for Digital Close-Range Photogrammetry.....</i>	<i>5</i>
Sagar Deshpande, Ferris State University	
<i>Performance Validation of an Energy Efficient Elevator Controller.....</i>	<i>14</i>
Muhammad Z Hasan, Texas A&M International University; Rainer Fink, Texas A&M University; Daryl James, Texas A&M University; Jesus Gamboa, Texas A&M International University	
<i>Matlab/Simulink Simulation of Arc Faults.....</i>	<i>23</i>
Amir Khan, California State University Fresno; Nagy Bengiamin, California State University, Fresno	
<i>Biomimetic Transfemoral Knee with a Gear Mesh Locking Mechanism.....</i>	<i>30</i>
Tyagi Ramakrishnan, University of South Florida; Millicent Schlafly, University of South Florida; Kyle B. Reed, University of South Florida	
<i>Feasibility of Using an Augmenting Solar Array for a Greenhouse in Northwest Ohio.....</i>	<i>39</i>
Sudershan Jetley, Bowling Green State University; Binisha Shrestha, Bowling Green State University	
<i>Using LIDAR and Animation to Develop Surveillance Strategies.....</i>	<i>46</i>
Roger Ruggles, Lafayette College; Karen Ruggles, DeSales University	
<i>Evaluating the Fidelity of a Scale Model Hexapod Motion Simulator.....</i>	<i>54</i>
Christopher Lawrence, Schmidt Peterson Motorsports; Andrew Borme, Indiana University Purdue University Indianapolis; Pete Hylton, Indiana University Purdue University Indianapolis	
<i>Pushover Analysis of the New Self-Anchored Suspension Bay Bridge Tower.....</i>	<i>62</i>
Abolhassan Astaneh-Asl, The University of California, Berkeley; Xin Qian, The University of California, Berkeley	
<i>Performance Analysis of Distributed Systems Involving Loops.....</i>	<i>74</i>
Toqeer Israr, Eastern Illinois University	
<i>Designing Standalone Microgrid and Grid-Connected Smartgrid Hybrid Solar/Wind Energy Systems.....</i>	<i>83</i>
Dominick Lauria, University of Hartford; Sarah Lamb, University of Hartford; Akram Abu-Aisheh, University of Hartford	
<i>Investigating the Effects of Stereo Camera Baseline on the Accuracy of 3D Projection for Industrial Robotic Applications.....</i>	<i>94</i>
Wutthigrai Boonsuk, Eastern Illinois University	
<i>Towards the Improvement of Highway-Rail Intersections with an Intelligent Grade Crossing System.....</i>	<i>99</i>
Osileke Osipitan, New York State Department of Transportation	
<i>Epistemic Uncertainty Effects on Resistance Factors Based on Static Top-Down Load Tests of Drilled Shafts Installed in Sands.....</i>	<i>106</i>
Jiliang Li, Purdue University Northwest; Jinyuan Zhai, University of Akron; Li Tan, Purdue University Northwest	
<i>A Comparison of the Minimum Resolutions of Two Digital Image Correlation-Based Tools in Making Strain Measurements.....</i>	<i>113</i>
Niranjan Desai, Purdue University Northwest; Christos Georgakis, Technical University of Denmark; Gregor Fischer, Technical University of Denmark	
<i>Instructions for Authors: Manuscript Requirements.....</i>	<i>127</i>

IN THIS ISSUE (P.46)

LIDAR: LIGHT DETECTION AND RANGING

Philip Weinsier, IJERI Manuscript Editor

It will be hard to do justice to the topic of LIDAR—what it is and all of its applications—in these two columns. But let me begin with the following flavor for its current and potential uses, though this is by no means an all-inclusive list. And even if I were to choose but one area to focus on, there would be no room left in this issue for the authors. So I encourage you to read the article on page 46 by Ruggles and Ruggles, then look around the Internet for additional reading on the topic.

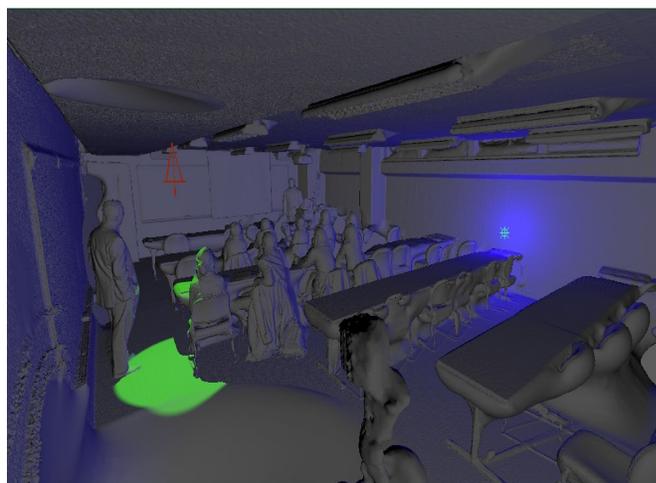
If you are only tangentially familiar with LIDAR, it's likely you know it from topography, with applications in agriculture, forest planning and management, parks and tourism management, environmental assessment and the protection of plants, biodiversity of birds and animals, watershed and stream delineation, ecological and land classification, and river surveys. But there is also plenty of interest in the air and water around us, where we can find applications in pollution modeling, coastline management, storm management, oceanography, and volume changes in glaciers. Digital elevation modeling (DEM) data can be used for mapping and transport planning for all types of infrastructure: railway, airports, roads, and building architecture; it can even generate information on slope. A somewhat related, or perhaps parallel, area might include accident scene investigation and forensic examination.

Oh, did I mention that several LIDAR systems are being deployed in self-driving cars? But from what I read about the current associated costs—over \$100,000 to outfit one car—most of us will be waiting a bit longer to get one.

Looking inward, we have oil and gas exploration by detecting atmospheric content of molecules. Also DIAL (Differential Absorption LIDAR) for sniffing out trace amounts of gases above hydrocarbon regions in order to find oil and gas deposits. Also related are mining, archeology (detecting micro-topography hidden by vegetation), and geology. Looking skyward, LIDAR is used in solar energy planning, astronomy, meteorology, and atmospheric physics, as it can be used to measure concentrations of gas particles in the upper atmosphere for creating density maps.

In hindsight, it might have been easier to talk about what LIDAR is NOT used for; I'm still thinking about that one and will get back to you!

The basic instrumentation of a LIDAR system includes a scanner, a laser, and a specialized GPS receiver. Systems can be ground-based or air-based. A topographical system would be used to map land-based structures, while a bathymetric one would use green light to penetrate the water to measure the elevations of riverbeds or seafloors. At a fundamental level, LIDAR is a method of remote sensing, similar to sonar and radar. But LIDAR uses light pulses from a laser for measuring distances, whereas sonar uses sound waves and radar uses radio waves. Light pulsing, or ranging, gathers very precise, 3-dimensional information. Some estimates put the sampling ability of LIDAR at over 150,000 pulses per second, creating millions of data points and accuracy within about 15 cm vertically and 40 cm horizontally. And while you end up with a massive amount of information to deal with, it's no wonder that the technology is finding so many applications, as noted earlier.



The authors of this current article describe the process of conducting the LIDAR scans, transferring the scanned information into an animation program, and the process of assessing points in the scene to establish the location for surveillance cameras. This site assessment process provides an analysis that reduces the cost of a surveillance system by decreasing the number of cameras required, while enhancing the surveillance coverage. In the image above, the authors present a visualization of two different light sources in Maya: a spotlight, the green emission on the left, and the point light, the purple emission on the right. For the rest of this fascinating story, I encourage you to read the full article.

Editorial Review Board Members

Mohammed Addallah	State University of New York (NY)	Shiyoung Lee	Penn State University Berks (PA)
Nasser Alaraje	Michigan Tech (MI)	Soo-Yen Lee	Central Michigan University (MI)
Aly Mousaad Aly	Louisiana State University (LA)	Chao Li	Florida A&M University (FL)
Jahangir Ansari	Virginia State University (VA)	Jimmy Linn	Eastern Carolina University (NC)
Kevin Berisso	Ohio University (OH)	Dale Litwhiler	Penn State University (PA)
Salah Badjou	Wentworth Institute of Technology (MA)	Guoxiang Liu	University of North Dakota (ND)
Pankaj Bhambri	Guru Nanak Dev Engineering (INDIA)	Louis Liu	University of New Orleans (LA)
Water Buchanan	Texas A&M University (TX)	Mani Manivannan	ARUP Corporation
Jessica Buck Murphy	Cal Poly State University (MS)	G.H. Massiha	University of Louisiana (LA)
John Burningham	Clayton State University (GA)	Thomas McDonald	University of Southern Indiana (IN)
Shaobiao Cai	Penn State University (PA)	David Melton	Eastern Illinois University (IL)
Vigyan Chandra	Eastern Kentucky University (KY)	Shokoufeh Mirzaei	Cal State Poly Pomona (CA)
Isaac Chang	Cal Poly State University SLO (CA)	Bashir Morshed	University of Memphis (TN)
Bin Chen	Purdue University Calumet (IN)	Sam Mryyan	Excelsior College (NY)
Wei-Yin Chen	University of Mississippi (MS)	Wilson Naik	University of Hyderabad (INDIA)
Hans Chapman	Morehead State University (KY)	Arun Nambiar	California State University Fresno (CA)
Rigoberto Chinchilla	Eastern Illinois University (IL)	Ramesh Narang	Indiana University-Purdue University (IN)
Phil Cochrane	Indiana State University (IN)	Anand Nayyar	Institute Management and Tech (INDIA)
Michael Coffman	Southern Illinois University-Carbondale (IL)	Stephanie Nelson	Cal State LA (CA)
Emily Crawford	Southern Wesleyan University (SC)	Hamed Niroumand	Universiti Teknologi (MALAYSIA)
Brad Deken	Southeast Missouri State University (MO)	Aurenice Oliveira	Michigan Tech (MI)
Z.T. Deng	Alabama A&M University (AL)	Troy Ollison	University of Central Missouri (MO)
Sagar Deshpande	Ferris State University (MI)	Reynaldo Pablo	Indiana University-Purdue University (IN)
David Domermuch	Appalachian State University (NC)	Basile Panoutsopoulos	Community College of Rhode Island (RI)
Ryan Dupont	Utah State University (UT)	Shahera Patel	Sardar Patel University (INDIA)
Marilyn Dyrud	Oregon Institute of Technology (OR)	Jose Pena	Purdue University Calumet (IN)
Mehran Elahi	Elizabeth City State University (NC)	Karl Perusich	Purdue University (IN)
Ahmed Elsayy	Tennessee Technological University (TN)	Thongchai Phairoh	Virginia State University (VA)
Rasoul Esfahani	DeVry University (OH)	Huyu Qu	Honeywell Corporation
Dominick Fazarro	Sam Houston State University (TX)	John Rajadas	Arizona State University (AZ)
Morteza Firouzi	University of Technology (MALAYSIA)	Desire Rasolomampionona	Warsaw University of Tech (POLAND)
Rod Flanigan	University of Nebraska-Kearney (NE)	Mulchand Rathod	Wayne State University (MI)
Ignatius Fomunung	University of Tennessee Chattanooga (TN)	Mohammad Razani	New York City College of Tech (NY)
Ahmed Gawad	Zagazig University (EGYPT)	Sangram Redkar	Arizona State University-Poly (AZ)
Daba Gedafa	University of North Dakota (ND)	Michael Reynolds	University of Arkansas Fort Smith (AR)
Ralph Gibbs	Eastern Kentucky University (KY)	Marla Rogers	Wireless Systems Engineer
Mohsen Hamidi	Utah Valley University (UT)	Dale Rowe	Brigham Young University (UT)
Mamoon Hammad	Abu Dhabi University (UAE)	Anca Sala	Baker College (MI)
Youcef Himri	Safety Engineer in Sonelgaz (ALGERIA)	Mehdi Shabaninejad	Zagros Oil & Gas Company (IRAN)
Xiaobing Hou	Central Connecticut State University (CT)	Ehsan Sheybani	Virginia State University (VA)
Shelton Houston	University of Louisiana Lafayette (LA)	Musibau Shofoluwe	North Carolina State University (NC)
Barry Hoy	St. Leo University (VA)	Siles Singh	St. Joseph University Tanzania (AFRICA)
Ying Huang	North Dakota State University (ND)	Ahmad Sleiti	University of North Carolina Charlotte (NC)
Charles Hunt	Norfolk State University (VA)	Jiahui Song	Wentworth Institute of Technology (MA)
Dave Hunter	Western Illinois University (IL)	Yuyang Song	Toyota Corporation
Christian Hyeng	North Carolina A&T University (NC)	Carl Spezia	Southern Illinois University (IL)
Pete Hylton	Indiana University Purdue (IN)	Michelle Surerus	Ohio University (OH)
Ghassan Ibrahim	Bloomsburg University (PA)	Vassilios Tzouanas	University of Houston Downtown (TX)
John Irwin	Michigan Tech (MI)	Jeff Ulmer	University of Central Missouri (MO)
Sudershan Jetley	Bowling Green State University (OH)	Mihaela Vorvoreanu	Purdue University (IN)
Rex Kanu	Ball State University (IN)	Phillip Waldrop	Georgia Southern University (GA)
Reza Karim	North Dakota State University (ND)	Abraham Walton	Purdue University (IN)
Tolga Kaya	Central Michigan University (MI)	Liangmo Wang	Nanjing University of Science/Tech (CHINA)
Satish Ketkar	Wayne State University (MI)	Jonathan Williams	Lake Erie College (OH)
Manish Kewalramani	Abu Dhabi University (UAE)	Boonsap Witchayangkoon	Thammasat University (THAILAND)
Tae-Hoon Kim	Purdue University Calumet (IN)	Alex Wong	Digilent Inc.
Doug Koch	Southeast Missouri State University (MO)	Shuju Wu	Central Connecticut State University (CT)
Sally Krijestorac	Daytona State College (FL)	Baijian Yang	Ball State University (IN)
Ognjen Kuljaca	Brodarski Institute (CROATIA)	Mijia Yang	North Dakota State University (ND)
Chakresh Kumar	Uttar Pradesh Tech University (INDIA)	Faruk Yildiz	Sam Houston State University (TX)
Zaki Kuruppallil	Ohio University (OH)	Yuqiu You	Morehead State University (KY)
Edward Land	Johns Hopkins Medical Institute	Jinwen Zhu	Missouri Western State University (MO)
Ronald Land	Penn State University (PA)		
Jane LeClair	Excelsior College (NY)		

EVALUATION OF SEMI-AUTOMATED TARGET EXTRACTION METHODS FOR DIGITAL CLOSE-RANGE PHOTOGRAMMETRY

Sagar Deshpande, Ferris State University

Abstract

Digital close-range photogrammetry is used for 3D reconstructions in many industries such as automobile and aircraft manufacturing, shipbuilding, and large-scale engineering construction. It commonly uses artificial targets to precisely locate a particular object on overlapping images. Image coordinates of the targets can be extracted manually or by implementing automated image processing algorithms. In this study, the author evaluated five target identification methods using a test field. Overlapping images were acquired of several targets placed on the test field using an off-the-shelf digital camera. These methods were then used to extract sub-pixel image coordinates of these artificial targets. Accuracy was evaluated based on 3D coordinates of these targets that were computed using the direct linear transform method. Based on the results, it was found that, among the five methods, the interpolation method produced the best results.

Introduction

In the manufacturing industry, high quality is achieved by fabricating parts or components as close to design specs as possible. Traditional methods can be used to make several measurements on a component to create a three-dimensional (3D) model, but they are time consuming and provide measurements at only a few locations. Similarly, advanced terrestrial scanners can also provide accurate 3D models. But, when compared to photogrammetry, they cannot be used in certain situations, due to high data acquisition times. On the other hand, digital close-range photogrammetry (DCRP) can be used to make several measurements on the object by capturing images from different locations. With the development in camera sensor technology, the quality and affordability of digital cameras have increased, resulting in increased use of close-range mapping applications. Therefore, DCRP systems, so-called vision metrology (VM) systems, have gained a footing for high-precision measurements on the commercial market.

Artificial targets are used in aerial photogrammetry and in close-range photogrammetry to facilitate precise photo-

graphic measurements. They are designed based on field conditions and camera properties. To achieve high photogrammetric accuracy, it is necessary to precisely identify the target's center. In this paper, five methods for extracting a target's center are presented. These methods use the digital number (DN) variations of targets on digital images. A close range photogrammetry setup was used to evaluate the accuracy achieved by these methods. Three sets of control targets (CTs) were used for camera calibration, using the direct linear transform (DLT) method. Thirty measure targets (MTs) were used for accuracy assessment by calculating standard errors.

Background Review

DCRP has been used in several applications for 3D reconstruction [1-7]. It is mainly used in mapping objects with limited time availability because of quick data acquisition capability [2-3, 8]. One of the primary, but tedious, tasks in photogrammetry is to identify common points on multiple overlapping images. Issues such as variable surface reflectivity, non-ideal illumination, and occlusion could make this task more complicated. Various image matching algorithms can be used to automate and expedite point-based matching. Generally, targets are used to facilitate and expedite matching. Targets can be classified as natural and artificial [4]. Natural targets are points or locations on an image with unique image properties, such as the corner of a road intersection. Artificial targets, which are typically symmetric in shape, are printed on a reflective material so that they can be easily identified in on an image [5], due to high contrast and shape properties.

Several researchers have experimented using of different target shapes such as circular [2, 4, 9-15], rectangular [2, 9], dots [3, 16], or other symmetric patterns [3, 4, 17]. Several researchers have implemented methods to extract the center of targets, depending on the shape of the target. A method to extract coordinates of crosses from analog continuous tone pictures using CCD (charged coupled device) cameras and comparing the mean error in image plane coordinates, was described by Jachimski and Trocha [18]. Based on spatial resolution, quantization accuracy, and noise, West and Clarke [19] experimentally analyzed and assessed four tech-

niques—centroid, weighted centroid, vernier, and interpolation. Use of a chess-board pattern is also a common target [20]. A chess-board pattern is extracted to the sub-pixel level by calculating the center of gravity of areas formed by connecting corners of high correlation [20]. Shortis et al. [17] presented coded targets that employed polar coordinate transformation and segment matching to automatically recognize and identify targets in digital images. Perez et al. [21] used a grid pattern provided by Photomodeler software for camera calibration. This software was designed to extract the predefined grid pattern for photogrammetric calculations.

Use of coded targets has been reported by several researchers to automatically label the identified targets [12] [11-15, 22]. They expedited the target matching process and removed error caused by human misinterpretation. Most of these methods are based on modeling the shape of the target on the image. Circular and square targets with thin cross-hairs at their centers were used. Five different semi-automatic methods, which were specifically developed for this target pattern, were used to extract the targets.

Methodology

Figure 1 shows that the methodology adopted for this study was divided into three stages: data acquisition, data reduction, and data presentation [5]. It should be noted that the overall accuracy of a photogrammetric system depends on the accuracy of target extraction. Therefore, this paper includes only detailed descriptions of the five methods.

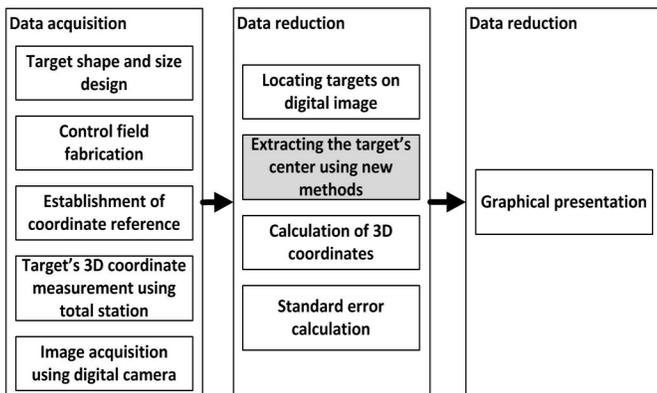


Figure 1. Methodology Adopted for This Study

Data Acquisition

In this stage, a 3D test field with dimensions of 1.0m (length), 0.6m (width), and 1.5m (height) was fabricated in the laboratory (see Figure 2).

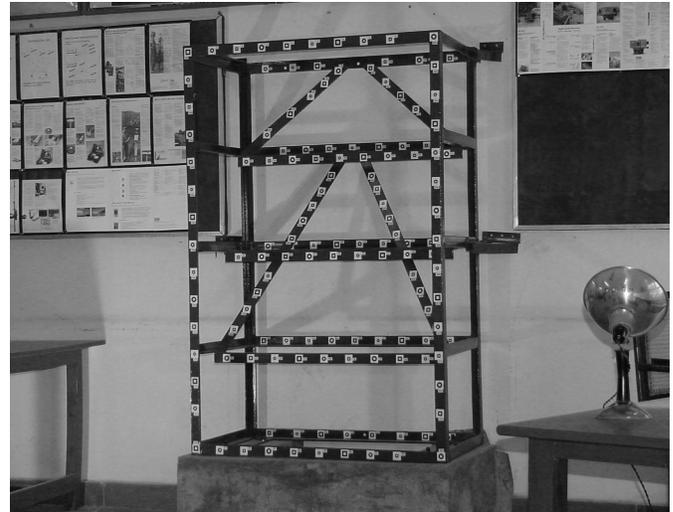


Figure 2. Fabricated Test Field in the Laboratory and Location of the Targets

Based on factors such as the nominal focal length of the camera (12.5 mm), the size of the CCD elements (6.3 microns), the object-to-camera distance (6.2 m), and the number of pixels that form a target on an image (12 to 15 pixels), three different sizes of targets were used: $l = 12$ mm, 16 mm, and 20 mm. These targets were printed on high-quality photographic paper, using a laser plotter and were fixed on the test field. The thickness, t , of the target was designed to be equal to 3 mm (see Figure 3).

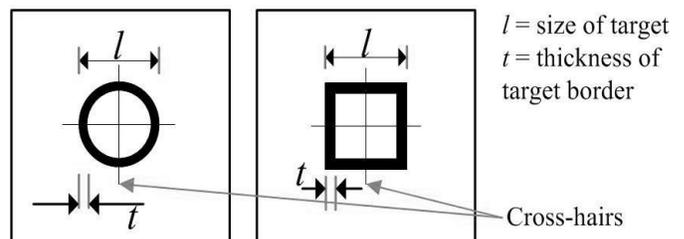
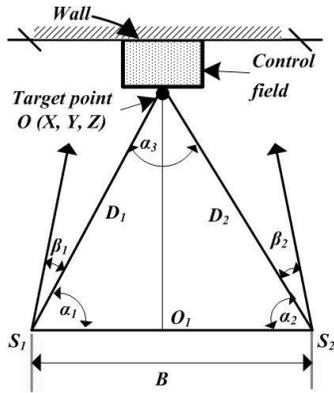


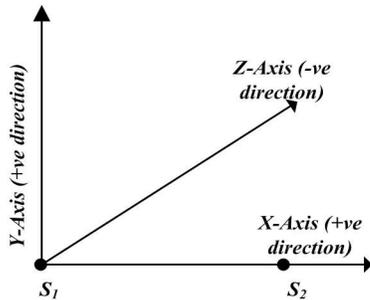
Figure 3. Shape and Dimensions of the Targets

The 3D coordinates of these targets were required for camera calibration and accuracy assessment. Therefore, the next step was to measure the coordinates of these targets. Figure 4(a) shows how a coordinate reference was established with its origin at point S_1 . The 3D coordinates of point S_1 was assumed to be 100m, 100m, and 100m in the X , Y , and Z directions, respectively. Figure 4(b) shows that the X axis passes through station point S_2 , the Y axis is parallel to the vertical plumb direction, and the Z axis is perpendicular to both X and Y directions. Vertical and horizontal angles to each target were measured from stations S_1 and S_2 , using Trimble's total station (model SR5600 DR+200). This instrument has an angular precision of ± 1 second. Crosshairs printed on the targets were used to make accurate

angular measurements. Distance B between stations S_1 and S_2 was measured using fiberglass tape and a diagonal scale. Using the horizontal and vertical angular measurements from stations S_1 and S_2 , and the distance B , 3D coordinates of all the targets were calculated using trigonometric relationships.



(a) Top View of the Two Points S_1 and S_2 and the Test Field



(b) 3D View of the Coordinate System Located at Station S_1

Figure 4. Schematic Representation of the 3D Coordinate System

Figure 5 shows how the images of the test field were then acquired from approximately 6.2m using an off-the-shelf digital camera (Sony MVC-FD95, a 2-megapixel camera). The field of view of cameras was kept such that the axes of the cameras converged onto the test field. These images were transferred to a computer to perform the steps in the data-reduction stage.

Data Reduction

Extraction of the image coordinates of the targets, calculation of object-space coordinates using image coordinates, and calculation of the standard error of the computed targets were the steps in this stage. The target-extraction step consisted of two sub-steps: to locate the targets on the digital image and to extract the center of each located target.

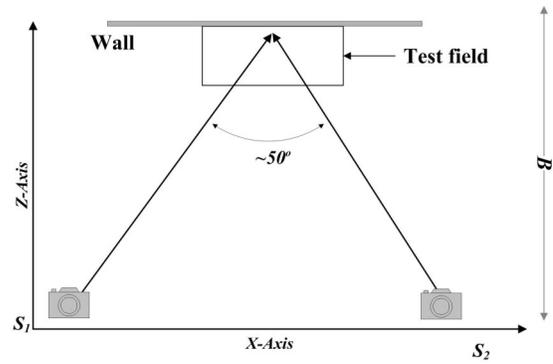
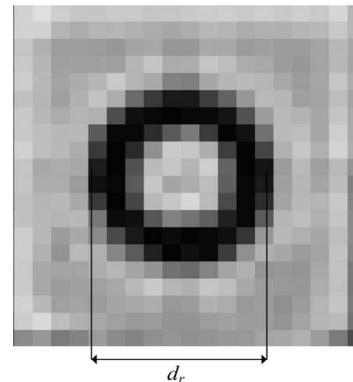
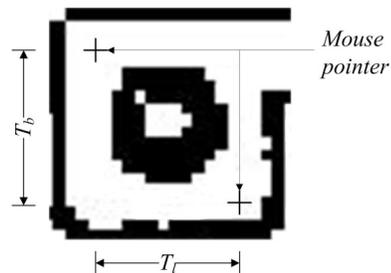


Figure 5. Top View of the Camera Positions and Test Field during Image Capture

In the first sub-step, edge detection [23] was performed on the original image [see Figure 6(a)] to create a binary image [see Figure 6(b)]. This binary image consisted of 0s (black color) at edge locations on the image and 1s (white color) for the rest of the image. A sample target was then manually defined using a mouse pointer, by specifying a representative target on a binary image [see Figure 6(b)]. This target was used for template matching, using cross-correlation [24] to locate other similar targets on the image. Since the entire test field had targets of different sizes and shapes, this step was repeated (usually 4-5 times) until all of the targets used in this study were located.



(a) Target Formed on the Original Image



(b) Target Formed on the Binary Image after Edge Detection

Figure 6. Sample Target

In the second sub-step, each located target was processed to find the central area of the target. In this process, the original image and the binary image were used interchangeably. Figure 7 shows the step-by-step procedure, which is explained below:

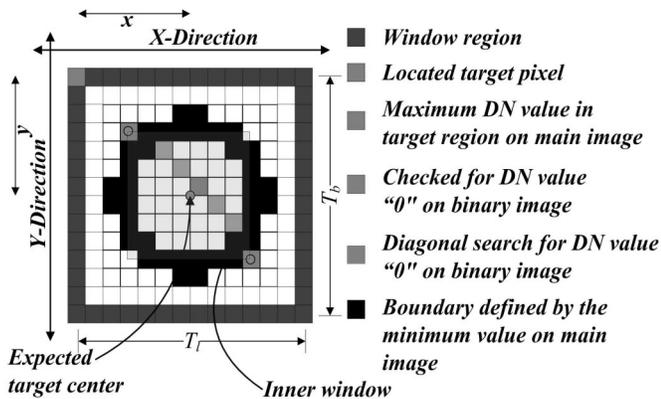


Figure 7. A Schematic Target Showing Different Features Used for Target Extraction

1. A window equal to the size of a manually specified target was placed around a located target. A diagonal search was performed on the binary image from the upper left and lower right corners of the window to locate the “0” DN value. An inner window was obtained by connecting the two “0” DN value locations.
2. As the area within this region contained only the target’s ring and the inner region, the maximum DN value within the inner window was identified using the original image. This search would point to a pixel location in the inner region of the target.
3. A “0” DN value was searched on the binary image up to a distance of half of the sample target ($T_b/2$ or $T_l/2$) diagonally away from the maximum DN value location. All rows and columns containing the pixel along the diagonal direction between the “0” DN value locations were used for extracting the target’s center.
4. For each row and column, the border pixels were identified on the original image by searching for the minimum DN value, up to a distance of half of the sample target ($T_b/2$ or $T_l/2$) on either side of the diagonal pixel.

After locating the inner extents of the target, all rows and columns within similar boundaries were used for identification of the target’s center. The following methods were developed to locate a target’s center to the sub-pixel level.

Method 1: Interpolation

The interpolation method was modified, based on the interpolation method by Jachimski & Trocha [18]. Compared to

their method, the modified method used fewer points along the profile. DN values in each row and column of a target, when plotted, resembled a bell-shaped curve. Figure 8 shows a sample profile in which the lower DN values at 4 and 13 locations along the horizontal axis are represented by the black border of the target, while small depressions in the central portion (at position 9) of profile was due to the crosshairs.

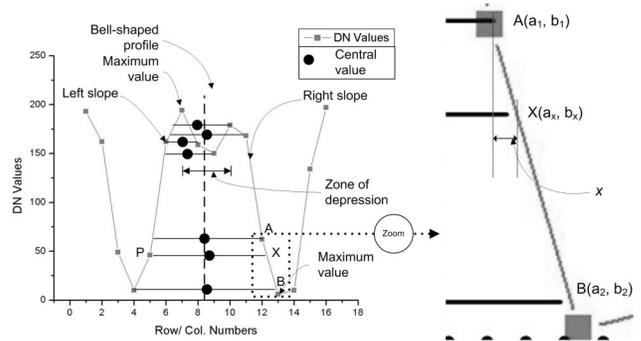


Figure 8. Profile of the DN Values in the Target Showing the Interpolation Method

To calculate the center, each point on the profile was processed individually. Point P in Figure 8 is used as an example to explain the process. It can be noted that the target’s approximate center is half way between points P and X. Point X, which is equal to the DN value at P, represents an interpolated location on the right profile between points A and B. Therefore, distance PX is equal to the sum of distances PA and AX, measured horizontally. Points P and A are located at 5 and 12, respectively. Using the gradient between points A and B, distance AX = x can be calculated. Therefore, distance PX can be determined using Equation (1):

$$PX = 12 - 5 + x \tag{1}$$

The center was considered to be located at a distance of $PX/2$, which is shown as the “Central value” in Figure 8. Several such center locations were calculated for every DN value in every row within the target’s boundary. The averaging and thresholding process was implemented on these calculated center locations to obtain the target’s center coordinates along the row. This procedure was also implemented for all DN values along the column within the target’s boundary to obtain the target’s center coordinates along the column.

Method 2: Cubic Spline

This method was adapted using the polynomial line-fitting approach [18] and utilized the cubic spline interpolation to estimate the center of the target along a row or column profile. The cubic spline function is a third-order curve em-

ployed to connect each pair of data points so that the connections between adjacent cubic equations are smooth. Therefore, this method was used to find the target's center. Figure 9 shows the profile of DN values within the target for a typical row. In this figure, a dip in DN value can be observed between pixel locations 3 and 7. This was due to the crosshair located at the center of target. Therefore, based on observation of the profiles, the middle third of the pixel values were excluded for this method. Figure 9 also shows how a cubic spline equation [25] was fitted to the remaining DN values. The location at which this cubic spline attained the maximum value was considered as the center for a particular row. Therefore, one center value was calculated for every row within a target. The averaging and thresholding process was implemented on these calculated center locations to obtain the target's center coordinates along the row.

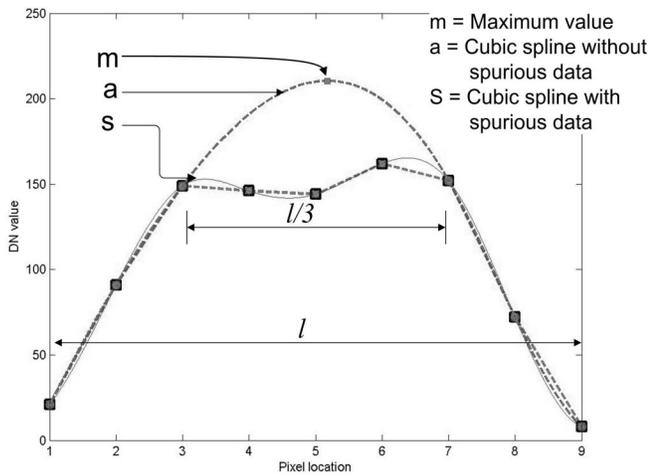
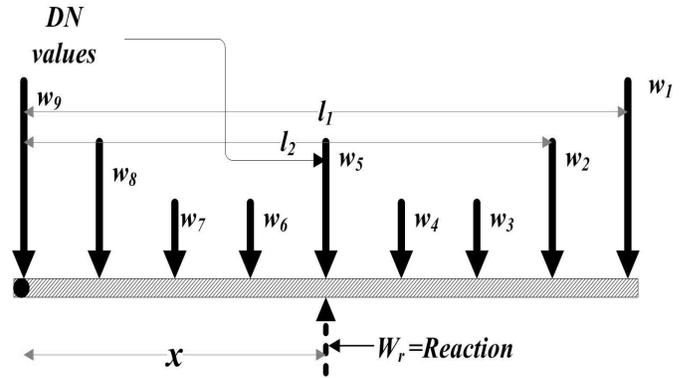


Figure 9. Graphical Representation of the Cubic Spline Method

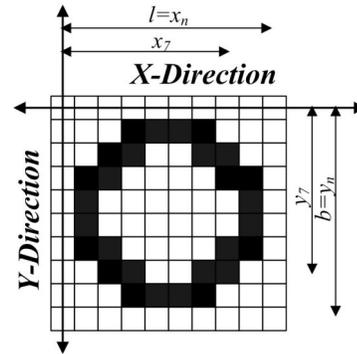
This procedure was also implemented for all columns within a target's boundary to obtain the target's center coordinates along the column.

Method 3: Moment of DN Values

This new method utilized a principle from engineering statics. In Statics, moment of a force is a measure of its tendency to cause a body to rotate about a specific point or axis. In order to keep a body stable, the magnitude of positive and negative moments should be equal. This principle was used to determine the target's center. This method was implemented for all of the pixels in the horizontal direction within the target's boundary. The left-most column containing the border pixel was considered as an axis about which positive and negative moments were calculated [see Figure 10(b)]. Each row in a target's boundary was assumed to be a horizontal beam with the DN values as forces acting vertically downward [see Figure 10(a)].



(a) DN Values Acting on a Beam



(b) Moments Calculated Along the X and Y Axes

Figure 10. Moment-of-Area Method

Figure 10(a) shows the axis as pivot point; W 's represent the n number of DN values within the target's border; l 's represent the distance to these pixels from the axis; and W_r , which is vertical, represents a reaction force equal to the sum of all W 's. By equating the positive and negative moments, the value of x was calculated using Equation (2):

$$x = \frac{\sum_{i=1}^n W_i l_i}{\sum_{i=1}^n W_i} \quad (2)$$

Similar calculations were implemented in the column direction to calculate the target's center.

Method 4: Average Number of Pixels below Threshold

This method was adapted based on similar center-of-gravity methods proposed by other researchers [14, 26]. In this method, pixels in a target region, which have DN values below a certain threshold, were used to calculate the center of the target.

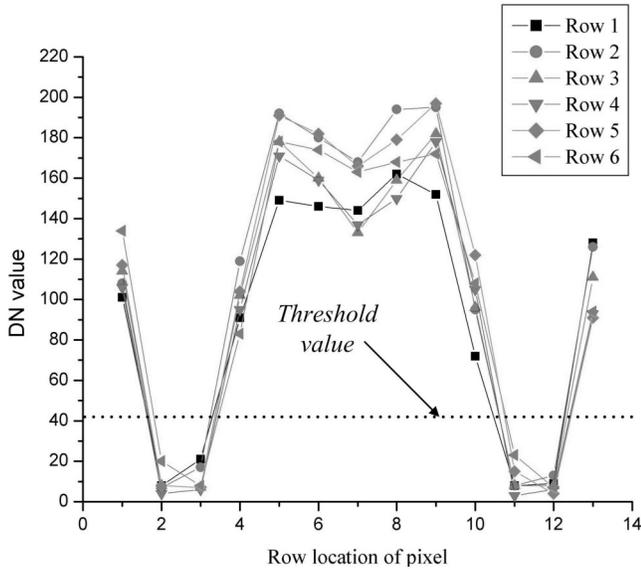


Figure 11. Implementation of the Average-Below-Threshold Method

Figure 11 shows typical DN values profiles along the rows within a target. Based on the observation that the pixels along the target's black border had DN values below 40, a minimum DN threshold value of +40 was adopted to shortlist the pixels that formed the target's border. The average of the row and column numbers for shortlisted pixel locations was calculated using Equation (3):

$$x = \frac{\sum_{i=1}^n C_i}{n} \quad y = \frac{\sum_{i=1}^n R_i}{n} \quad (3)$$

where, x and y are the image coordinates of the target's center; C is the column number of shortlisted pixels; R is the row number of shortlisted pixels; and, n is the number of shortlisted pixels.

The calculated x and y values were considered as the target's center image coordinates along the rows and columns, respectively.

Method 5: Weighted Average-Below-Threshold

This method was an extension of the aforementioned method. In this method, weighted average values of the pixel locations below the threshold value were calculated using Equation (4):

$$x = \frac{\sum_{i=1}^n W_i C_i}{\sum_{i=1}^n W_i} \quad y = \frac{\sum_{i=1}^n W_i R_i}{\sum_{i=1}^n W_i} \quad (4)$$

where, W represents the DN values in n number of pixels.

The calculated x and y values were considered as the target's center image coordinates along the rows and columns, respectively. In the interpolation and spline methods, several center locations were calculated in row and column directions for a target. The averaging and thresholding process was used to calculate the average value for the target's center in the horizontal and vertical directions. This method was modified based on the work by Jachimski and Trocha [18]. Figure 12 shows the method used to calculate the center.

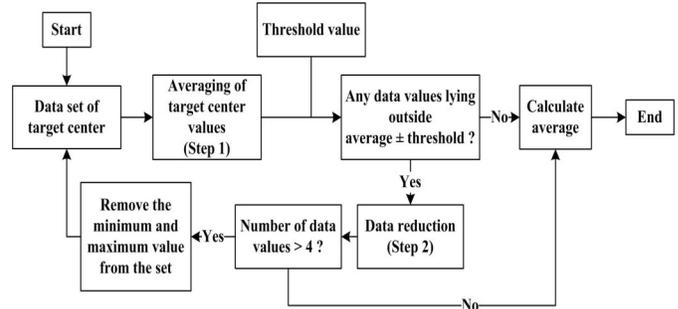


Figure 12. Flowchart Describing the Averaging and Thresholding Method

In the first step, the average (μ) of the set of multiple center coordinates along the rows was calculated. A check was performed to find values in the set that were greater than $\mu \pm$ the threshold. If any values were greater than $\mu \pm$ the threshold then the minimum and maximum values were removed from the set. This process was repeated until either the number of coordinates in the set was more than four or if no values were outside the threshold criteria. The average of all the retained values in the set was considered as the center value along the rows. The same procedure was implemented along the columns. Based on experimentation, a threshold of 0.2 pixels was used in this study.

In this study, the direct linear transform (DLT) calibration approach [27] was used to calibrate the non-metric cameras and to calculate object-space coordinates. DLT establishes a direct linear relationship between coordinates of image points and the corresponding object-space coordinates. Equation (5) shows how DLT can be derived from the standard collinearity equations:

$$\begin{aligned} x &= \frac{L_1 X + L_2 Y + L_3 Z + L_4}{L_9 X + L_{10} Y + L_{11} Z + 1} \\ y &= \frac{L_5 X + L_6 Y + L_7 Z + L_8}{L_9 X + L_{10} Y + L_{11} Z + 1} \end{aligned} \quad (5)$$

where, L_1 to L_{11} represent the 11 DLT parameters; X , Y , and Z are the 3D coordinates of a point; and, x and y are the image plane coordinates.

Using the five methods discussed here, image coordinates of the targets were extracted. Thirty targets were used as measuring targets (MTs), while three sets consisting of 6, 10, and 15 targets were used as control targets (CTs). Using the image coordinates and the ground 3D coordinates of CTs in the DLT method, camera calibration parameters L_1 to L_{11} were calculated. Using these parameters and the image coordinates of the 30 MTs, 3D coordinates of the MTs and their corresponding standard error were calculated. Standard variance of the adjusted quantities [28] is given by Equation (6):

$$\Sigma = \sigma_o^2 N^{-1} \quad (6)$$

where, N^{-1} is the variance-covariance matrix of the computed coordinates, shown in Equation (7), and σ_o^2 is the reference standard variance.

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_z^2 \end{bmatrix} \quad (7)$$

The square root of the diagonal elements of the matrix in Equation (7), σ_x^2 , σ_y^2 , and σ_z^2 represent the standard error values in the X , Y , and Z directions, respectively.

Results and Discussion

Results were presented in the third stage. The average of the standard error values in the X , Y , and Z directions were computed for the 30 MTs. Results from the previous study [9] using manually picked targets were also included in the comparison with the results obtained from the five methods. These results were presented in Table 1 and were graphically represented in Figure 13. The effects of the number of control points in the X , Y , and Z directions were also studied.

The following observations were made from results of this study.

1. Performance of the interpolation method was consistently better than the remaining four methods. This may have been due to the fact that several center locations were calculated in this method to calculate the sub-pixel target's center coordinates. The lowest average standard error values in the X , Y , and Z directions were 63 μm , 27 μm , and 113 μm , respectively, for 10 CTs. The performance of the interpolation method was least affected by the variation in number of CTs.

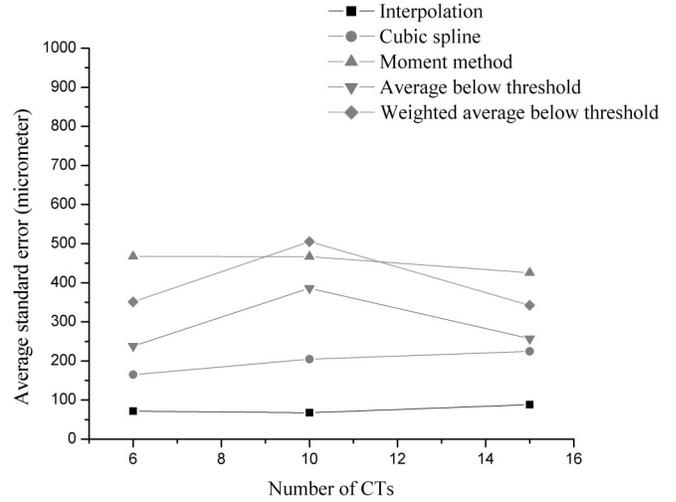


Figure 13. Graphical Representation of the Average Standard Error

Table 1. Standard Error in Micrometer (μm) for 30 MTs

Number of CTs→	6			
Method #↓	σ_x	σ_y	σ_z	σ_{Avg}
1	67	29	119	72
2	153	73	269	165
3	448	175	779	467
4	228	96	390	238
5	332	149	573	351
Manual	682	544	4309	1845
Number of CTs→	10			
Method #↓	σ_x	σ_y	σ_z	σ_{Avg}
1	63	27	113	68
2	189	85	340	205
3	433	198	769	467
4	358	158	642	386
5	464	213	838	505
Manual	492	385	3086	1321
Number of CTs→	15			
Method #↓	σ_x	σ_y	σ_z	σ_{Avg}
1	83	35	147	88
2	207	93	373	224
3	393	185	698	425
4	243	103	426	257
5	322	139	567	343
Manual	374	293	2434	1034

2. The cubic spline and the average-below-threshold methods performed similarly. The lowest error achieved by the cubic spline method were 153 μm , 73 μm , and 269 μm , in the X , Y , and Z directions, respectively, for six CTs.

3. Performance of the weighted average and the moment methods were the poorest with the lowest σ_{Avg} values of 425 μm and 343 μm , respectively. The poor performance could have been because these methods calculated a single value compared to the interpolation or cubic spline method, which calculated sets of values and used the averaging and thresholding method.
4. Maximum error was observed in the Z direction for all of the five methods. The σ_z values for the average-below-threshold, weighted average, and the moment method were more than 400 μm in the Z direction. The effect of change in number of CTs was particularly observed in the Z direction for all of the methods, except for the interpolation method.
5. For all of the five methods, values of σ_y were least affected by the change in number of CTs.
6. Results of all the five methods were better than that of manually selected targets. A significant improvement was observed in the Z direction.

Conclusion

In this paper, the author described five methods for extracting targets to the sub-pixel level using an off-the-shelf digital camera. These methods used cross-correlation to locate the targets and then perform analyses using DN values to extract targets at the sub-pixel level. Three sets of CTs were used to analyze the impact of the number of control points on the accuracy. Based on the results, it was concluded that using an off-the-shelf camera and precise methods for target extraction, micron-level accuracy can be achieved. It should also be noted that if the targets are extracted at sub-pixel level then the impact of number of CTs was not significant. Further, these methods were semi-automatically implemented. Therefore, errors caused by human misinterpretation of the target's center were removed.

References

- [1] Bertin, S., Friedrich, H., Delmas, P., & Chan, E. (2013). On The Use of Close-Range Digital Stereo-Photogrammetry to Measure Gravel-Bed Topography in a Laboratory Environment. In *Proceedings of 35th IAHR Congress* (pp. 8-13).
- [2] Bernat, K., & Tokarczyk, R. (2013). Automation of Measurements of Selected Targets of Photopoints in Application to Photogrammetric Reconstruction of Road Accidents. *Geomatics and Environmental Engineering*, 7.
- [3] Maas, H. G., & Hampel, U. (2006). Photogrammetric techniques in civil engineering material testing and structure monitoring. *Photogrammetric Engineering & Remote Sensing*, 72(1), 39-45.
- [4] Guarnieri, A., Vettore, A., El-Hakim, S., & Gonzo, L. (2004). Digital photogrammetry and laser scanning in cultural heritage survey. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35, B5.
- [5] Deshpande, S. (2004). *Measurement of displacement using digital close range photogrammetry (DCRP)*. (Master's thesis, IIT Kanpur, India).
- [6] White, D. J., Take, W. A., Bolton, M. D., & Munachen, S. E. (2001). A deformation measurement system for geotechnical testing based on digital imaging, close-range photogrammetry, and PIV image analysis. In *Proceedings of the international conference on soil mechanics and geotechnical engineering* (pp. 539-542). AA Balkema Publishers.
- [7] Dai, F., Feng, Y., & Hough, R. (2014). Photogrammetric error sources and impacts on modeling and surveying in construction engineering applications. *Visualization in Engineering*, 2(1), 2.
- [8] Habib, A. F., Ghanma, M. S., & Tait, M. (2004, July). Integration of LIDAR and photogrammetry for close range applications. In *Proceedings of the ISPRS XXth Conference*, Istanbul, Turkey B.
- [9] Kadari, K. (2003). *Network designing of a digital close range photogrammetric system*. (Master's thesis, IIT Kanpur, India).
- [10] Csanyi, N., & Toth, C. K. (2007). Improvement of lidar data accuracy using LiDAR-specific ground targets. *Photogrammetric Engineering & Remote Sensing*, 73(4), 385-396.
- [11] Hattori, S., Akimoto, K., Fraser, C., & Imoto, H. (2002). Automated procedures with coded targets in industrial vision metrology. *Photogrammetric engineering and remote sensing*, 68(5), 441-446.
- [12] Van Den Heuvel, F. A., Kroon, R. J. G. A., & Le Poole, R. S. (1993). Digital close-range photogrammetry using artificial targets. *International Archives of Photogrammetry and Remote Sensing*, 29, 222-222.
- [13] Li, L., & Liu, M. (2010). Research on decoding method of coded targets in close-range photogrammetry. *Journal of Computational Information Systems*, 6(8), 2699-2705.
- [14] Ahn, S. J., Rauh, W., & Kim, S. I. (2001). Circular coded target for automation of optical 3D-measurement and camera calibration. *International journal of pattern recognition and artificial intelligence*, 15(06), 905-919.

-
- [15] Mills, D., & Carty, G. (2005). Semi-automated crush determination using coded and non-coded targets with close-range photogrammetry. Retrieved from <http://www.PhotoModeler.com/applications/documents/MillsCodedTargetsCrush.pdf>.
- [16] Chen, J., & Clarke, T. A. (1993). The automatic recognition, location and labelling of targets in digital photogrammetric engineering measurement. *International Archives of Photogrammetry and Remote Sensing*, 29, 686-686.
- [17] Shortis, M. R., Seager, J. W., Robson, S., & Harvey, E. S. (2003, January). Automatic recognition of coded targets based on a Hough transform and segment matching. *Electronic Imaging*, 202-208. International Society for Optics and Photonics.
- [18] Jachimski, J., & Trocha, W. (1993). Determination of the Position of Crosses with Subpixel Accuracy on the Image Taken with the CCD Camera. *International Archives of Photogrammetry and Remote Sensing*, 29, 391-391.
- [19] West, G. A. W., & Clarke, T. A. (1990, August). A survey and examination of subpixel measurement techniques. *Close-Range Photogrammetry Meets Machine Vision*, 1395, 456-463.
- [20] Wu, B., Hu, H., Zhu, Q., & Zhang, Y. (2013). A Flexible Method for Zoom Lens Calibration and Modeling Using a Planar Checkerboard. *Photogrammetric Engineering and Remote Sensing*, 79(6), 555-571.
- [21] Pérez, M., Agüera, F., & Carvajal, F. (2011). Digital camera calibration using images taken from an unmanned aerial vehicle. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(1), C22.
- [22] Clarke, T. A. (1994, October). Analysis of the properties of targets used in digital close-range photogrammetric measurement. *Photonics for Industrial Applications*, 251-262. International Society for Optics and Photonics.
- [23] Gonzales, R. C., & Woods, R. E. (2002). *Digital Image Processing*, 6, 681.
- [24] Schenk, T. F. (1999). *Digital Photogrammetry: Backgrounds, Fundamentals, Automatic Orientation Procedures*. TerraScience.
- [25] Chapra, S. C., & Canale, R. P. (2010). *Numerical methods for engineers*. New York: McGraw-Hill Higher Education.
- [26] Li, Z., & Liu, M. (2010). Research on decoding method of coded targets in close-range photogrammetry. *Journal of Computational Information Systems*, 6(8), 2699-2705.
- [27] Abdel-Aziz, Y. I., & Karara, H. M. (1971). Direct linear transformation from comparator coordinates into object-space coordinates in close-range photogrammetry. *Proceedings of the Symposium on Close-Range Photogrammetry*. Falls Church, VA: American Society of Photogrammetry.
- [28] Wolf, P. R., & Ghilani, C. D. (1997). *Adjustment computations: statistics and least squares in surveying and GIS*. New York, NY, USA: John Wiley & Sons.

Biographies

SAGAR S. DESHPANDE is an assistant professor at Ferris State University, Department of Surveying Engineering. He earned his MTech degree in Civil Engineering from the Indian Institute of Technology Kanpur, India, in 2004 and MS degree in Geodetic Science from the Ohio State University in 2008. His current research interests include digital photogrammetry, remote sensing, and GIS. Professor Deshpande may be reached at sagardeshpande@ferris.edu

PERFORMANCE VALIDATION OF AN ENERGY EFFICIENT ELEVATOR CONTROLLER

Muhammad Z Hasan, Texas A&M International University; Rainer Fink, Texas A&M University;
Daryl James, Texas A&M University; Jesus Gamboa, Texas A&M International University

Abstract

In smart buildings, measures are being taken to conserve energy without degradation in performance. Elevators are an integral part of many smart, high-rise buildings. To achieve notable energy savings in elevator systems, the authors of this study experimentally tested an algorithm that was developed earlier. Simulation results from an analytical model running the algorithm were encouraging. An elevator controller is capable of saving energy by manipulating elevator speed during operation. Speed is varied based on the difference between load carried and the counterweight. A miniature elevator model was used to carry out the experiments and to collect energy data. The algorithm was tested on both pre-determined and random traffic patterns. Voltage and current samples were collected from the miniature elevator model. The final energy consumption of the proposed variable-speed system was calculated and compared with energy data of a constant-speed system. This proposed method resulted in a 3.715% energy savings for pre-determined traffic pattern and 8.7612% energy savings for random traffic patterns.

Introduction

Recently, energy savings in electrical systems is being explored as one of the prime factors in sustainable development. The use of elevator systems has increased rapidly in urban areas with the growing population. A study of the number of elevators employed in several countries [1] found that Italy had the highest number. In general, any effort towards reducing energy consumption tends to degrade the performance of the system. Thus, there is a necessity to optimize performance and energy with an ideal trade-off. One type of elevator is the traction elevator that has become widely prevalent in recent years. In such systems, the car is suspended by ropes wrapped around a sheave that is driven by an electric motor. A counterweight that equals the mass of the car plus 45% to 50% of the rated load is used to balance the weight of the car. The counterweight is used to ensure that a constant potential energy is maintained in the system [2].

Traction elevators are of two types: 1) geared lifts, typically used in midrise applications, where high speed is not

an important factor and a reduction gear is utilized to reduce the speed of the motor; and, 2) gearless lifts, used in high-rise applications, where the sheave is driven directly by the motor eliminating the losses in the gear. In such a case, both motor and sheave rotate at the same speed [2]. The purpose of the counterweight is to maintain sufficient tension in the suspension system. This ensures adequate traction between ropes and drive sheave. The counterweight also maintains a near-constant potential energy level in the whole system, heavily reducing energy consumption [2].

Significant development has been achieved in optimizing elevator controllers in terms of energy efficiency and reducing average waiting and transit time. This involves implementation of artificial intelligence and fuzzy logic in controllers to optimize the service parameters. In this study, the authors dealt with experimental speed manipulation of the elevator, based on load to achieve energy savings and considering its traffic intensity. A trade-off between speed and load torque was proposed, such that their product was constant. This paper includes experimental verification of the proposed idea using a miniature elevator model.

Related Work

Recently, many ideas have been proposed for energy savings in elevator systems. Following is a summary of achievements in single-elevator systems. Effective energy savings can be made through various means such as regenerative energy feedback and loss reduction in order to obtain optimum utilization of energy. Effective energy savings can also be made through energy storage and discharge using capacitors [3] as well as speed manipulation, as considered in this study. A study proved that, for a prolonged duration, a traffic pattern exists that repeats day to day in multi-storied buildings [4]. Based on this fact, a simulation method was developed to analyze the energy consumption of elevators under varying load and traffic patterns. This method was employed to compare energy savings of various drive systems and machinery as well as control systems. This work provided an initial base for the current research on energy consumption of elevator systems. Furthermore, efforts have been reported attempting to reduce waiting and transit time of the elevators. Current research implemented artificial intelligence (AI) and neural networks in elevator group control systems. This involved assigning the elevator

cart based on their demand at that instant of time. In these efforts, elevators underwent a period of training with a definite traffic pattern corresponding to the building. This has eased the process of prioritizing the floors based on the requirement during up-peak and down-peak hours of the day. Energy consumption was reduced, even in optimal assignment of an empty cart, based on the need of the hour [5, 6].

Energy conservation can also be made more effective in elevator group systems with intelligent algorithms. These algorithms control all of the lifts in the system, ensuring optimal assignment of each cage to a particular floor [7]. This effort resulted in saving energy by reducing the average number of stops from 1600 to 1400 for a traffic pattern. One recent study involved development of an ant colony algorithm in elevators towards energy conservation during peak traffic flow [8].

There were attempts to strike a balance between energy savings and performance. A recent study reported development of a genetic algorithm towards energy savings in elevators [9]. The algorithm was 23.6% effective in energy conservation. However, the authors reported increased waiting time and service time by 64.9% and 39.5%, respectively. This reveals that energy savings beyond a particular limit may degrade the performance of the elevator. As such, the aim is towards energy savings that does not affect the performance of the system significantly. A study with such simulations was reported [10-12] that used traffic patterns from several other studies [13-15]. Simulation results showed a trade-off between energy savings and travel time. Authors of this current study considered energy conservation in traction elevators driven by electric motors. Normally, traction elevators move at a constant speed, irrespective of the load it carries. This leads to inefficient use of energy. Adjustment of speed based on load is needed for more energy-efficient operation [12].

The Proposed Algorithm, Analytical Model, and Simulation Results

The authors of this study considered energy conservation in traction elevators. With this focus, speed of the elevator was varied, considering the load carried by the elevator cart. In determining the load, the authors considered the counterweight of the elevator and also measured the travel time to estimate its overall performance.

The initial idea was proposed as follows:

1. To achieve maximum possible efficiency, operate the motor at its rated full-load power.

2. To maintain this efficiency, vary the steady speed of the elevator, based on the weight carried, such that the motor operates closer to its rated full-load power [14].

Elevator motor rating, R , is related to its out-of-balance load, B , its rated linear speed, v , and its efficiency, ζ , as shown in Equation (1) [15]:

$$R = 9.81 (B v) / (\zeta) \quad (1)$$

Thus, any change in out-of-balance load gets adjusted by a subsequent variation in the speed of the elevator in order to maintain a fixed efficiency.

3. The upper limit on speed is decided by the following relationship, taking passenger comfort into consideration.

$$\text{TimeConstant} \geq 4\text{TimeAccDecel}$$

which is termed as the time-constraint equation. In this equation, TimeConstant is the time the elevator runs at a constant speed, and TimeAccDecel is the time for acceleration or deceleration. The upper limit on speed is found by dividing the distance to travel by the TimeConstant.

This proposed algorithm is referred to as variable-speed algorithm in later parts of this paper. In this algorithm, the lower limit for the elevator speed is the speed at which the elevator would run if the variable-speed algorithm were not applied. Figure 1 shows the diagram representing the traction system considered in the simulation [13]. The maximum load on the elevator was 400 kg with a counter weight of 300 kg. T_M is the motor torque (in N-m), v is the speed of the elevator (in m/s), and r is the radius of the pulley (in m). Equations (2)-(7) represent various relationships used to calculate the energy [13].

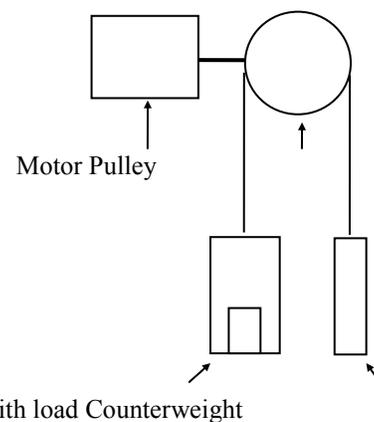


Figure 1. Schematic Representation of the Elevator System

$$LT = (Mu + Mv - Mc)gr \quad (2)$$

where, LT is the load torque (in N-m) about the center of the motor shaft; Mu is the mass of the load; Mv is the mass of the elevator cart; Mc is the mass of the counterweight; and g is the gravity.

$$v = \text{linear velocity} = 2\pi \text{ rps} \quad (3)$$

$$\Omega M = \text{angular velocity} = 2\pi \text{ rps} \quad (4)$$

$$\Omega M, \text{ Angular velocity (in radians/second)} = v / r \quad (5)$$

$$\theta = \text{Angular displacement} = \Omega M \text{ travel time} \quad (6)$$

$$E = \text{Energy} = TL \times \Omega M \quad (7)$$

The following three algorithms were simulated under multiple traffic patterns, and the simulation results were tabulated and compared.

1. *Constant-speed case*: elevator running at a constant speed of 1m/s. This was used as the reference for comparison of energy consumption and travel time.
2. *Speed variation I*: elevator changing its speed based on load but with a fixed acceleration.
3. *Speed variation II*: elevator changing its speed based on the load as well as changing acceleration based on the number of levels moved. (a modified version of *speed variation I*).

Figure 2 show the basic operation of a constant-speed elevator system. According to the calls, the elevator controller determines the direction of movement and satisfies the calls, based on a fixed priority (floors along the direction of movement are given priority). Figure 3 shows the modified version of the algorithm. This algorithm uses the “load on the elevator” to determine the speed of movement and the “number of floors to move” to determine acceleration. The elevator then satisfies the calls similar to the constant-speed case. Figure 3 shows how the calls are satisfied in a five-floor building using this algorithm.

Case 1: Consider the elevator of Figure 1 running in a 10-level building during peak-hour traffic. The three algorithms were simulated. Table 1 shows the simulation results of all the algorithms over the same traffic intensity (peak hour).

This table also shows that the *Speed variation II* algorithm reduces both energy consumption and travel time, as compared to the constant-speed type. Hence, energy savings are achievable at a reduced travel time for peak-hour traffic patterns using *speed variation II*.

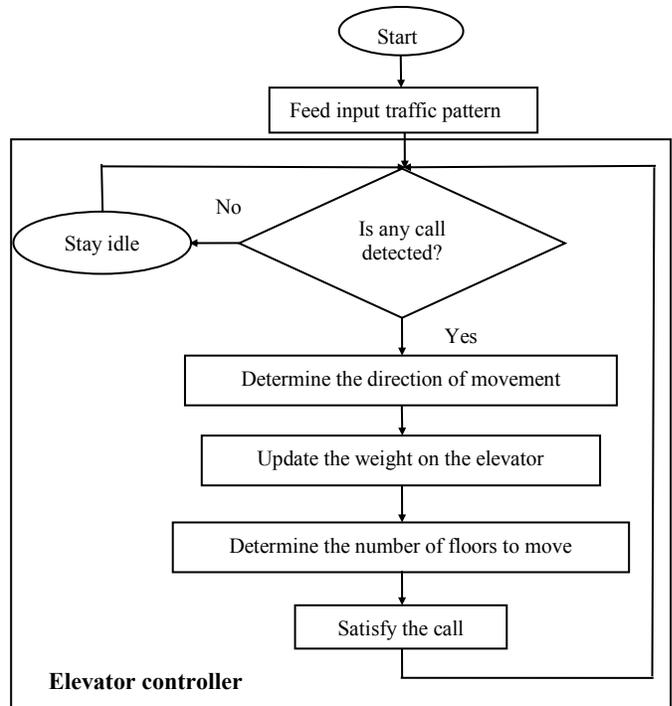


Figure 2. Flow Diagram Elaborating the Algorithm of the Constant-Speed Case

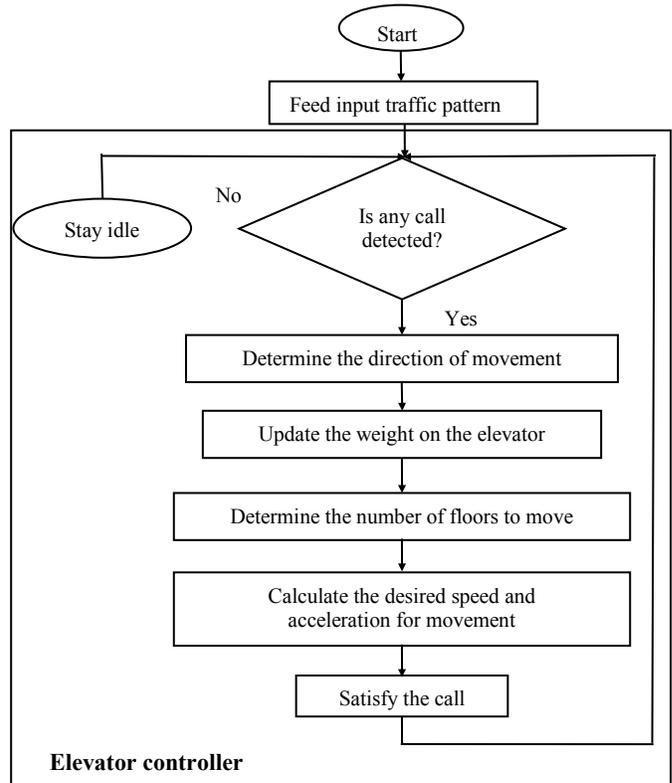


Figure 3. Flow Diagram Elaborating the Algorithm of the Modified Version

Table 1. Simulation Results for 10-Floor Peak-Hour Traffic

	Constant Speed	Speed Variation I	Speed Variation II
Consumption (KJ)	3.45×10^3	3.65×10^3	3.27×10^3
Losses (KJ)	388.83	360.98	394.22
Regeneration (KJ)	-2.88×10^3	-2.92×10^3	-2.81×10^3
Equivalent energy (kJ)	570.1	732.1	465
Total travel time (s)	4.45×10^3	4.21×10^3	4.39×10^3
Efficiency (%)	88.72	90.10	87.96

Case 2: Consider the same elevator with a maximum load of 400 kg, running in a 10-level building during non-peak-hour traffic. The three algorithms were simulated. Table 2 shows the simulation results of all the algorithms over the same traffic intensity (non-peak-hour).

Table 2. Simulation Results for 10-Floor Non-Peak-Hour Traffic

	Constant Speed	Speed Variation I	Speed Variation II
Consumption (KJ)	4.19×10^3	3.96×10^3	3.68×10^3
Losses (KJ)	400.62	337.83	365.00
Regeneration (KJ)	-3.21×10^3	-3.79×10^3	-3.17×10^3
Equivalent energy (kJ)	0.99×10^3	0.18×10^3	0.50×10^3
Total travel time (s)	5.27×10^3	4.88×10^3	4.9761×10^3
Efficiency (%)	90.44	91.46	90.04

For the *speed variation II* algorithm, both energy consumption and travel time were reduced significantly, when compared to the constant-speed type. Hence, energy savings are achievable at a reduced travel time for non-peak hour traffic pattern using *speed variation II*.

Case 3: Consider an additional case of an elevator with a maximum load 400 kg, running in a 20-level building using a non-peak-hour traffic pattern. Table 3 shows the simulation results of all the algorithms over the same traffic intensity (non-peak-hour).

Table 3. Simulation Results for 20-Floor Non-Peak-Hour Traffic

	Constant Speed	Speed Variation I	Speed Variation II
Consumption (KJ)	6.01×10^3	5.99×10^3	5.27×10^3
Losses (KJ)	552.17	418.91	405.86
Regeneration (KJ)	-4.73×10^3	-4.84×10^3	-4.71×10^3
Equivalent energy (kJ)	1.28×10^3	1.14×10^3	0.56×10^3
Total travel time (s)	7.29×10^3	5.86×10^3	5.96×10^3
Efficiency (%)	90.78	93.00	91.90

From Table 3, it is clear that the *speed variation II* algorithm consumed less energy than the other three for a fixed traffic pattern, irrespective of the number of floors in the building. Similar results were obtained from the simulation of a 5-floor building. Thus, from the analyses, the *speed variation II* algorithm produced significant energy savings in both peak and non-peak hours of traffic intensity. Additionally the travel time of the elevator was reduced, thereby enhancing its overall performance.

Miniature Model Implementation

The model system was built for a 5-floor building. It was made up of plywood with open front and back. The motor was mounted on top of the ceiling and the motor shaft was attached to a gear. A chain rotated over this gear as well as two other gears. The chain held the elevator cart on one end and the counter weight on the other. With the rotation of the motor shaft, the chain enabled the cart and the counter-weight to move linearly along the vertical axis. Figure 4 shows how the motor was driven by a motor driver, which was also mounted on the ceiling next to the motor. The three-phase induction motor used for this project was driven by driver hardware. The driver was capable of changing the speed of the motor and reading the voltage and current of the motor. The driver could accept commands for motor

speed from a host via a serial communication interface. The driver could also send the voltage and current readings of the motor to the host over the same interface. The host was a microcontroller-based system that ran embedded software specifically written for the purpose of running the elevator in single-speed or variable-speed modes of operation.



Figure 4. The Miniature Model Elevator

The embedded software written for the microcontroller was intended to run the elevator motor at a desired speed. The software was required to read the motor current and voltage and convert the data into appropriate units in order to estimate the energy consumption. The software received its input commands from switches that told it to choose either a single-speed or a variable-speed system. For a variable-speed system, the weight applied to the elevator cart was read by the software. Based on this information, the software chose the appropriate speed for the elevator motor (as well as the corresponding frequency for the motor driver) from a look-up table. The software then sent the required command frame to the motor driver. Once the motor started moving, the driver collected motor voltage and current samples as well as sent the samples to the microcontroller software. When the elevator completed a predetermined motion, the microcontroller software uploaded the data to a PC running LabView. The data were stored in a spreadsheet for calculating energy consumption of the elevator.

Analysis and Experimental Results from the Miniature Model

It was decided to collect the motor voltage and current every 100 ms during elevator operation in order to calculate energy consumption. The motor was a three-phase induction

motor [16] connected in a Y (wye) configuration. The motor driver manual enlisted the following: phase voltage = V_p (V); phase current = I_p (mA); time difference (sample interval) = t (0.1s, constant); and, motor power factor = PF (0.55 constant). As such, the formulas to be used for exact energy calculation are given here as Equations (8)-(10):

$$\text{Line voltage } V_L \text{ (V)} = 1.73 \cdot V_p \quad (8)$$

$$\text{Line current } I_L \text{ (A)} = I_p / 1000 \quad (9)$$

$$\begin{aligned} \text{Power } P \text{ (W)} &= 1.73 \cdot V_L \cdot I_L \cdot \text{PF} \\ &= 1.73 \cdot 1.73 \cdot V_p \cdot I_p \cdot \text{PF} \\ &= 3 \cdot V_p \cdot I_p \cdot \text{PF} \end{aligned} \quad (10)$$

where, V_L is line voltage; I_L is line current; and, P is power.

In order to convert W to kW and s to hr, use energy (kW · hr) = power / (1000) · (t/60²). Total energy per trip equals the sum of all energy measurements made during the trip. The overall energy equals the sum of the total energy per trip during a complete traffic pattern. And, as PF was constant, $V_p \cdot I_p$ gave an approximation of the energy consumed, which was assumed to be the case here. Energy data were collected automatically by the control hardware and the software. The elevator was run in both single-speed and variable-speed modes with different weights on the elevator cart. For each trip (start to stop), voltage and current samples were collected by the motor driver. Sampling was carried out at 100-ms intervals. The trips were designated as T_{xy} , where x represented the starting floor and y represented the ending floor. The data table, Table 4, was created by LabView from the collected data items for each trip.

Table 4. Data Items Collected from the Model

Speed system	Weight (lbs)	Trip	Number of voltage samples	Number of current samples
Single	10	T_{12}	120	120
Variable	30	T_{31}	80	80

From these tables, the averages of the voltage and current samples for each trip were calculated with a particular weight on the elevator cart. The average values for voltage and current were multiplied to obtain the power. When trips T_{12} (representing floor 1 to 2) and T_{21} (representing floor 2 to 1) were combined, round trip T_{121} was obtained. This was called a one-level round trip. Other one-level round trips were: T_{232} , T_{343} , and T_{454} . Table 5 shows the power (in milliwatts) consumed for each of these one-level round trips.

Table 5. Average Power Consumed (milliwatts) in Each One-level Round Trip with Zero Pounds on the Cart

Speed System	Trips				
	T ₁₂₁	T ₂₃₂	T ₃₄₃	T ₄₅₄	Average
Single Speed	23.27	23.23	21.44	23.26	22.80
Variable Speed	9.89	9.80	8.75	8.63	9.26

Table 5 reveals a power, P, reduction of 59.36% in favor of the variable-speed mode (for zero pounds on cart). Such tables were created for all possible weights, ranging from 0 pounds to 80 pounds. By multiplying the average power by the trip time, the energy ($E = P \cdot t$) consumed for that particular trip (in Joules) was obtained. This is shown in Table 6 for 0 pounds. Table 6 also shows an energy reduction of 18.73% in favor of the variable-speed mode.

Table 6. Average Trip-Energy (Joules) for a One-Level Round Trip with Zero Pounds on the Cart

Applied Weight	Speed System	One-level round trips
		Average
0 Pounds on Cart	Single	54.72
	Variable	44.47

Similarly, additional tables were created with the energy data of two different speed systems for all possible loads. From such tables, and for all possible loads on the elevator cart, an energy savings chart was created. Figure 5 shows that (where the x-axis represents the load), a positive value in energy savings represents less energy consumption under the variable-speed algorithm, whereas a negative value represents more energy consumption using the same algorithm. It can be concluded that significant energy savings is possible for 2-level round trips, when the load on the elevator cart is 0, 10, 20, 70, or 80 pounds. Thus, it appears that, for weights larger or smaller than the counterweight, energy savings is significant. However, for weights that are closer to the counterweight (out of balance load was almost zero), there is little or no energy savings. For these loads, the speed chosen was high and the energy consumed was significant under variable-speed operation.

Figures 6-8 present additional energy-savings data for one-, two-, and three-level round trips, respectively, when the load on the elevator cart was 5, 15, 25, 35, 45, 55, 65, and 75 pounds (where the x-axis represents the load). Observation of these figures clearly reveals more instances of energy savings when the elevator is operated under the variable-speed algorithm.

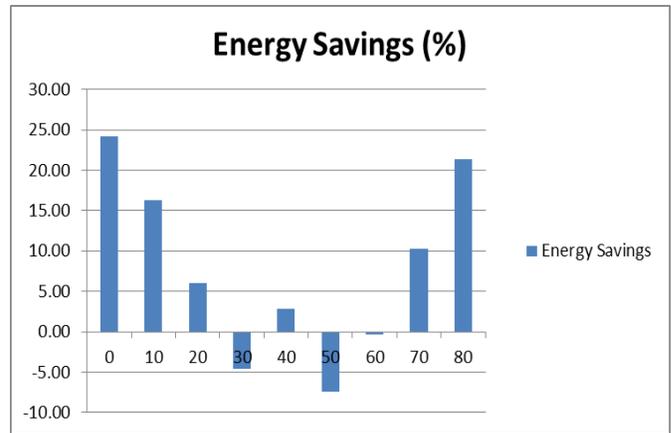


Figure 5. Energy Savings (%) for the Variable-Speed System for Two-Level Round Trips (even weights)

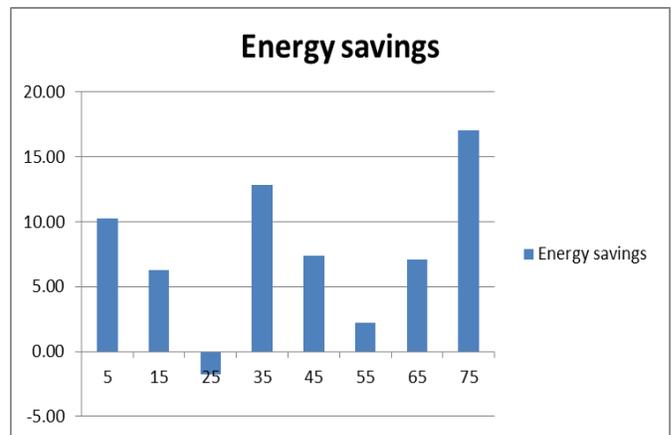


Figure 6. Energy Savings (%) for the Variable-Speed System for One-Level Round Trips (odd weights)

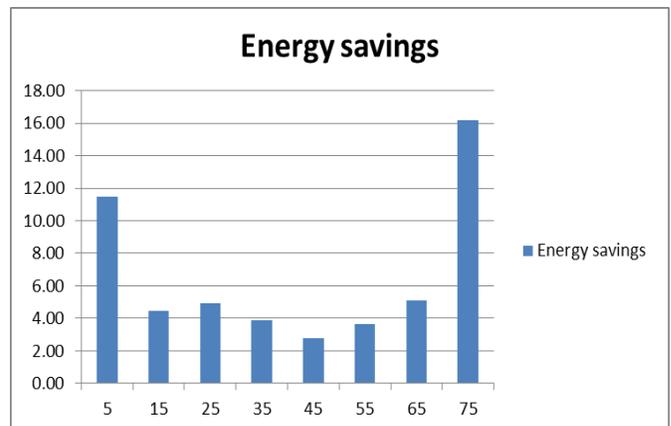


Figure 7. Energy Savings (%) for the Variable-Speed System for Two-Level Round Trips (odd weights)

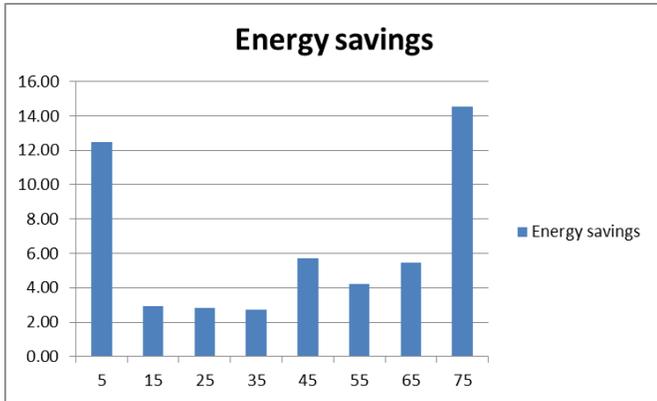


Figure 8. Energy Savings (%) for the Variable-Speed System for Three-Level Round Trips (odd weights)

Synthesized Random-Traffic Energy Data Analysis

The authors created a few random traffic patterns with different sequences of round trips and weights. Energy data were obtained for each single round trip and the added to find the total energy for that pattern. Tables 7 and 8 show two such patterns [15, 17] and the associated energy (in Joules) for the two algorithms (single speed and variable speed). For the first traffic pattern, about 18.48% of the energy was saved under the variable-speed algorithm. However, in the second traffic pattern, there was a loss of energy (4.95%) under the variable-speed algorithm.

Table 7. Energy Comparison of Single Speed and Variable Speed under Random Traffic Pattern 1

Trips	Applied Weight	Average for SS	Average for VS
2 level trip	10 pounds	1130.64	946.31
4 level trip	0 pounds	2319.90	1855.93
3 level trip	80 pounds	1733.24	1413.59
1 level trip	60 pounds	54.12	54.31
	TOTAL	5237.90	4270.15
	Difference (%)		18.48

The authors then created 10 such random traffic patterns. Experiments with these 10 random patterns showed energy savings in nine of the 10 under the variable-speed operation, as shown in Figure 9 (where, the x-axis represents traffic patterns). A positive value represents less energy consumption under the variable-speed algorithm, whereas a negative value represents more energy consumption under the same algorithm.

Table 8. Energy Comparison of Single Speed and Variable Speed under Random Traffic Pattern 2

Trips	Applied Weight	Average for SS	Average for VS
2 level trip	70 pounds	1105.24	991.73
3 level trip	0 pounds	1211.96	1407.13
4 level trip	50 pounds	2240.65	2390.01
1 level trip	20 pounds	54.93	52.13
	TOTAL	4612.78	4841.00
	Difference (%)		-4.95

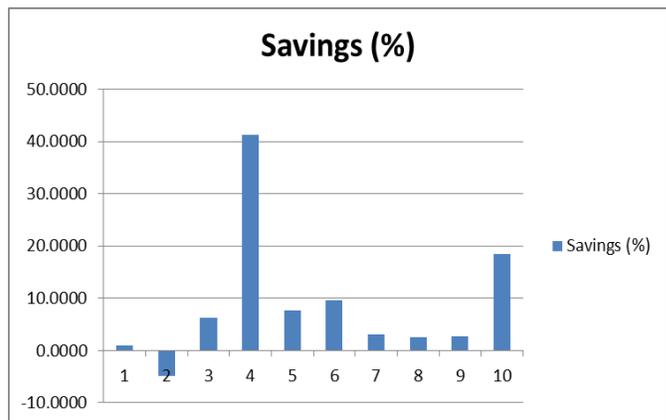


Figure 9. Energy Savings (%) for the Variable-Speed System under Various Random Traffic Patterns

As such, it can be concluded that (in general) the variable-speed operation provides energy savings in such random traffic patterns. The average of these energy savings equaled 8.76%. These results support earlier findings of energy savings with pre-determined traffic (round trips) under the variable-speed mode of elevator operation. Other studies [18-20] were referenced during subsequent analyses. The average energy savings in variable speed for various round trips with different loads was also computed. These round trips were considered to be pre-determined traffic. Table 9 shows the energy savings in percentage for various round trips. The average of these energy savings was 3.72%.

Table 9. Average Energy Savings (%) for the Variable-Speed System for Different Round Trips

1-Level Round trip	2-Level Round trip	3-Level Round trip	4-Level Round trip
7.11	3.71	2.48	1.56

When one compares the data of Figure 9 and the data of Table 9, it appears that the energy savings is higher, in general, for random traffic patterns (between 2% to 40% in Figure 9) than for pre-determined traffic patterns (between 1.5% and 7% in Table 8). This indicates that the variable-speed operation, even with random traffic, is expected to conserve energy. In an earlier study [15], it was shown by simulation that the variable-speed operation of an elevator yielded 5.06% energy savings under peak-hour traffic. The experimental results of this current study from the miniature elevator model support that conclusion. In fact, the average energy savings under random traffic was 8.76% (from Figure 9) in variable-speed operation of the miniature elevator model. This looks promising and should motivate us to undertake further research in this direction.

Comparison of Simulation and Implementation Results

In general, both the simulation and experimental results showed that energy savings are possible in elevator operation under the variable-speed algorithm. The amount of energy savings obtained in simulation was slightly different from the amount of energy savings obtained in the experiment with the miniature model. One reason for this difference may have been the use of a different type of energy form. In simulation, the mechanical energy of the elevator system was considered, whereas in the actual experiment electrical energy was measured. Also, the experimental model was for a 5-floor building, whereas the simulation model was for a 10-floor building. In addition, the actual height of the elevator shaft and the actual weight carried were scaled down in the miniature model with a ratio of 25:1. All of these factors may have led to some differences between simulation and the experiment. Nonetheless, the percentage of electrical energy savings measured in the experiment was higher than the percentage of mechanical energy savings obtained in simulation.

Conclusions

Energy conservation in elevators has gained much importance recently [18-20]. In this paper, the authors presented an algorithm for elevator operation and its verification in order to conserve energy. The focus was to prove earlier simulation results by means of experiments. As such, a miniature elevator model was built for running the elevator in both single-speed and variable-speed modes. Energy data were collected from the model for various traffic patterns (both pre-determined and random). Results showed that the average energy savings under pre-determined traffic was 3.72% and under random traffic was 8.76% in the variable-

speed mode (as compared to the single-speed mode). This supports the simulation results published earlier [14]. Possible future work would be to code peak-hour traffic into the elevator controller and run this traffic under the two different modes of operation. The energy data collected from the model under such traffic would be analyzed and compared to verify the effectiveness of the proposed mode of elevator operation during peak traffic.

References

- [1] Elevators Employed in Several Countries. (n.d.). Retrieved from http://tian.terc.edu/TIAN_statestandardsAZ.html
- [2] Almeida, A. T., Patrão, C., Fong, J., Araújo, R., & Nunes, U. (2010, March). Energy Efficient Elevators and Escalators. Report prepared for ISR-University of Coimbra, Portugal.
- [3] Tominaga, S. S., Araki, I., Ikejima, H., Kusuma, H., & Kobayashi, M. (2002). Development of Energy-saving Elevator using Regenerated Power Storage System. *Proceedings of the IEEE Power Conversion Conference*, (pp. 890–895).
- [4] Hakala, H., Siikonen, M., Tyni, T., & Ylinen, J. (2012). Energy Efficient Elevators for tall buildings. Retrieved from http://www.kone.com/countries/SiteCollectionDocuments/MP/2001_energy_efficient_elevators_tall.pdf
- [5] Zhang, T., Mabu, S., Yu, L., Zhou, J., Zhang, X., & Hirasawa, K. (2009). Energy saving Elevator Supervisory Group Control System with Idle Cage Assignment using Genetic Network Programming. *Proceedings of the ICCAS-SICE International Joint Conference*, (pp. 939–944).
- [6] Brand, M., & Nikovski, D. (2004). Optimal Parking in Group Elevator Control. Retrieved from <http://www.merl.com>.
- [7] Dagdelen, U., Bagis, A., & Karaboga, D. (2005). An Efficient Control Method for Elevator Group Control System. *Proceedings of the 2005 international conference on Computational Intelligence and Security* (pp. 1042-1047), Springer-Verlag Berlin Heidelberg.
- [8] Jing-Long, Z., Jie, T., Qun, Z., & Jun-fang, L. (2010). Energy saving Scheduling Strategy for Elevator Group Control System Based on Ant Colony Optimization. *Proceedings of the IEEE YC-ICT*, (pp. 37–40).
- [9] Hu, Z., Liu, Y., Su, Q., & Huo, J. (2010). A Multi-Objective Genetic Algorithm Designed for Energy saving of the Elevator System with Complete Information. *Proceedings of the IEEE International Energy Conference*, (pp. 126–130).

-
- [10] Hasan, M. Z., Fink, R., Raj, M., & Baskaran, M. (2012). Assessment and Improvement of Intelligent Controllers for Elevator Energy Efficiency. *Proceedings of the IEEE International Conference on Electro/Information Technology (EIT)*, (pp. 1–7), Indianapolis, Indiana.
- [11] Hasan, M. Z., Fink, R., Raj, M., & Baskaran, M. (2016). Evaluation of Intelligent Controllers for Improving Elevator Energy Efficiency. *International Journal for Engineering Research and Innovation*, 7 (2), 5-12.
- [12] Hasan, M. Z., Fink, R., Raj, M., Baskaran, M., James, D., & Gamboa, J. (2015). Performance Evaluation of Energy Efficient Intelligent Elevator Controllers. *Proceedings of the IEEE International Conference on Electro/Information Technology*, (pp. 44–49), Naperville, Illinois.
- [13] Matlab/Simulink in Drives and Power Electronics. (n.d.). Retrieved from http://www.lgep.supelec.fr/uploads/images/lgep/Mocosem/Cocodi/Pages_personel/gremy/Hebron2009/Practice2_Intro_Simulink.pdf
- [14] Cortés, P., Larrañeta, J., & Onieva, L. (2004). Genetic Algorithm for Controllers in Elevator Groups: Analysis and Simulation during Lunch-peak Traffic. *Applied Soft Computing*, 4, 159-174.
- [15] Barney, G. C. (2003). *Elevator Traffic Handbook: Theory and Practice*. Spon Press, London-New York.
- [16] Reliance Induction Motors. (n.d.). Retrieved from http://www.reliance.com/prodserv/motgen/b7097_2.htm
- [17] Barney, G. C. (2012). Towards Low Carbon Lifts. Retrieved from, <http://www.cibseliftsgroup.org/docs/Towards%20Low%20Carbon%20Lifts.pdf>
- [18] Kiker, P. (2015, June). Study Finds Global Opportunities for Greater Elevator Efficiency. Report prepared for American Council for Energy Efficient Economy (ACEEE).
- [19] Lee, S., & Bahn, H. (2005). An Energy-aware Elevator Group Control System. *Proceedings of the 3rd IEEE International Conference on Industrial Informatics (INDIN)*, (pp. 639–643).
- [20] Sachs, H., Kwatra, S., & Misuriello, H. (2015, January). Advancing Elevator Energy Efficiency. Research Report A1501 prepared for American Council for Energy Efficient Economy (ACEEE).

MUHAMMAD ZAFRUL HASAN received his BSc in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology. He received his Master of Electronic Engineering from Eindhoven University of Technology (the Netherlands) under a Philips post-graduate scholarship program. He subsequently held several faculty positions in an engineering college and in a universi-

ty in Malaysia. He obtained his PhD in Computer Engineering from the New Jersey Institute of Technology. He was awarded the NJIT Hashimoto Fellowship. He taught at Texas A&M University in College Station. He is currently an assistant professor in the School of Engineering at Texas A&M International University. His research interests include the design, implementation, and testing of embedded systems for energy conservation, performance evaluation of computer architectures, and behavioral synthesis and testing of digital systems. Dr. Hasan may be reached at muhammad.hasan@tamiu.edu

RAINER J. FINK received his PhD in Biomedical Engineering in 1995, MS in 1992, and BS in 1988, all from Texas A&M University. He is currently an associate professor in the Department of Engineering Technology and Industrial Distribution, Texas A&M University. He holds several patents in bio-medical devices. His semiconductor test educational efforts have been continuously supported by Texas Instruments (Dallas, TX), Teradyne Inc. (Boston, MA), National Instruments (Austin, TX), National Semiconductor (Santa Clara, CA), and IBM (Austin, TX). His research interests include analog electronics, semiconductor testing, medical device design and testing, product development, and entrepreneurship. Dr. Fink may be reached at fink@tamu.edu

DARYL JAMES received his BS degree in Electronic Systems Engineering Technology in 2013 from Texas A&M University. Mr. James may be reached at daryl.james2010@gmail.com

JESUS GAMBOA received his BS degree in Systems Engineering in 2014 from Texas A&M International University. Mr. Gamboa may be reached at jesus.gamboa@dusty.tamiu.edu

MATLAB/SIMULINK SIMULATION OF ARC FAULTS

Amir Khan, California State University Fresno; Nagy Bengiamin, California State University, Fresno

Abstract

Many models have been developed to characterize electric arcs in the computer simulation software package, Electromagnetic Transients Program—Alternative Transients Program (EMTP/ATP). EMTP/ATP software has traditionally been the preferred simulation tool for electric power applications. However, Matlab/Simulink has been gaining popularity among researchers and engineers for its versatility and ease of use. In this paper, the software tools (Simulink and EMTP/ATP) are compared for their abilities and viability to simulate arc models. The Cassie arc model was simulated in two circuit schematics using identical settings in both software tools. The arc voltages produced by both tools were compared in the time and frequency domains. It was shown that Simulink provides the needed computational capability to model arcs accurately; as compared to the results of EMTP/ATP.

Introduction

Power system protection schemes have traditionally focused on bolted fault-current protection, such as protection against phase-to-phase faults or line-to-ground faults. Recently, fault research has shifted to studying arc faults. These faults also occur from phase-to-phase or line-to-ground, but the current conducts through air. The conductance through air causes several complexities that are of particular interest to researchers. Because of the impedance of the arc, these currents are often smaller than bolted faults. As a result, their currents sometimes are not immediately detected by protection devices. Furthermore, if the arc experiences a high enough current, it can produce a flash of high-temperature plasma and pressure, otherwise known as an arc flash.

The voltages and currents measured from arcs are often multifaceted and complex. Many black-box models have been created to characterize arcs. The digital simulation of these arc models gives researchers a way to study arc characteristics without creating an actual high-energy arc experimental setup. The long-established program used to simulate arcs is EMTP/ATP, a program often used to study transients within power systems. A popular preprocessor of EMTP is a program named ATPDraw. This program integrates a visual circuit designer and plotting tools with the EMTP program.

In recent years, Simulink has expanded its library to include actual circuit components for simulation. As a result, it has been used more widely as an alternative to EMTP in simulating arc faults [1]. In this study, the authors attempted to compare the simulations from both programs in order to investigate whether Simulink is a suitable alternative in the simulation of arc faults, which can compare two situations involving an arc using identical settings in Simulink and ATPDraw. Their characteristics were then compared in time and frequency. Lastly, any software limitations from either program were evaluated.

Simulation of the Cassie Model in a Simple Circuit

It is preferable to represent arcs as electrical elements in a circuit [2]. Equation (1) shows how the Cassie model defines arc conductance [3]:

$$\frac{dG}{dt} \frac{1}{G} = \frac{d(\ln(G))}{dt} = \left(\frac{1}{\tau}\right) \left(\frac{u^2}{u_c^2} - 1\right) \quad (1)$$

where, G is the arc conductance; τ is the time constant associated with the arc; u is the arc voltage; and, u_c is the average arc voltage.

The model assumes that the power loss of the arc is constant and due to “deformation processes” [3, 4]. It also assumes that the arc voltage is constant and best used to represent the arc during most of the current-carrying period [3]. This model was first implemented in series with one resistor and a power supply, uS . Both programs were set to a fixed-step of $1 \mu s$ and sampled for 0.2 seconds. The following parameters were used in both programs:

$$uS = 12kV, R = 200\Omega, u_c = 3.850KV, \tau = 5 \times 10^{-5} s$$

Software Implementation

ATPDraw opens in an interface in which circuit components can be put together visually, similar to circuit simulation programs like PSIM and Multisim. The arc can be represented as either a time-varying resistance, $R(TAC)$ Type 91, or a time-varying voltage source ($TACS$ Source). That time-varying component is then defined using a `MODEL` block. This model defines the time-varying component through each iteration, and it can read voltages or currents from the circuit as inputs. The actual language and syntax

used to code these models are defined in the ATP Rule Book. Figure 1 shows the schematic used in this study. Figure 2 shows the content of the MODELS block. Part of the code was influenced by the work of Kizilcay and Hoidalen [5]. A time-varying resistor was used to define the arc.

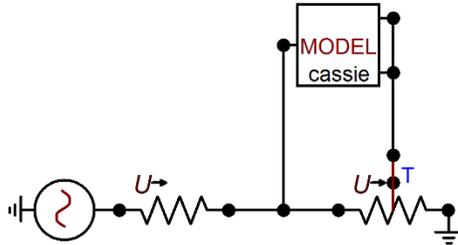


Figure 1. The Cassie Model Implemented in ATPDraw Using a Time-Varying Resistor

```

MODEL CASSIE
DATA tau,uc
INPUT v2
OUTPUT o1,g
VAR o1,g,G1,i1
HISTORY i1 {dflt: 0}
G1 {dflt: 0}
g {dflt: 10000}
INIT
G1:=0
ENDINIT
EXEC
if t=0 then
  g:=10000
else
  COMBINE ITERATE AS group1
  i1:=v2*g
  G1 := (i1*v2) / (uc**2)
  LAPLACE (g/G1) :=1.0|/(1.0|+tau|S)
  ENDCOMBINE
endif
o1 := recip(g)
ENDEXEC
ENDMODEL

```

Figure 2. Contents of the MODEL Block that Define the Time-Varying Resistance According to the Cassie Model

The Cassie model was implemented in Simulink using the same parameters. The overall circuit was implemented using only mathematical blocks; no components from SimPowerSystems were used in this case. Figure 3 shows the schematic for this block diagram.

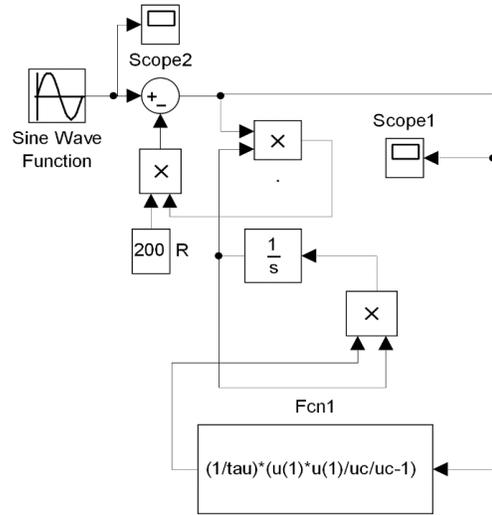


Figure 3. Cassie model implemented in Simulink

Results and Comparisons in the Time Domain

Figures 4 and 5 show the results from both programs. Figure 6 shows the two voltages from both programs on the same plot.

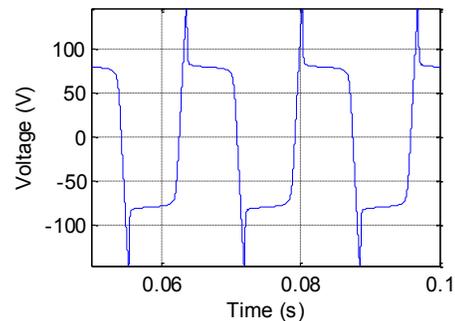


Figure 4. Simulink-Generated Arc Voltage for the Cassie Arc Model

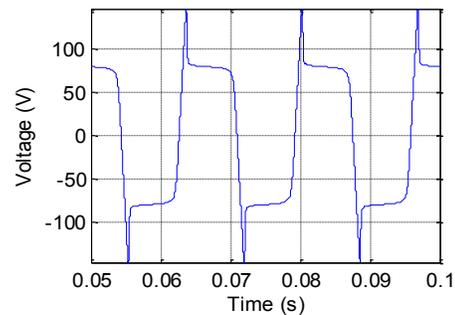


Figure 5. ATPDraw-Generated Arc Voltage for the Cassie Arc Model

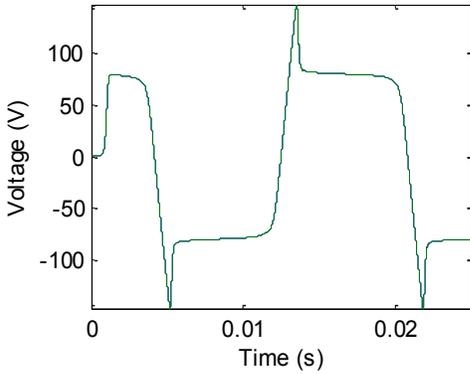


Figure 6. Two Overlapping Signals

The two signals in the time domain were compared within Matlab by calculating their cross-correlation using the `xcorr` command. This command takes two matrices as inputs and continually shifts the second matrix across the first. For each time shift, the two signals are multiplied and the sum of those array elements is calculated. The highest value of this array indicates the point in time in which the two signals are most similar.

However, the shifting in time is not of importance, because the signal from Simulink and the signal from ATPDraw will already be in-phase. Of greater interest is the actual maximum cross-correlation value, which would be the two signals multiplied together and then all array elements summed together, when the two signals overlap. This value can be compared to the individual energies of both signals to determine signal similarities. If both signals are identical, this maximum cross-correlation value should be equal to the individual energies of the two signals.

Since the signals had identical time steps, a new array was created that calculated the difference between the two signals for every array element. From this array, the maximum difference between any two individual array elements was identified, and the total number of array values that had a difference of 5% or less was calculated as a percentage of the total number of array elements. Table 1 presents those findings.

The maximum of the cross-correlation indicated that the two signals were nearly identical, since its value was equal to energies of each individual signal. Furthermore, the maximum difference between any two points was very small relative to the maximum and minimum values that the signals reach. Most notably, no array value ever exceeded a difference greater than 5%. The time domain information from Table 1 indicates that the signals were not significantly different.

Results and Comparisons in the Frequency Domain

The frequency spectra were plotted using the FFT command in Matlab. Figures 7 and 8 show the spectra for the arc voltages from the two programs. Similar comparisons were again made in Table 2. Since the signals were virtually identical in the time domain, it was predicted that they would be similarly identical in the frequency domain. Table 2 shows that to be true. The percent difference in the maximum values between the two in the frequency domain was 0.00269%. The maximum difference between any two array values was 0.0050, and the maximum of the cross-correlation, when compared to the energies of each individual signal, indicated that they were virtually identical. Only 0.04% of the frequency array values differed by more than 5%.

Table 1. Comparison of Two Signals in the Frequency Domain

	ATPDraw Voltage	Simulink Voltage
Minimum value	- 147.2625 V	- 147.1222 V
Maximum value	147.2625 V	147.2222 V
Energy	1.2733×10^9 J	1.2733×10^9 J
Maximum of cross-correlation	1.2733×10^9 J	
Maximum difference between any two array elements	0.3346 V	
Percentage of array elements in which signals differ by 5% or less	100.00 %	

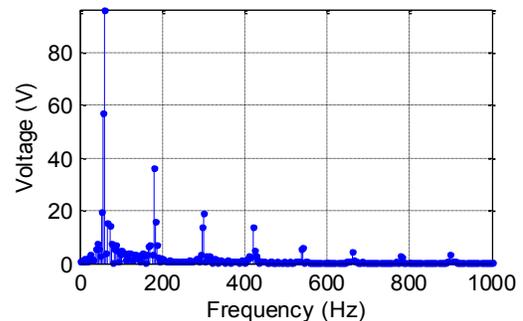


Figure 7. Frequency Spectrum for Arc Voltage Produced by ATPDraw

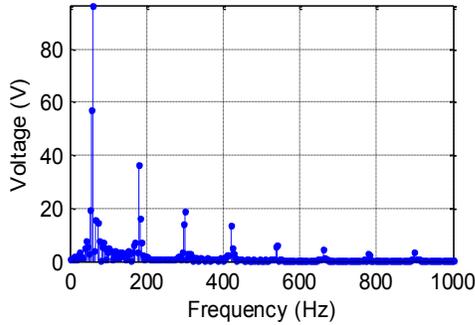


Figure 8. Frequency Spectrum for Arc Voltage Produced by Simulink

Table 2. Comparison of Two Signals in the Time Domain

	ATPDraw Voltage	Simulink Voltage
Maximum value	96.4789 V	96.4815 V
Energy	8.3445×10^3 J	8.3445×10^3 J
Maximum of cross-correlation	8.3445×10^3 J	
Maximum difference between any two array elements	0.0050 V	
Percentage of array elements in which signals differ by 5% or less	99.96 %	

Comparisons in the time and frequency domains demonstrated that this simpler schematic with a voltage source, resistor, and the Cassie arc model produced nearly identical results. Subsequently, a more elaborate circuit schematic was developed using the same arc model, and the performance of the arc simulations in both programs was compared.

Simulation of the Cassie Model and Time Domain Analysis

The same Cassie arc model was placed inside of another circuit that was more reflective of an actual transmission line. Figure 9 portrays that schematic. This circuit was replicated from the block diagrams previously produced by Schavemaker and Van der Sluis [6]. Figure 9 defines the Cassie model. The content of the MODELS block is identical to the code used in Figure 2. A time-varying resistor was used to define the arc. The system was simulated and then exported to Matlab for further analysis.

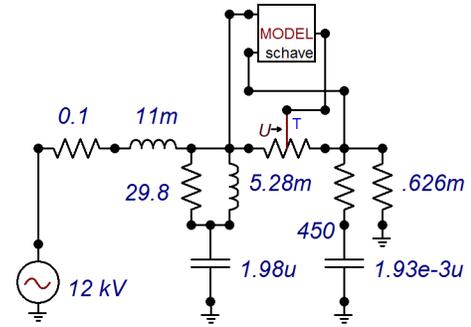


Figure 9. New Schematic for the Cassie Model in ATPDraw

The model was created in Simulink using the same parameters. The circuit was implemented using SimPowerSystems, a toolbox that allows the simulation of actual electrical components. Figures 10 and 11 show the Simulink schematic for this block diagram. Figures 12 and 13 show the arc voltage results from both programs. Figure 14 shows the results from both programs on the same plot for the first few cycles. The blue signal is ATPDraw and the green is Simulink. The two signals appear to be very similar. The same characteristics from Table 1 were calculated for these two signals. Table 3 shows the results. The maximum of the cross-correlation indicated that the two signals were extremely similar. Its value differed from the ATPDraw signal's energy by 0.020%, and from the Simulink signal's energy by 0.017%. The energies of the ATPDraw signal and the Simulink signal differed by only 0.0034%. Only 4.73% of the array indexes differed by more than 5%. The maximum difference calculated between array elements was a large value, but it was misleading in comparing the two signals.

This large value was a result of an apparent time delay between the two signals that was only apparent at the first falling edge and first rising edge of the arc voltage. The reason for this time shift was not identified; however, since this was not evident in the previous schematic, it was likely associated with how each individual program processes inductors and capacitors, not the arc modeling. The voltage sources of each program did not contain this characteristic when overlapped with one another. Figures 15 and 16 show these. In those figures, the dashed red line is the Simulink signal; the solid green dotted line is the ATPDraw signal; and the solid blue line is the difference between the two.

After about two cycles, the two signals were in-phase, so it is safe to assume that this occurrence does not appear in steady state values after a few cycles have passed; this is evident in Figure 13. However, the width of the time shift is not significant in changing the characteristics of the signal.

Neglecting the first few cycles, the maximum difference between the two signals in each array element then becomes 270.89 V, and this occurs at one of the sharp peaks of the arc voltage. This value is very close to the difference in the maximum values of the two signals, which is about 212.9 V. The value of 270.89 V is about 4.64% of the maximum value. The information from Table 3 and Figure 14 demonstrates that the signals were virtually the same in the time domain.

Table 3. Comparison of Two Signals in the Time Domain

	ATPDraw Voltage	Simulink Voltage
Minimum value	-6.0750×10^3 V	-5.8652×10^3 V
Maximum value	6.0450×10^3 V	5.8321×10^3 V
Energy	2.9488×10^{12} J	2.9487×10^{12} V
Maximum of cross-correlation	2.9482×10^{12}	
Maximum difference between any two array elements	2.2533×10^3 V	
Percentage of array elements in which signals differ by 5% or less	95.272 %	

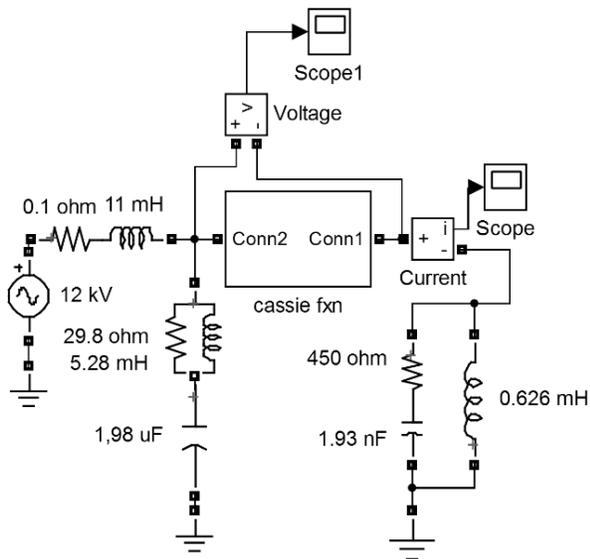


Figure 10. Block Diagram for a New Model in Simulink

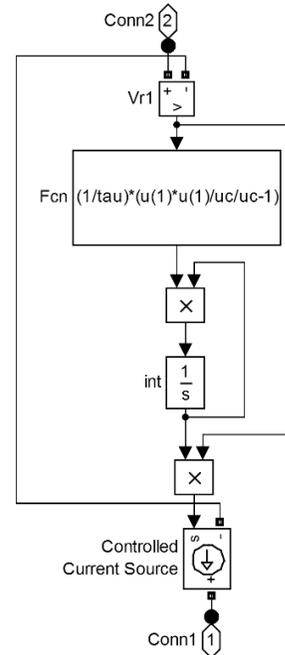


Figure 11. Sub-model from Figure 10 that Defines the Arc as a Controlled Current Source

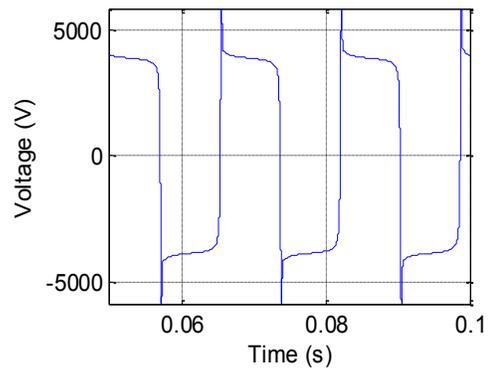


Figure 12. Simulink-Generated Arc Voltage for Schematic 2

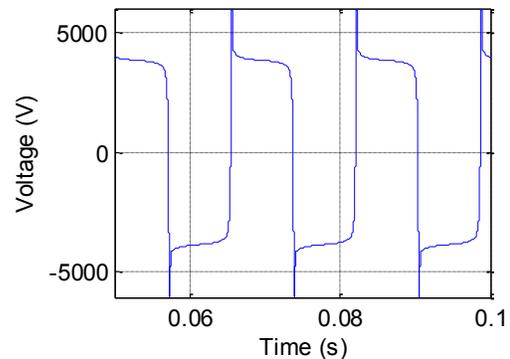


Figure 13. ATPDraw-Generated Arc Voltage for Schematic 2

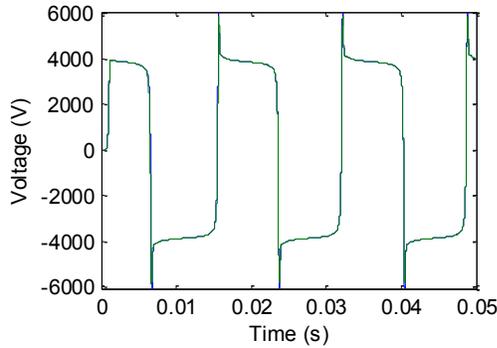


Figure 14. Output Voltage from Both Programs on the Same Plot

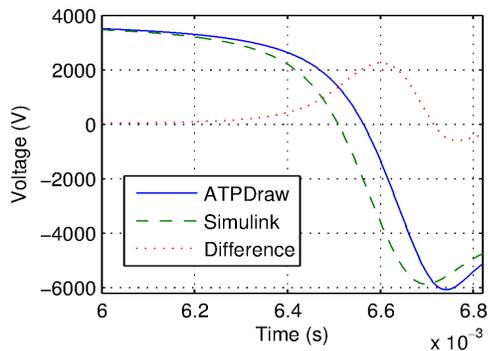


Figure 15. Demonstration of Time Shift on the First Falling Edge

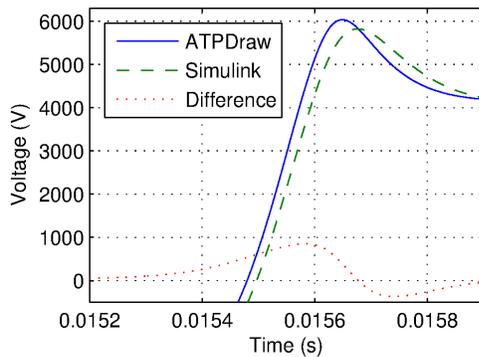


Figure 16. Demonstration of Time Shift on the First Rising Edge

Frequency Domain Analysis

The frequency spectra were plotted using the FFT command in Matlab, similar to what was done for the previous schematic. Figures 17 and 18 show the spectra for the arc voltages from the two programs. Table 4 presents a comparison of the two. The percent difference in the maximum values between the two in the frequency domain was 0.0343%.

The maximum difference between any two array values was 0.1565, and the maximum of the cross-correlation was extremely close to the powers calculated from the frequency domain. Only 1.85% of the total array indexes differed by greater than 5%. All of these features demonstrate that both programs produce indistinguishable frequency spectra for the arc voltage.

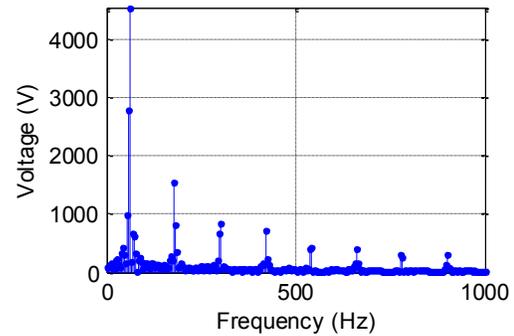


Figure 17. Frequency Spectrum Produced by ATPDraw

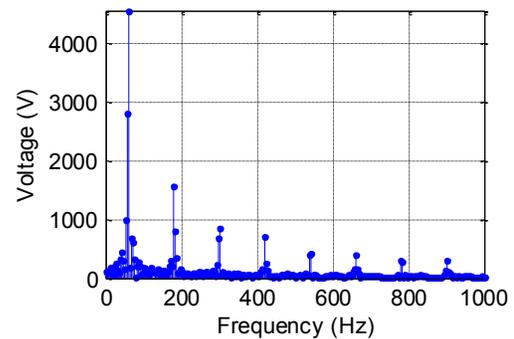


Figure 18. Frequency Spectrum Produced by Simulink

Table 4. Comparison of Two Signals in the Time Domain

	ATPDraw Voltage	Simulink Voltage
Maximum value	$4.5563 \times 10^3 \text{ V}$	$4.5545 \times 10^3 \text{ V}$
Energy	$1.9325 \times 10^7 \text{ J}$	$1.9324 \times 10^7 \text{ J}$
Maximum of cross-correlation	1.9323×10^7	
Maximum difference between any two array elements	5.1275	
Percentage of array elements in which signals differ by 5% or less	98.05 %	

Limitations of Simulink and ATPDraw

Arc models in ATPDraw/EMTP are represented as time-varying resistors or controlled voltage sources. In Simulink, the arc is represented as either a controlled current source for series arcs, or controlled voltage sources for parallel arcs. However, using a controlled current source in Simulink does prevent the use of inductors in series with these current sources. This is due to the inductor having its own current characteristic that is dependent on the voltage across it. ATPDraw does not encounter this problem in modeling an arc. As a result, modifications need to be made in Simulink models to bypass this limitation. Furthermore, EMTP is restricted to a fixed time-step solver, while Simulink has the ability to use a fixed or variable time-step with multiple solvers. This gives Simulink an advantage in debugging solver issues. Although ATP is widely used as a standard for arc model simulations, the ATPDraw software has much more limited support in comparison to Simulink. ATPDraw has a small forum community on its website, and EMTP has a mailing list, where users can ask other licensed users questions. However, responses to queries are not guaranteed. Simulink, on the other hand, has a large community of users and an extensive help service. For software issues, more assistance is available for Simulink than ATPDraw.

Conclusions

Based upon the similarities that ATPDraw and Simulink produced in the arc model simulations in this study, the authors concluded that Simulink is a suitable alternative to EMTP in the study of arc faults. The characteristics of the arc in the time and frequency domain were consistent in both programs, and both could be used to model an arc to a similar precision. In the second circuit schematic that was modeled, a small phase difference was found in the time domain during the first two cycles. However, this was likely due to differences between the two programs in their modeling of capacitors and inductors rather than differences in their abilities to model arcs. Although a small time shift was identified between the two, its duration was negligible, and it did not change the characteristics of the model in the time domain. This time shift was not caused by the actual arc model, but rather the passive components external to it. The frequency domain characteristics for both comparisons were also similar for signals from both programs. Thus, both programs produce results that are similar in both the time and frequency domains. Therefore, either program can be used to simulate and study arcs in the time and frequency domains.

References

- [1] Yu, R., Fu, Z., Wang, Q., Sun, S., Chen, H., Xu, Q., & Chen, X. (2011). Modeling and simulation analysis of single phase arc grounding fault based on MATLAB. *Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology*, vol. 9 (pp. 4607 - 4610). IEEE Xplore (10.1109/EMEIT.2011.602400).
- [2] Browne, T. E. (1958). An Approach to Mathematical Analysis of A-C Arc Extinction in Circuit Breakers. *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems Trans.*, 77(3), 1508–1514.
- [3] Yuan, L., Sun, L., & Wu, H. (2013). Simulation of Fault Arc Using Conventional Arc Models. *Energy and Power Engineering*, 5(4), 833–837.
- [4] Browne, T. E. (1948). A Study of A-C Arc Behavior Near Current Zero by Means of Mathematical Models. *Trans. Am. Inst. Electr. Eng. Transactions of the American Institute of Electrical Engineers*, 67(1), 141–153.
- [5] Kizilcay, M., & Hoidalén, H. K. (2015). EMTP-ATP. In *Numerical Analysis of Power System and Transients and Dynamics* (pp. 47-101). UK: CPI Group.
- [6] Schavemaker, P., & Van der Sluis, L. (2002). The Arc Model Blockset. *Proceedings of the Second IASTED International Conference POWER AND ENERGY SYSTEMS (EuroPES)*, (pp. 644-648). Crete, Greece.

Biographies

AMIR KHAN earned his MS degree in engineering, Electrical Engineering Option, in 2016, from California State University, Fresno. He earned his BS degree from the University of California, Davis. His interests include power and control systems. He may be reached at amir454@mail.fresnostate.edu

NAGY BENGIAMIN is a professor of electrical engineering at California State University, Fresno. He holds an MEngr degree in systems engineering (1976) from Carleton University and PhD in electrical engineering (1979) from The University of Calgary, Canada. Dr. Bengiamin's current interests include power systems, power electronics, and control systems. Dr. Bengiamin is a senior member of IEEE and may be reached at bengiami@csufresno.edu

BIOMIMETIC TRANSFEMORAL KNEE WITH A GEAR MESH LOCKING MECHANISM

Tyagi Ramakrishnan, University of South Florida; Millicent Schlafly, University of South Florida;
Kyle B. Reed, University of South Florida

Abstract

The innovation presented in this paper is a passive knee locking mechanism that incorporates a cross-linked, four-bar mechanism similar to the sagittal plane configuration of the anterior cruciate ligament (ACL) and posterior cruciate ligament (PCL). The flexible four-bar mechanism guides the motion of the knee and aids in the return of the knee from full flexion to extension. The flexible four-bar mechanism also connects the femoral spur gear to the tibia spur gear. The gears are based on a circular radius, derived from femoral condyle dimensions. The gears are connected using a parallel link to keep the femur and tibia from moving away from each other when the knee is in motion. This prosthetic knee design is based on the anatomical dimensions of a human knee, which allows the design to be scaled from a large adult male to a small pediatric knee.

Introduction

Human gait consists of a synchronized and cyclic movement of each leg that helps a person move forward [1]. Walking is a complex and coordinated process that recruits a range of muscles to actuate the motion. This coordination is disrupted by a limb amputation. The knee and ankle joints are a vital part of human locomotion and are responsible for articulation, load bearing, and the general dynamic control of an overall stable gait [2]. Therefore, an amputation that causes a loss of either joint is detrimental to a person's gait [3]. Improving the design of prostheses can greatly increase the quality of life of a person with an amputation by increasing their potential mobility. There are about seven million transfemoral amputees around the world [4], and each amputee is unique and, hence, requires a custom prosthesis. With the advent of 3D printing technology on the rise, it is becoming possible to customize a prosthesis to a specific individual's size and gait pattern [5]. It is also possible to tailor-make a prosthesis that has anatomically similar dimensions to the person, while making the design inexpensive and passive.

Current prosthetic knees are either passive or active mechanisms [6, 7]. Active knee mechanisms are considered state-of-the-art and the designs incorporate complex mechanisms that enable the actuators to mimic human walking [5]. Active knee mechanisms cause more lower metabolic strain

than passive knees in tasks such as walking, stair ascent, traversing slopes, and ambulatory tasks [3, 8, 9]. Active knees use variable control algorithms to adjust for terrain and environmental conditions. However, active knees are expensive, and transfemoral amputees typically use their passive knees more than their active knees [10]. Also, additional training is required to properly fit and fine-tune active knees, which hinders the widespread adoption of active knees [11].

The human knee is categorized as a condylar joint. It can be closely represented by a polycentric mechanism. Polycentric mechanisms are one of the five forms of passive knee mechanisms: manual, single axis, weight activated, polycentric, and knee with exterior hinges [12, 13]. The knees are designed to assist amputees with various control levels given by the K level (K is an arbitrary letter assigned by HCFA) [13, 14]. Low-control amputees of the scale K0-K2 rely on manual locking mechanisms. The manual locking knee relies on user input to lock and unlock the knee joint during gait, thereby giving full control to the user. The widely used polycentric knees are for a user with medium to high control [15]. A four-, five-, or six-bar mechanism [12, 16] can be used for a polycentric knee. These mechanisms shift the instantaneous center throughout the gait cycle to improve locking and unlocking of the knee joint. Other mechanisms, such as a single-axis knee, are used in conjunction with hydraulic or weight-activated locking systems to aid in the control and return to extension [13]. This biomimetic knee design uses a polycentric, cross-linked four-bar mechanism that is designed to mimic the anatomical movement of the human knee. This knee design can be tuned to offer a wide range of control from K0 to K4.

There have been several attempts to recreate the human knee joint in a prosthetic mechanism. Figure 1 (a1-c1) shows the ACL and PCL, which play an integral part of knee kinematics, are modeled as a cross-linked four-bar mechanism [17-20]. Anatomically, the mechanism consists of the femur, tibia, ACL, and PCL. The simplistic representation as a four-bar mechanism excludes the fact that the knee joint is controlled by several muscles to execute refined motion [21]. It only focuses on the sagittal plane kinematics. The knee joint rolls, slides, and rotates as it goes from extension to flexion, which is called knee roll back [22]. This motion is controlled by the ACL and PCL that stabilize the knee at every position. A tear in either ligament

causes detrimental effects to the motion of the knee, which in turn greatly affects a person's gait [23].

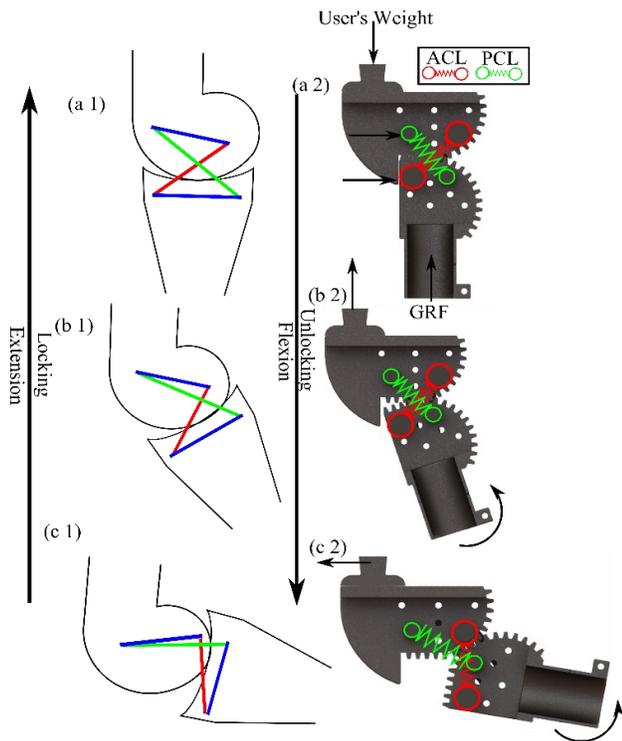


Figure 1. Comparison of Human (a1-c1) and Biomimetic (a2-c2) Knee Motions

There is also a need for prosthetics to be unique to every amputee's condition. Current prosthetic products are designed to be of similar dimensions, due to manufacturing, design, and cost constraints. Fitting prostheses that are not suitable to the user's dimensions causes an array of problems with their gait dynamics, which in turn leads to physical pain. The mechanism detailed in this paper bridges the gap of anatomically similar prostheses and scaling the design with the dimensions of the amputee limb. This customization can be easily achieved with additive manufacturing. The prosthetic design presented here can be made available in a way such that a person with access to a 3D printer can make a viable product, thereby driving down the cost of manufacturing.

Design

Figure 2 shows how the knee design consists of four major components: femur gear, tibia gear, spring holders, and outer linkages. The femur and tibia gears are modified to aid in locking the knee. The circular gear ends with a flat rack on both gears that absorb the load during weight bearing and prevents hyperextension. The radius of the circular gear

can be scaled to fit any amputee's anatomical femoral dimension. The spring holders can be fit with a range of springs that vary in stiffness, depending on the control level of the amputee. There are also other holes provided on the femur and tibia that allow for different configurations of initial spring stiffness to better aid in the dynamics of the knee joint. The outer links provide lateral support and keep the knee mechanism intact. Depending on the configuration of the springs, they can act as returning and stabilizing mechanisms.

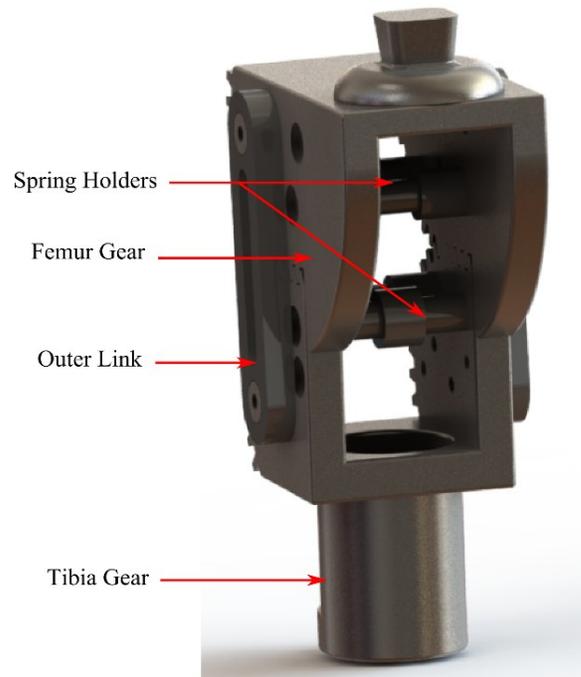


Figure 2. Parts of the Knee Mechanism

The function of the knee is straightforward, and Figure 1 (a1-c2) shows how the kinematics of the mechanism were designed to mimic the human knee. When the user's weight is applied to the top of the femoral spur gear, it locks with the tibia spur gear. The weight of the user is applied towards the anterior of the femoral spur gear in order to prevent buckling of the knee. The knee stays locked throughout the stance phase. At toe-off, the weight of the user shifts to their opposite leg and no weight is applied to the prosthetic knee. This shifting of weight allows the knee to flex and is guided by the four-bar mechanism that behaves like the ACL and PCL in a normal knee. In essence, the flexible links are stretched to a certain length, thereby increasing the force from the springs. Figure 3 (a-f) shows how, at terminal swing phase, just before heel strike, the flexible links in the four-bar mechanism snap back to the extended locked position and lock completely once the user applies weight on the knee.

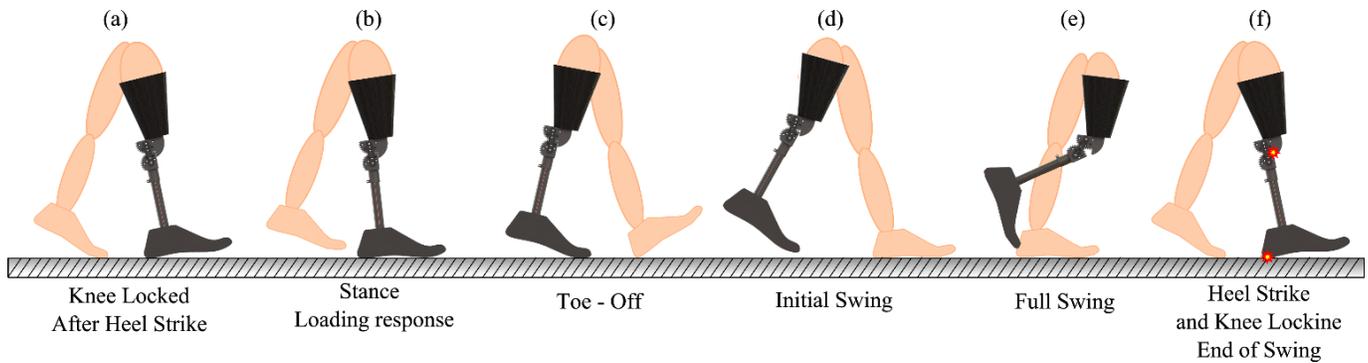


Figure 3. Gait with the Prosthetic Knee

This knee prosthesis has very few parts, and thus is a good candidate for sustainable additive manufacturing, which can offer the highly customized products needed by amputees. The prototypes used for this study were made of 3D-printed ABS and nylon. Functional pediatric knees can be made out of these materials and the child amputee can get a new knee when they outgrow their old. It is also a viable candidate to be produced using metal additive manufacturing processes for adult sizes, when added strength is needed. An MRI image of the amputee's intact femur can be used to obtain the pitch diameter of the gear used in the prosthetic knee mechanism. In case of bilateral amputees, a person of similar proportion can be used for sizing purposes. The pitch diameter of the gear will be twice the average radius of the femoral condyles. Femoral condyles radii generally range from 20–30 mm for adults [24-27]. The pitch radius used for this design was 28.5 mm, which is on the larger side of the condylar radii range, because the kinematics match better.

Table 1 shows the gear design parameters for two different sizes. The femur and tibia spur gear have the same gear pitch diameter. This versatility and biomimetic design make this knee unique and highly customizable. The knee also facilitates modification to add control elements that can benefit amputees with lower control, since it operates using simple spur gears and springs. For example, a high-functioning amputee may desire low stiffness for an instantaneous response from the knee and a lower functioning amputee may require higher stiffness for more control. The specific cross-linked four-bar mechanism used for this design, represented in Figure 4, was designed similar to the anatomical mechanism consisting of the ACL and PCL, as described from the sagittal plane. The mechanism presented here has slightly different dimensions to simplify the design and make it easier to model. The mechanism consists of two pivot points that can be seen as P_1 and P_2 in Figure 4(a). P_1 is the pivoting joint at the Femur and P_2 is the pivoting joint for the Tibia. The blue link represents the shank and is perpendicular to link c-d.

Table 1. Gear Design Parameters for Different Sizes

Gear parameter	Adult	Child (assuming half the size)
Pitch radius (Condylar radius)	28.5 mm	14.25 mm
Teeth pressure angle	14.5°	14.5°
Number of teeth in full gear	25	25
Number of teeth in the rack	4	4
Addendum	1.14	0.57
Diametric pitch	0.877	1.754
Module	1.14	0.57
Velocity ratio	1	1

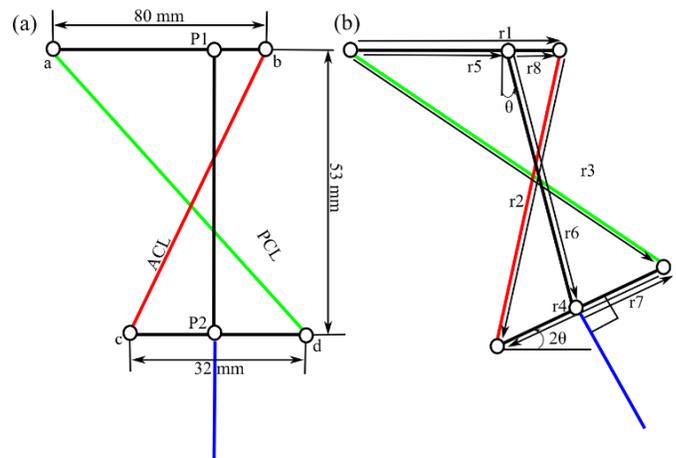


Figure 4. Prosthetic Knee Mechanism (a) Dimensions and (b) Vectors

The analysis of the four-bar mechanism was performed using position vector-based equations. Figure 4(b) shows the vector representation and consists of five vectors, one for each link, and three smaller vectors defining a section of

a link. The constraints applied to this mechanism are that vectors $r_1, r_4, r_5, r_6, r_7,$ and r_8 are of a fixed length, while vectors r_3 and r_2 represent the springs in the design and are allowed to change in length, as shown in Table 2. Another constraint is that r_4 is constrained to move twice the angle of r_6 , which is imposed by the gear and four-bar mechanism configuration.

Table 2. Position Equation Parameters

$r_1 = 80$ mm	$\theta_1 = 0$
$r_4 = 32$ mm	$\theta_4 = 180 + 2\theta$
$r_5 = 60$ mm	$\theta_5 = 0$
$r_8 = 20$ mm	$\theta_8 = 0$
$r_6 = 53$ mm	$\theta_6 = 170 + \theta$
$r_7 = 0.5 * r_4 = 16$	$\theta_7 = 2\theta$

Equations (1)-(10) were solved using Matlab to obtain the positions of the moving links and joints. Some variables were fixed to constrain the size of the design. Solving Equations (1)-(4) or Equations (3)-(6) will give the solutions for the four unknown variables, which are $r_2, r_3, \theta_1,$ and θ_2 . Figure 5 shows the motion of the mechanism with the given inputs for a flexion from 0° to 90° , where P_1 and link a-b are fixed. Figure 6 shows how the ACL and PCL change in length with respect to the corresponding flexion angle. The change of prosthetic ACL and PCL link length was compared to the change of length of the ACL and PCL ligament obtained from anatomical data [28]. Figure 6 shows that this prosthetic knee mechanism has the potential to exhibit biomimetic behavior with respect to the ACL and PCL behavior of the anatomical data [28].

$$r_1 \cos(\theta_1) + r_2 \cos(\theta_2) = r_3 \cos(\theta_3) + r_4 \cos(\theta_4) \quad (1)$$

$$r_1 \sin(\theta_1) + r_2 \sin(\theta_2) = r_3 \sin(\theta_3) + r_4 \sin(\theta_4) \quad (2)$$

$$r_5 \cos(\theta_5) + r_6 \cos(\theta_6) + r_7 \cos(\theta_7) = r_3 \cos(\theta_3) \quad (3)$$

$$r_5 \sin(\theta_5) + r_6 \sin(\theta_6) + r_7 \sin(\theta_7) = r_3 \sin(\theta_3) \quad (4)$$

$$r_8 \cos(\theta_8) + r_2 \cos(\theta_2) + r_7 \cos(\theta_7) = r_6 \cos(\theta_6) \quad (5)$$

$$r_8 \sin(\theta_8) + r_2 \sin(\theta_2) + r_7 \sin(\theta_7) = r_6 \sin(\theta_6) \quad (6)$$

$$\dot{r}_2 = \dot{\theta} (r_6 \sin(\theta_6 - \theta_2) - 2r_7 \sin(\theta_7 - \theta_2)) \quad (7)$$

$$\omega_2 = -\frac{\dot{\theta}}{r_2} (r_6 \cos(\theta_6 - \theta_2) + 2r_7 \cos(\theta_7 - \theta_2)) \quad (8)$$

$$\dot{r}_3 = -\dot{\theta} (r_6 \sin(\theta_6 - \theta_3) + 2r_7 \sin(\theta_7 - \theta_3)) \quad (9)$$

$$\omega_3 = \frac{\dot{\theta}}{r_3} (r_6 \cos(\theta_6 - \theta_3) + 2r_7 \cos(\theta_7 - \theta_3)) \quad (10)$$

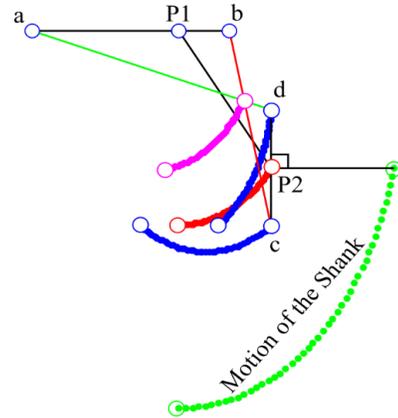


Figure 5. The Motion of the Mechanism

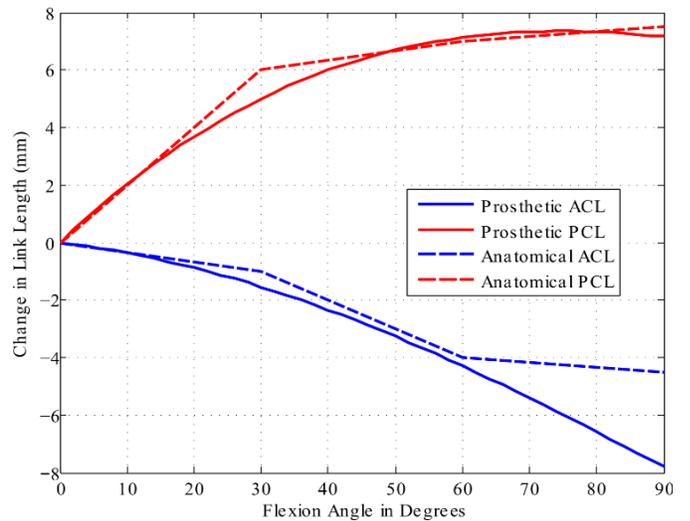


Figure 6. Changes in Length of ACL and PCL

The linear and angular velocity equations for change in length of ACL (r_2) and PCL (r_3) are defined in Equations (7)-(10). Figure 7 is a plot of the resultant linear velocities, where a constant velocity of 10 rad/s was assumed, as it is reported to the peak angular velocity during gait [29]. The rate of change of PCL length decreases as the knee approaches full flexion. This is compensated by the ACL, which changes length at a more rapid frequency as the knee is in flexion. Figure 8 shows the angular velocity of the rate of change of the angles that ACL (θ_2) and PCL (θ_3) make with the positive x-axis changes over flexion. While there is a negative change in angular velocity for the ACL, it is a positive change for the PCL link.

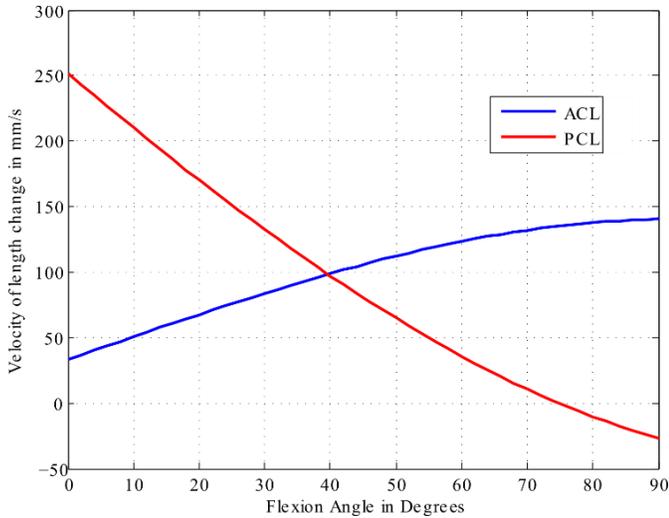


Figure 7. Linear Velocity Profiles of Prosthetic ACL and PCL

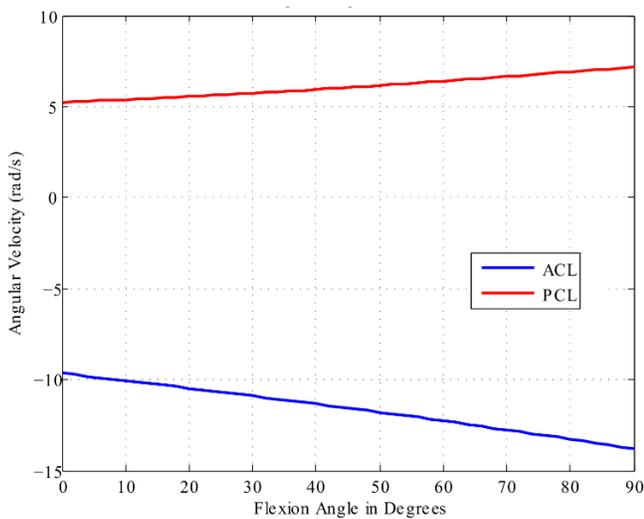


Figure 8. Angular Velocity Profiles of Prosthetic ACL and PCL

Results

Figure 9 shows a single subject fitted with a prosthetic simulator attached to the biomimetic knee to test the knee mechanism. Figure 10 shows the trial being conducted in the Computer Assisted Rehabilitation ENvironment (CAREN) by Motek Medical. The CAREN system is equipped with a Bertec split-belt treadmill, a 6-degree-of-freedom motion base, a 10-camera Vicon motion capture system, Bertec continuous force plates, and a panoramic screen for virtual interaction. The knee's motion was recorded using three reflective markers placed on the prosthesis to obtain the knee angles of the biomimetic knee as the subject walked on the treadmill.



Figure 9. 3D Printed Prosthetic Knee Fitted on the Simulator

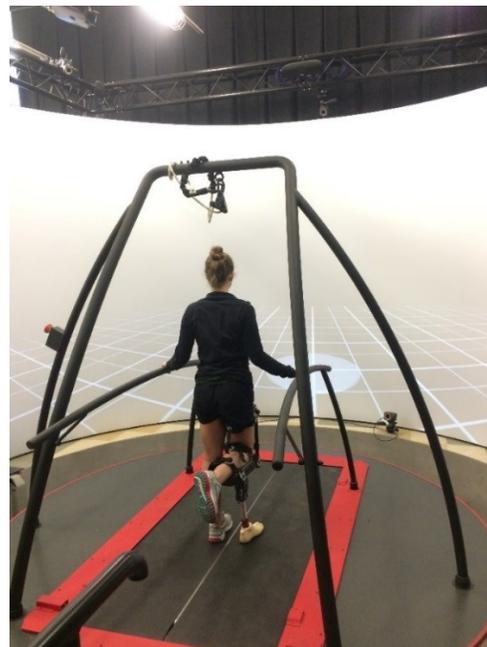


Figure 10. Subject on the CAREN System

The gait data were processed using a Matlab script to calculate the knee angles during gait. The results were compared to the standard able-body knee angle data by Winter [30]. Figure 11 shows the prosthetic gait data obtained with an Ossur Total Knee 2000 with a hydraulic return mechanism obtained by Ramakrishnan et al. [31]. The Biomimetic

knee shows knee angle trends more similar to that of Winter's data than does the Ossur Total knee. This is because the biomimetic knee was designed to have the same dimensions as a human knee and the flexible four-bar mechanism helps in stabilizing the motion. However, there is a clear difference at toe-off between Winter's data and the biomimetic knee, because the biomimetic knee prosthesis is completely passive and, hence, generates less push-off torque. This explains the drop in knee angle just before flexion. The Ossur Total knee has a hydraulic return mechanism and, due to the hydraulic resistance, has a smoother transition from push-off to full flexion [21].

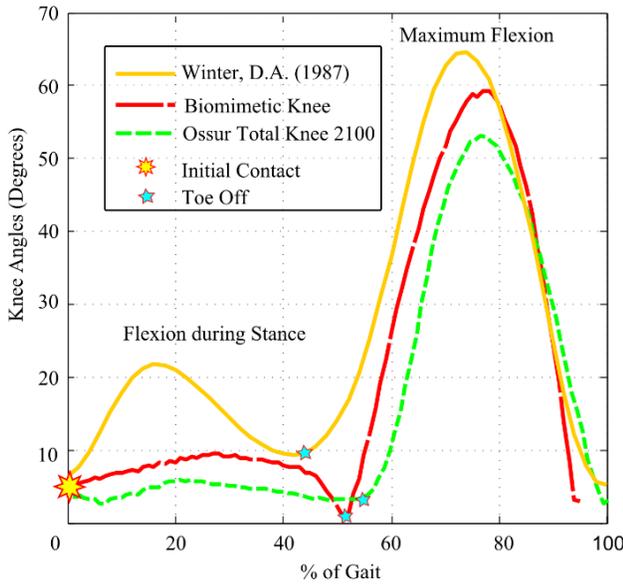


Figure 11. Comparison of Knee Angles

The results of this experiment were obtained using a 3D-printed prototype. The prototype demonstrates the kinematics of the proposed design using the minimum viable product. As stated previously, the kinematics can be tuned to better fit normal human knee motion. This can be done by adjusting the stiffness of the links or by an addition of dampers to make the motion smoother. The knee angles of the two prosthetic knees were compared to published gait data in Figure 11. Figure 12 shows the error between the standard data and the two knees. The average error of the biomimetic knee (6.46°) was lower than the average error of the Ossur Total knee (10.7°). This is because the biomimetic knee uses spring-based, four-bar stabilization, while the Ossur Total knee uses hydraulic resistance that disrupts the natural dynamic motion of the knee by adding excessive damping. This causes lower maximum flexion, as seen in Figure 11. The Ossur knee also had a larger standard deviation error (6.78°) than the biomimetic knee (4.22°).

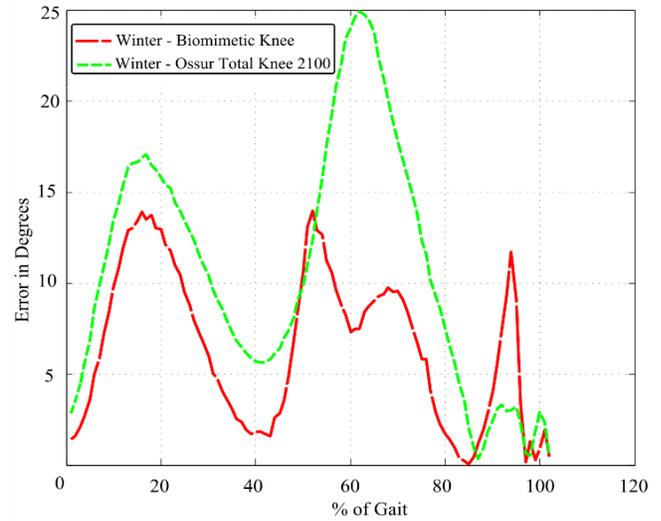


Figure 12. Difference Between Able-Body and Prosthetic Knee Angles

Discussion

This knee design, unlike conventional prosthetic knees, can be scaled to fit any individual, regardless of age, weight, height, or gender. This is because the knee is based on anatomical dimensions and scaling will not affect the function of the mechanism. This design can also be extended into orthotics and humanoid robotics. Orthotics already use similarly geared mechanisms to lock and unlock [32, 33], but the addition of the flexible four-bar links can aid in the stability of the mechanism and can assist the limb to move more like a human knee in the sagittal plane. In humanoid robotics, it could provide the necessary human-like gait motions that are lacking, since most robotic knees are single-axis joints. An actuated version of this mechanism could offer the stability and function that is required by humanoid robots.

This knee mechanism offers the flexibility to design the prosthetic knee to subjective requirements. The designer can start with the condylar measurements obtained from a scan to generate the gear profile. The designer is not constrained to utilize the gear parameters exactly as presented in Table 1. The designer can then model the kinematics of the mechanism to fit the amputee's level of control. This is important because many modern prostheses are rejected by users because the prosthetic components cannot be tuned to their individual specifications [10, 11]. The biomimetic knee prosthetic design has the potential to behave much like an actual human knee in the sagittal plane. The results show that it follows the human knee angle kinematics more closely than does an existing prosthetic knee [31]. The biomimet-

ic knee uses a polycentric mechanism like many popular mechanisms. This offers several advantages compared to a single-axis or simple weight-actuated mechanism, because it helps the shank and foot clear the ground to avoid tripping. In this specific case, the tibia gear rolls on top of the femur gear that helps it move in the vertical direction as well as the horizontal direction, which helps the foot clear the floor during terminal swing phase.

Another major advantage of this design is its ability to be customized. Tuning the prosthetic knee according to a person's body and gait helps in managing their quality of gait [34]. This is important to amputees in order to avoid long-term injuries, due to their physical asymmetry. The process of tuning the prosthesis may be an interesting avenue for future research, because it is important to understand the factors that influence amputee gait. Passive mechanisms such as this can also have simple control systems that may offer long-term benefits to amputees and relieve them from expensive, loud, and inefficient active prosthetic knees. Customization may lead this design to be used by amputees with various levels of control. This is an important factor to address with this knee design, because in current prosthetic technology there are certain types of knees that are designated for each of the K levels. This is a disparate system that can be streamlined with a highly customizable base platform. Further, this design's ability to be scaled to any size will offer better treatment protocols and faster iterations in order to provide the best prosthetic fit for the amputee.

The biomimetic knee described here bridges an important gap in current prosthetic technology trends. This simple design can be mass-produced using both traditional and modern manufacturing processes. This knee is also designed to reduce the cost of manufacturing, since it consists of only two major parts. Further testing is required to evaluate the full capacity of the design. This design can be used with a simple configuration of springs to a more complex fully actuated system for amputees with low muscle control that requires the knee to provide more assistance.

Conclusion

The biomimetic knee design presented here is simple and functional. The assembly consists of the femur gear and tibia gear that are easy to manufacture, which, in turn, will reduce the cost of manufacturing. The rest of the materials can be off-the-shelf components. This, combined with the ability of the design to scale for different amputees, makes this design unique. The results also show that the knee exhibits similar kinematics to the standard human knee model. In future iterations, the knee will be fine-tuned to exhibit more human-like kinematics.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. MRI-1229561

References

- [1] Perry, J., & Burnfield, J. M. (1992). *Gait Analysis: Normal and Pathological Function*, 271-279.
- [2] Morrison, J. B. (1970). The Mechanics of the Knee Joint in Relation to Normal Walking. *Journal of Biomechanics*, 3(1), 51-61.
- [3] Highsmith, M. J., Kahle, J. T., Bongiorni, D. R., Sutton, B. S., Groer, S., & Kaufman, K. R. (2010). Safety, Energy Efficiency, and Cost Efficacy of the C-Leg for Transfemoral Amputees: A Review of the Literature. *Prosthetics and Orthotics International*, 34(4), 362-377.
- [4] Sup, F., Varol, H. A., Mitchell, J., Withrow, T. J., & Goldfarb, M. (2009). Self-Contained Powered Knee and Ankle Prosthesis: Initial Evaluation on a Transfemoral Amputee. *IEEE International Conference on Rehabilitation Robotics, 2009*, 638-644.
- [5] Sup, F., Bohara, A., & Goldfarb, M. (2008). Design and Control of a Powered Transfemoral Prosthesis. *The International Journal of Robotics Research*, 27(2), 263-273.
- [6] Cantos, M. (2005). *Pirates & Peg Legs: A Historical Look at Amputation and Prosthetics*. History of Medicine Days, 16.
- [7] Childress, D. S. (1985). Historical Aspects of Powered Limb Prostheses. *Clinical Prosthetics & Orthotics*, 9(1), 2-13.
- [8] Highsmith, M. J., Kahle, J. T., Carey, S. L., Lura, D. J., Dubey, R. V., Csavina, K. R., et al. (2011). Kinetic Asymmetry in Transfemoral Amputees While Performing Sit to Stand and Stand to Sit Movements. *Gait & Posture*, 34(1), 86-91.
- [9] Boonstra, A. M., Schrama, J., Fidler, V., & Eisma, W. H. (1995). Energy Cost during Ambulation in Transfemoral Amputees: A Knee Joint with a Mechanical Swing Phase Control vs a Knee Joint with a Pneumatic Swing Phase Control. *Scandinavian Journal of Rehabilitation Medicine*, 27, 77-77.
- [10] Narang, Y. S. (2013). *Identification of Design Requirements for a High-Performance, Low-Cost, Passive Prosthetic Knee through User Analysis and Dynamic Simulation*. Master's thesis, Massachusetts Institute of Technology.
- [11] Goldfarb, M., Lawson, B. E., & Shultz, A. H. (2013). Realizing the Promise of Robotic Leg Prostheses. *Science Translational Medicine*, 5(210), 15.

- [12] Radcliffe, C. W. (1994). Four-Bar Linkage Prosthetic Knee Mechanisms: Kinematics, Alignment and Prescription Criteria. *Prosthetics and Orthotics International*, 18(3), 159-173.
- [13] Michael, J. W. (1999). Modern Prosthetic Knee Mechanisms. *Clinical Orthopaedics and Related Research*, 361, 39-47.
- [14] Shurr, D. G., Michael, J. W., & Cook, T. M. (2002). *Prosthetics and Orthotics*. Prentice Hall.
- [15] Mukul, P., Sadler, J., & Thorsell, E. (2010, May). Stanford-Jaipur Knee Joint for Transfemoral Amputees. *Proceedings of the 13th World Congress of the International Society for Prosthetics and Orthotics*, 1179-80.
- [16] Jin, D., Zhang, R., Dimo, H. O., Wang, R., & Zhang, J. (2003). Kinematic and Dynamic Performance of Prosthetic Knee Joint using Six-Bar Mechanism. *Journal of Rehabilitation Research and Development*, 40(1), 39.
- [17] Ponce-Saldias, D. A., Martins, D., de Mello-Roesler, C. R., Teixeira-Pinto, O., & Fancello, E. A. (2015). Relevance of the Hyperelastic Behavior of Cruciate Ligaments in the Modeling of the Human Knee Joint in Sagittal Plane. *Revista Facultad de Ingenieria Universidad de Antioquia*, 76, 123-133.
- [18] Lührs, B., Theile, F., Griesser, M., Richard, H. A., & Kullmer, G. (2004). U.S. Patent No. 6,749,640. Washington, DC: U.S. Patent and Trademark Office.
- [19] Etoundi, A. C., Burgess, S. C., & Vaidyanathan, R. (2013). A Bio-Inspired Condylar Hinge for Robotic Limbs. *Journal of Mechanisms and Robotics*, 5(3), 031011.
- [20] Zavatsky, A. B., & O'Connor, J. J. (1992). A Model of Human Knee Ligaments in the Sagittal Plane: Part 1: Response to Passive Flexion. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 206(3), 125-134.
- [21] Hamon, A., Aoustin, Y., & Caro, S. (2014). Two Walking Gaits for a Planar Bipedal Robot Equipped with a Four-Bar Mechanism for the Knee Joint. *Multibody System Dynamics*, 31(3), 283-307.
- [22] Massin, P., & Gournay, A. (2006). Optimization of the Posterior Condylar Offset, Tibial Slope, and Condylar Roll-Back in Total Knee Arthroplasty. *The Journal of Arthroplasty*, 21(6), 889-896.
- [23] Lewek, M., Rudolph, K., Axe, M., & Snyder-Mackler, L. (2002). The Effect of Insufficient Quadriceps Strength on Gait After Anterior Cruciate Ligament Reconstruction. *Clinical Biomechanics*, 17(1), 56-63.
- [24] Siebold, R., Axe, J., Irrgang, J. J., Li, K., Tashman, S., & Fu, F. H. (2010). A Computerized Analysis of Femoral Condyle Radii in ACL Intact and Contralateral ACL Reconstructed Knees Using 3D CT. *Knee surgery, Sports Traumatology, Arthroscopy*, 18(1), 26-31.
- [25] Siu, D., Rudan, J., Wevers, H. W., & Griffiths, P. (1996). Femoral Articular Shape and Geometry: A Three-Dimensional Computerized Analysis of the Knee. *The Journal of Arthroplasty*, 11(2), 166-173.
- [26] Yue, B., Varadarajan, K. M., Ai, S., Tang, T., Rubash, H. E., & Li, G. (2011). Gender Differences in the Knees of Chinese population. *Knee surgery, Sports Traumatology, Arthroscopy*, 19(1), 80-88.
- [27] Monk, A. P., Choji, K., O'Connor, J. J., Goodfellow, J. W., & Murray, D. W. (2014). The Shape of the Distal Femur. *Bone Joint J*, 96(12), 1623-1630.
- [28] Li, G., DeFrate, L. E., Sun, H., & Gill, T. J. (2004). In Vivo Elongation of the Anterior Cruciate Ligament and Posterior Cruciate Ligament during Knee Flexion. *The American Journal of Sports Medicine*, 32(6), 1415-1420.
- [29] McGibbon, C. A. (2012). A Biomechanical Model for Encoding Joint Dynamics: Applications to Transfemoral Prosthesis Control. *Journal of Applied Physiology*, 112(9), 1600-1611.
- [30] Winter, D. A. (1987). *The Biomechanics and Motor Control of Human Gait*, Waterloo.
- [31] Ramakrishnan, T., Muratagic, H., & Reed, K. B. (2016). Combined Gait Asymmetry Metric. *In Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference*, 2165-2168.
- [32] Thompson, J. W. (1978). U.S. Patent No. 4,090,264. Washington, DC: U.S. Patent and Trademark Office.
- [33] Kubein-Meesenburg, D., & Naegerl, H. (2002). U.S. Patent No. 6,443,994. Washington, DC: U.S. Patent and Trademark Office.
- [34] Kark, L., & Simmons, A. (2011). Patient Satisfaction Following Lower-Limb Amputation: the Role of Gait Deviation. *Prosthetics and Orthotics International*, 35(2), 225-233.

Biographies

TYAGI RAMAKRISHNAN is a doctoral candidate in mechanical engineering at the University of South Florida. He earned his BE in Mechanical Engineering from Anna University, Chennai, India, and MSME from the University of South Florida. His interests include biomechanics, human machine interfaces, wearable robotics, and interdisciplinary clinical rehabilitation measures. He may be reached at tyagi@mail.usf.edu

KYLE REED is an associate professor of mechanical engineering at the University of South Florida. He received the BS degree from the University of Tennessee and the MS and PhD degrees from Northwestern University, all in mechanical engineering. He was a postdoctoral fellow in the Laboratory for Computational Sensing and Robotics at the Johns Hopkins University. His interests include haptics, human-machine interaction, rehabilitation, prosthetics, and engineering education. He may be reached at kylereed@usf.edu

FEASIBILITY OF USING AN AUGMENTING SOLAR ARRAY FOR A GREENHOUSE IN NORTHWEST OHIO

Sudershan Jetley, Bowling Green State University; Binisha Shrestha, Bowling Green State University

Abstract

The idea of growing plants in environmentally controlled areas has existed since Roman times. By the first century A.D., the Romans were growing fruits and vegetables in simple greenhouses or cold frames. A green house is an example of a green building, making use of renewable energy and sunlight to grow plants during winter, when the temperature is too cold to grow the plants out in the open at northern latitudes. However, there is often not enough sunlight to sufficiently warm the greenhouse in some areas without additional heating. Such is the case in Northwest Ohio. One such greenhouse uses natural gas to provide this additional heating. The burning of natural gas incurs extra costs, and is a non-renewable energy source that emits carbon dioxide, thereby contributing to global warming. Therefore, there is a need to eliminate or reduce the use of natural gas for this greenhouse. One solution proposed by the owners of the greenhouse was to use a solar collector array placed by the side of the green house. In this paper, the authors describe an analysis carried out in collaboration with the greenhouse owners to evaluate the feasibility of using a solar array. The analysis involved using the design provided to the owners by a consulting company. The study evaluated the design from the perspective of cost savings per year, due to reduced gas usage and against the cost of installation and maintenance of the solar array system. The results were then used to calculate the payback period.

Introduction

An agricultural greenhouse consists of frames of metal or wood covered with a transparent material, which provides a suitable environment for the intensive production of various crops. All greenhouses collect solar energy. The basic operational principles of greenhouses include: collection of light and conversion to heat; storage of thermal energy; release of thermal energy; temperature moderation; a controlled environment; and, protection from severe weather and animals. A greenhouse is essentially an enclosed structure that traps short-wavelength solar radiation and stores long-wavelength thermal radiation to create a favorable microclimate for plant growth. It is not a problem to maintain the temperature inside the greenhouse when there is sun during most times of the day, except in the Northwest Ohio location, where there is a lot of snowfall during the winter and there are

fewer sunny days. Hence, there is a need for additional thermal energy to heat the greenhouse and maintain the required temperature for plants to grow.

Different renewable energy sources can be applied in heating the greenhouse, such as geothermal, solar, and biomass instead of using fossil fuels, which have predominantly been used thus far. The intent of this study was to analyze the new design of a greenhouse located at the J.C. Reuthinger Memorial Preserve on Oregon Road in Perrysburg, Ohio. The heat demand in the greenhouse was provided by heaters that burn natural gas. The main problem with this type of heater is the cost of the fuel. That is why the owners decided to utilize a solar energy storage system in the greenhouse. As only low-temperature heat was needed for pre-heating the air in the greenhouse, the use of solar panels was proposed in order to partially reduce heating demand of the heater, thereby reducing the amount of fuel consumption. The purpose of this study was to evaluate a design of a new heating system for the greenhouse that incorporated solar panels for economic viability.

The Greenhouse

The greenhouse was 30' x 96' and used predominantly for growing perennial plants. Figures 1 and 2 show how the greenhouse was built with an aluminum frame system that was shrouded by clear plastic sheets. It was a double poly structure with air pumped between two layers. The temperature maintained inside the greenhouse for these plants was 60°F. Greenhouse use starts the first week of January and goes until the first week of June. Since this area of the country faces a severe climate during the winter, which starts from December and lasts until March, the greenhouse needed to be heated during this period. In the proposed design, solar panels would be used to collect the heat from sun as the first stage of heating. Solar energy collected by solar panels would then be used to heat the fluid stored in a drain-back tank, which would then give power to a natural gas heater to operate and heat the tube in the heat exchanger, thereby heating the floor of the greenhouse.

Description of the System

Figure 3 shows a schematic diagram of the proposed heating system. The proposed design consisted of five flat-plate

solar thermal collectors manufactured by Alternate Energy Technologies (AET). The selected solar collector system was chosen based on the availability in the local market and domestic demand. Aluminum flat-plate solar collectors were popular because of their low investment costs. The number of solar thermal collector panels was determined according to the ground surface area of the greenhouse.



Figure 1. Exterior View of the Greenhouse



Figure 2. Interior View of the Greenhouse

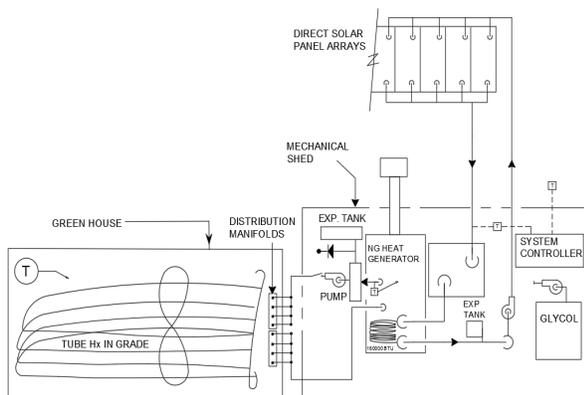


Figure 3. Proposed System Flow Diagram (Source: greenhouse owners)

Theory of Operation

As the sun starts to shine on the solar collectors, the collectors begin to heat up. This system had a differential controller that sensed temperature differences between water leaving the solar collector and the coldest water in the tank. When the water in the collector was about 12°F warmer than the water in the tank, the controller turned on the pumps. When the temperature difference dropped to about 3-5°F, the pump turned off. In this way, the water always gained heat from the collector when the pump was operating. The solar installer manually set the differential temperature of the controller at the time of system installation. The pumps were very quiet.

The system collected heat by circulating the water in the collector loop through the collectors, the drain-back reservoir, and a heat exchanger. When the pumps turned on, the water in the collector loop was circulated through the solar collectors where it was heated. The return water from the solar collectors passed through the drain-back tank and to the heat exchanger. The heat exchanger gave up the collected heat to the water in the main underground heat exchanger. The now cooler water was returned to the collectors to continue to collect heat. When there was no longer a marked difference between the temperature at the solar collector and at the solar storage tank (4°F), the controller automatically turned the pumps off—since there was no heat to be gained. This process was repeated continuously throughout the day. This was all done automatically and required no interaction on the part of the system owner. Drain-back systems provide a fail-safe method of ensuring that collectors and collector-loop piping never freeze, by removing all water from the collectors and collector-piping loop when the system is not collecting heat. Freeze protection is provided when the system is in the drain mode. Water in the collectors and exposed piping drains into the insulated drain-back reservoir tank each time the pump shuts off. A slight tilt of the collectors is required in order to allow complete drainage. A sight glass attached to the drain-back reservoir tank showed when the reservoir tank was full and the collectors had to be drained. For this reason, the drain-back system was protected from freeze damage at all temperatures.

Methodology

The basic idea was to compare two greenhouse heating systems: a conventional fossil-fuel system and a hybrid system in which part of the heating demand was covered by solar collectors, heat storage, and a heat pump. The economic feasibility of heating the greenhouse using these two methods was analyzed.

Data Collection

The data collection for the evaluation of the project involved the following areas:

- The monthly averaged, mean maximum, and mean minimum temperatures from January 2014 to July 2014 were collected. These data were used to calculate the greenhouse's thermal need. These data were obtained from a commercial database and from the Internet source AccuWeather.
- The price of natural gas during the recommended period of time was also obtained from Columbia Gas of Ohio, the local gas supplier company.
- Equipment prices for the installation of solar panels were obtained from the manufacturer recommended by the designer. For all of the systems, mean costs with reference to a greenhouse ground-surface area of 2700 sq. ft. were used. This was the actual ground surface of the greenhouse that was being analyzed.
- For the evaluation of the proposed design, the solar system costs, installed collector surface area, and operating costs were obtained from the manufacturer and the suppliers.

Using all of the parameters mentioned, the total thermal needs, the parts covered by the solar panel system and the lifecycle costs of two heating systems were calculated. The lifecycle cost is the sum of all the costs associated with an energy system; in this case, investment, fuel, electricity, and maintenance, during its lifetime at current dollar value.

Analysis and Results

Table 1 shows the average temperature during the six months and the remaining temperature increase needed to be attained in order to keep the greenhouse at 60°F. Figure 4 shows the location of the two gas heaters currently used to heat the greenhouse in winter. Figure 5 shows the fuel consumed for the six months of 2014. The information about the fuel consumed was obtained from monthly gas bills for 2014. It was also found that the unit price of the gas supplied by Columbia Gas of Ohio varied from month to month. Data on the fuel consumed every month and the estimated unit rate of natural gas were used to calculate the cost incurred each month for the consumption of gas (see Table 2). February, the coldest month during that year, had the highest consumption of fuel.

Only about 80-85% of the heat produced by the heaters was used for heating the space; the remaining 15-20% was lost through the ground, covering material, and other related avenues [1]. Since the greenhouse being analyzed had two

layers of plastic covering, based on published literature [1], it was assumed that 15% of the heat was wasted. Using the information on the temperature required, the actual temperature, the fuel consumption, the heat-loss percentage, and the amount of gas used and its cost were determined to maintain the temperature regardless of the fuel used for heat loss, as shown in Table 3. The actual gas used, A , was 85% of the total gas used, T .

Table 1. Temperature Difference Each Month

Months (2014)	Temperature needed to maintain (°F)	Actual average temperature (°F)	Remaining temperature needed to be raised (°F)
January	60	25.84	34.16
February	60	18.92	41.08
March	60	29.75	30.25
April	60	49.75	10.25
May	60	61.69	-1.69
June	60	72.25	-12.25



Figure 4. Location of Heaters inside the Greenhouse
(Source: J.C Reuthinger Preserve, Oregon Road, OH)

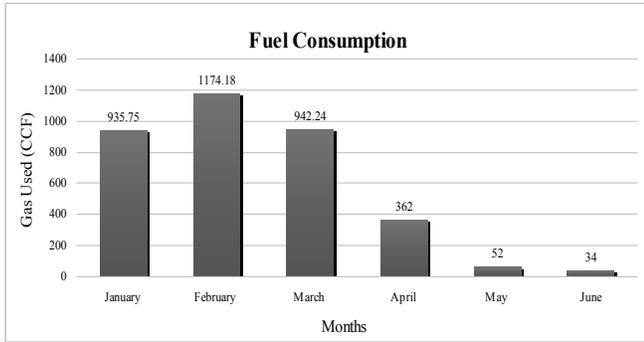


Figure 5. Fuel Consumption for the First Six Months of 2014

Table 2. Historic Rates in \$/CCF (Source: Columbia Gas of Ohio, 2014)

Billing period end date (2014)	Rate (\$/CCF)	Fuel consumed (CCF)	Total amount (\$)
January	\$0.5697	935.74	\$533.09
February	\$0.6547	1174.18	\$768.73
March	\$0.6145	942.24	\$579.00
April	\$0.5984	362	\$216.62
May	\$0.5821	52	\$30.26
June	\$0.58	34	\$19.72

Table 3. Temperature and Fuel Consumption Considering Heat Loss

Month	Temperature required to raise (°F)	Total Gas used = T (CCF)	Actual Gas used for raising temperature = A (CCF)	Cost incurred in actual Gas (A) used to raise temp.
January	34.16	935.74	795.37	\$ 453.12
February	41.08	1174.18	998.05	\$ 653.42
March	30.25	942.24	800.90	\$ 492.15
April	10.25	362	307.7	\$ 184.12
May	-1.69	52	44.2	\$ 25.72
June	-12.25	34	28.9	\$ 16.76

Table 3 shows how much fuel was needed to raise the required temperature to 60°F in each month, assuming a 15% heat loss. It is clear from Table 3 that February, being the coldest month, required the most fuel, whereas June, being the hottest month, had the least amount of fuel con-

sumption. No fuel was used for heating during the months of May and June; the numbers for cost and fuel consumption associated with these months were the lowest cost the owners had to pay, even when no fuel was used at all. These data were used in further calculations and analyses during the later part, where the heating was partially replaced by the solar panels.

Heat Calculations Using the Solar Panels

The use of five 4' x 8' solar panels was based on the ground coverage of the greenhouse, which was about 2700 sq. ft. The storage capacity of the 150-gallon drain-back tank was also based on the surface that required heating (i.e., 27,000 sq. ft.). To calculate the reduction in fuel due to the use of solar panels, it was necessary to determine how much heat one solar panel could produce. This information was provided by the expert engineers from Radiant Tech Flooring, who have worked in the industry for more than 20 years. According to them, an AET 4' x 8' solar panel, when exposed to continuous sun for five hours a day, can produce about 30,000 BTUs of energy. Being in Northwest Ohio, it is impossible to get that amount of sunlight every day during winter.

From 1961 to 1990, the National Renewable Energy Laboratory monitored the hourly values of direct beam and diffuse horizontal solar radiation in various cities across the United States. They published a table for various cities to show how much direct sunlight a solar flat-plate collector received for a given tilt per day [2]. The values for Toledo, which is adjacent to Perrysburg, are given for each month. Values for Toledo were used in this study. Figure 6 shows the amount of heat generated each month and how it could be calculated.

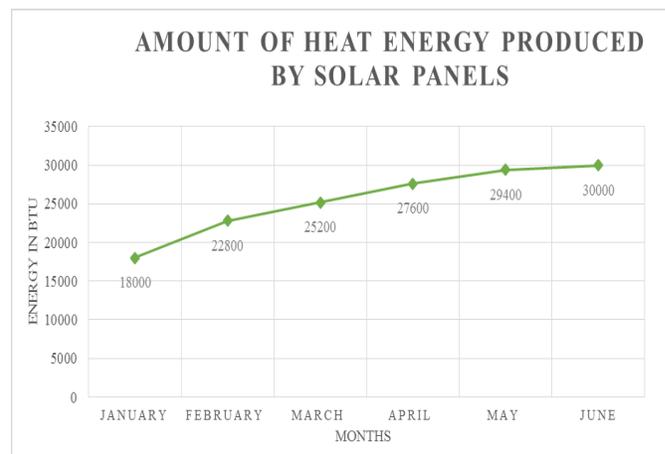


Figure 6. Amount of Heat Energy Produced by the Solar Panel Each Month

The next step was to calculate the amount of heat energy that was required every month to heat the water in the drain back tank in order to maintain the required temperature. An example of this calculation for the month of January is as follows:

Temperature required to maintain (T) = 60°F
 Amount of water needed to be heated (W) = 150 gallons
 1 gallon of water = 8.33 pounds (lbs.)
 150 gallons of water (W) = 1249.5 lbs.
 Initial temperature in January (T_0) = 25.84°F
 Temperature differences (ΔT) = $T - T_0 = 60 - 25.74 = 34.16^\circ\text{F}$
 Percentage of heat loss to be considered (H_0) = 15%

Let the energy required to heat the water to the desired temperature without heat losses = Z
 $Z = W * \Delta T$

hence, $Z = 1249.5 * 34.16 = 42682.92$ BTUs

Let the energy required to heat the water to the desired temperature with heat losses = P . Adding the 15% heat loss factor, $P = 49085.35$ BTUs

This was the actual amount of heat required in January to maintain a temperature of 60°F. On the other hand, Figure 6 shows that the amount of energy produced by the solar panel during this month was only 18,000 BTUs. The gas heaters provided the rest of the energy—31,085.35 BTUs. From the calculation shown below, it is clear how many degrees 18,000 BTUs can raise the temperature, which is the amount the solar panel could produce in January.

Energy produced (E) = 18,000 BTUs
 Water weight (W) = 1249.5 lbs.
 Temperature = TT
 $TT = E/W = 18,000/1249.5 = 14.40^\circ\text{F}$

For January, the average temperature was 25.84°F (from Table 1). So the temperature achieved through solar panels = Average temperature + $TT = 5.84 + 14.40 = 40.24^\circ\text{F}$

It was clear that 18,000 BTUs could only raise the temperature to 40.24°F from 25.84°F, and the temperature still needed to be raised an additional (60-40.24) = 19.76°F. To achieve this, additional heat was then supplied by the existing natural gas heaters. The amount of gas and the associated costs were then determined to be 460.01 CCF for a cost of \$262.06. Table 4 shows the energy required for the remaining months and the associated costs. Using the findings noted above, the costs and savings would be comparable if a solar heating system were installed. Table 5 shows a summary of the data, and Figure 7 shows a graphical representation. From Table 5, it is clear that after the installation of the solar panels, the amount of fuel consumption decreased and, in six months, the owners were able to save \$1036.43 on their gas bills. However, the installation costs of the solar panel and its components need to be taken into account to calculate the payback period. Figure 7 shows the comparative chart.

The cost of the project, which in this case was the installation cost of the solar panel and its components, was obtained from the manufacturers, as shown below:

Five 4' x 8' solar collectors	\$7826
Mounting hardware and strut	\$600
150-gallon storage tank	\$1500
Plumbing mechanical package	\$525
Pumps	\$500
Passive solar heat exchanger	\$2500
Labor costs for installation	\$4500
Miscellaneous components	\$1549
Total	= \$19,500

Table 4. Heat Energy Required, Fuel Consumed, and the Associated Costs for Each of the Six Months

Month	Heat required to maintain the temp. + 15% heat loss (BTUs)	Actual heat produced by Solar Panel (P) (BTUs)	Temp. raised by Heat P (°F)	Remaining temp. (Tr) (°F)	Fuel consumed to raise Tr (CCF)	Cost associated with fuel (\$)
Jan.	49,085.35	18,000	40.24	19.76	460.01	262.06
Feb.	51,329.46	22,800	37.16	22.84	554.78	362.82
Mar.	37,797.37	25,200	49.91	10.09	267.08	163.98
Apr.	12,807.37	27,600	71.83	-11.83	0	0
May	-2428.40	29,400	85.21	-25.21	0	0
June	-17602.33	30,000	96.25	-36.25	0	0

Table 5. Comparison between the Two Heating Systems

Month	Conventional heating system (A)		Solar panel-based heating system (B)		Total savings
	Fuel consumed	Total cost	Fuel consumed	Total cost	Saving cost (A-B)
Jan.	795.37 Ccf	\$453.12	460.01 Ccf	\$262.06	\$191.06
Feb.	998.05 Ccf	\$653.42	554.78 Ccf	\$362.82	\$290.60
Mar.	800.90 Ccf	\$492.15	267.08 Ccf	\$163.98	\$328.17
Apr.	307.7 Ccf	\$184.12	0	0	\$184.12
May	44.2 Ccf	\$25.72	0	0	\$25.72
June	28.9 Ccf	\$16.76	0	0	\$16.76
Total					\$1,036.43

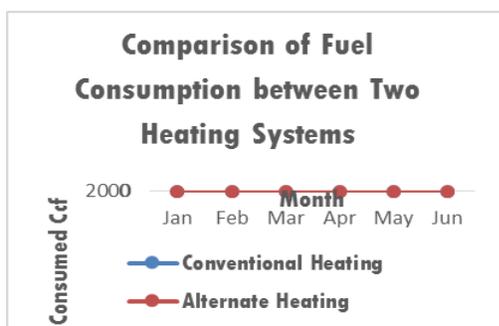


Figure 7. Comparative Chart of the Fuel Consumed

In general, the discounted payback period is the amount of time it takes to break even, based on the initial investment, the profit being generated, and taking into account the time value of money. The time value of money is accounted for by a discount rate, or the rate at which the value of money decreases in the future. The discounted cash inflow for each period can be calculated using Equation (1) [3]:

$$\text{Discounted cash flow} = CF/(1+i)^n \quad (1)$$

where, *CF* is actual cash flow; *i* is the discount rate; and, *n* is the period to which the cash inflow relates.

The rest of the procedure is similar to the calculation of simple payback period except that the discounted cash flows, as calculated above, need to be used instead of actual cash flows. The cumulative cash flow was replaced by the cumulative discounted cash flow [3]. Discounted payback period was calculated using Equation (2):

$$A + \frac{B}{C} \quad (2)$$

where,

A = the last period with a negative discounted cumulative cash flow

B = the absolute value of discounted cumulative cash flow at the end of period *A*

C = the discounted cash flow during the period after *A*

Now, calculating the discounted payback period using these values,

Initial investment = \$19,500

Annual saving each year = \$1036.43

Discount rate/interest rate = 5% (most commonly used)

Years = 18 years (from simple payback period)

Figure 6 shows the calculated discounted cash flow of each period and the cumulative cash flow. From Table 6, the last period with a negative discounted cumulative cash flow was (*A*) = 17 years. The absolute value of the discounted cumulative cash flow at the end of period *A* (*B*) = \$7823.57. The discounted cash flow during the period after *A* (*C*) = \$430.118. The discounted payback period Equation (2). So, the discounted payback period was $17 + (7823.57/430.118) = 35.18$ years. Thus, the discounted payback period was about 35 years.

Conclusions

With the installation and operation of solar panels, it can be seen that for the first three months of 2014, the owners were able to reduce overall fuel consumption. But since the temperature rose after April, there was a need to vent excess heat to maintain the temperature. Thus, it may be better to start heating with the solar panels from November instead of heating the greenhouse from January, given that tempera-

Table 6. Cumulative Discounted Cash Flow

Year (N)	Cash flow (CF) (\$)	Present value factor (\$) $PV= 1/(1+i)^n$	Discounted cash flow = CF x PV (\$)	Cumulative discounted cash flow
0	-19,500	1	-19500	-19500
1	1036.43	0.952	986.6814	-18513.31864
2	1036.43	0.907	940.042	-17573.27663
3	1036.43	0.863	894.4391	-16678.83754
4	1036.43	0.822	851.9455	-15826.89208
5	1036.43	0.783	811.5247	-15015.36739
6	1036.43	0.746	773.1768	-14242.19061
7	1036.43	0.71	735.8653	-13506.32531
8	1036.43	0.676	700.6267	-12805.69863
9	1036.43	0.644	667.4609	-12138.23771
10	1036.43	0.613	635.3316	-11502.90612
11	1036.43	0.584	605.2751	-10897.631
12	1036.43	0.556	576.2551	-10321.37592
13	1036.43	0.53	549.3079	-9772.06802
14	1036.43	0.505	523.3972	-9248.67087
15	1036.43	0.481	498.5228	-8750.14804
16	1036.43	0.458	474.6849	-8275.4631
17	1036.43	0.436	451.8835	-7823.57962
18	1036.43	0.415	430.1185	7393.46117

ture starts to fall in Northwest Ohio from these months. However, the payback period of more than 35 years is excessive, and so the owners may not want to invest in this venture.

References

- [1] Karlsson, M., & Wayne, V. (2014). Controlling the Greenhouse Environment. Kenai, AK: University of Alaska.
- [2] Marion, W., & Wilcox, S. (1994). Solar Radiation Data Manual for Flat- Plate and Concentrating Collectors. Golden, Colorado: National Renewable Energy Laboratory.
- [3] Accounting Explained (2013). Retrieved from <http://accountingexplained.com/managerial/capital-budgeting/discounted-payback-period>

Biographies

SUDERSHAN JETLEY is an associate professor in the Department of Engineering Technologies at Bowling Green State University. He has over 30 years of experience teaching a wide variety of courses at both undergraduate and graduate levels. He is the former head of the department and has published many articles in technical journals and conferences. Dr. Jetley may be reached at sjetley@bgsu.edu

BINISHA SHRESTHA recently graduated with master’s degree in Construction Management from Bowling Green State University. She is currently working as a Project Engineer at Lake Erie Electric, Inc. in Cleveland, Ohio. She also holds a bachelor’s degree in Architecture from Nepal, where she did her research in Sustainable architecture for her thesis in undergraduate degree. She has over two years of experience as a senior architect and site supervisor for many residential and small-scale commercial projects. Binisha Shrestha may be reached at bnsasth@gmail.com

USING LIDAR AND ANIMATION TO DEVELOP SURVEILLANCE STRATEGIES

Roger Ruggles, Lafayette College; Karen Ruggles, DeSales University

Abstract

Stationary ground-based Light Detection And Ranging (LIDAR) is used to measure points in three-dimensional (3D) space, and the processing of these points can generate an extremely accurate 3D model of the scanned area. One 360-degree scan from a LIDAR instrument can generate more than four million points. Using the combined points from multiple scans, an accurate 3D scene can be generated with computer software. By creating a virtual camera with select parameters, the viewer can virtually move to any point in the scene in order to display the scene from the chosen vantage point. This capability is used in conjunction with animation to assess the placement of surveillance cameras. The resulting animation highlights areas that are visible to the camera as well as those areas that the camera is unable to detect. The animation provides additional assessment for systems with multiple cameras to clearly visualize overlapping areas of coverage and to avoid blind spots.

In this paper, the authors describe the process of conducting the LIDAR scans, transferring the scanned information into an animation program, and the process of assessing points in the scene to establish the location for surveillance cameras. This site assessment process provides an analysis that reduces the cost of a surveillance system by decreasing the number of cameras required, while enhancing the surveillance coverage.

Introduction

Development of surveillance strategies in the placement of security cameras has received increased interest, as security concerns have escalated. Most current methods of siting security cameras are founded on a two-dimensional (2D) footprint of the area of security concern. The 2D ground plane of the surveillance area in some models is extruded to develop a three-dimensional (3D) area of surveillance. Extrusions such as this provide vertical planes that simulate walls, barriers, or other impediments to the line of site of the surveillance camera. Recent developments in LIDAR (Light Detection And Ranging) technology have provided instruments that have the capability of creating a 3D surveyed representation of any site. These representations are a collection of points referred to as a point cloud. Each point in the point cloud is accurate to within three millimeters, while

the number of points obtained from a LIDAR scan can be several million. When the point cloud is displayed, the sheer number and accuracy of the points in the point cloud appears to be a photorealistic image of the area. The photorealism and accuracy of the points provides the ability to move anywhere within the point cloud and visualize the 3D scene from that point.

In this study, the authors coupled the point cloud images obtained from LIDAR scans and the manipulation of digital cameras in animation software to locate surveillance cameras. The advantage of this process is that an actual 3D model of the surveillance area is used to place surveillance cameras. Within this 3D image, the 3D view of the camera defined by the field of view (FOV), depth of field (DOF), and other characteristics of the selected camera is inserted to establish the 3D space visible to the camera. Animation is then used to assess the placement of cameras for best coverage of the area of security concern. This assessment provides strategic placement of security cameras in order to maximize coverage and minimize cost. The animated scene can also have lighting applied to simulate the actual lighting within the FOV of the camera. Additional lights can be added to the scene to assess increased visibility. The animated scene can also introduce obstacles to the FOV of the camera in the form of vehicular traffic, pedestrian traffic, or vegetative growth to simulate a variety of actual conditions.

Current Practice

The current practice in placing surveillance cameras typically follows one of two design scenarios: 1) minimize the number of cameras, while covering the region of interest (ROI), and 2) maximizing the ROI using the minimum number of cameras. In each case, the economics of the total cost of the system is typically a driving factor in the ultimate design. Initial camera acquisition and installation costs, in addition to system maintenance and data storage costs, are all factors driving the cost of the security system. Interviews with individuals from institutions and municipalities that currently own, operate, and maintain surveillance systems indicate that the placement of the cameras of the system primarily follow guidelines identified by the Department of Homeland Security that are designed to secure and regulate physical access to facilities, enhance safety, prevent crimes, and assist in the investigation of criminal acts. Video surveillance also supports facility protection with its visi-

ble presence, and detects and deters unauthorized intrusion at facilities [1]. The design of camera placement within a security system varies dramatically in practice. The process can be as simple as standing at a particular location and observing the area to establish what can be viewed by a camera, or it can be a complex process, as described by Ghanem et al. [2], which uses a convex binary quadratic program to site cameras. This method allows the user to provide user-defined parameters in the design process in order to optimize the camera placement solution. The primary limitation to all current models is the limitation of placing cameras in a true 3D scene of the ROI and analyze and display the coverage in a 3D perspective.

Problem Definition

In this paper, the authors detail the use of LIDAR 3D scans of ROI in the development of a camera surveillance system through animation. Current practice in siting security cameras, as described in the previous section, has a very wide range of sophistication. Even the most sophisticated methods provide limited 3D visualization of the region of interest. LIDAR scans of areas provide a 3D representation of the area and allow the viewer to digitally move anywhere within the scene and view of the region of interest from that vantage point. Figure 1 shows an example LIDAR scan. The scan shown in Figure 1 consists of millions of points. Each point is plotted within the established coordinate system. A viewer (or camera) can be moved to any location within the scan and from that position observe all points to which there is a line of site. This process allows the actual siting of a camera and the determination of the coverage that results.

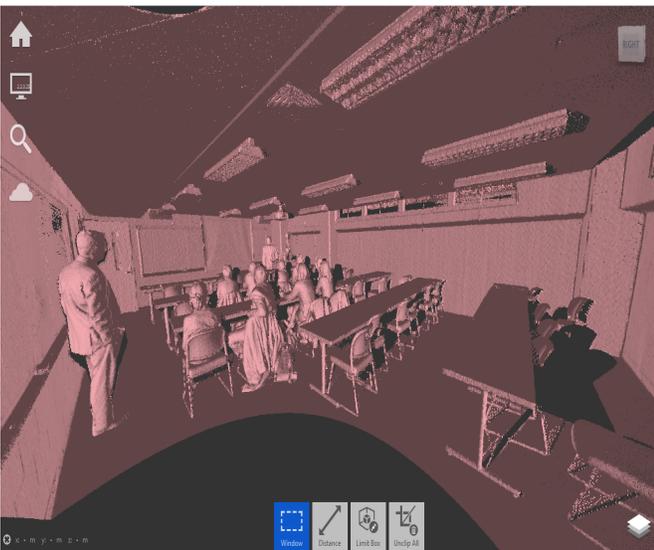


Figure 1. Example of a LIDAR Scan with the Point Cloud showing a True 3D Representation of the Area

The objective of this study was to link the LIDAR 3D scanned image to animation software in order to assess the visual coverage of the ROI and in doing so minimize the number of cameras, while maintaining or increasing the FOV. This assessment will decrease costs and enhance coverage of the security system. In addition, the animation assesses lighting within the ROI and provides a simulation of vehicular traffic, pedestrian traffic, or vegetative growth.

LIDAR Scanning to Obtain a 3D Image of the Surveillance Area

LIDAR is a remote sensing method that uses light in the form of a pulsed laser to measure ranges (variable distances) to objects. These light pulses are combined with other data recorded by the LIDAR instrument to generate precise, 3D information about the shape of objects and other surface information [3]. The laser pulse from a LIDAR instrument is similar to radar technology, which uses radio waves (instead of light) to range to an object by measuring the time delay between the transmission of a pulse and the detection of the return signal. LIDAR instruments can collect an enormous number of points in a very short period of time. The image acquired and shown in Figure 1 had a collection rate of approximately 120,000 points per second. The 360-degree scan obtained in order to generate Figure 1 was made in about four minutes, which resulted in approximately 29 million points in this one scan. As describe above, LIDAR uses a laser to measure distance and, therefore, will only record distances to objects that are within the line of site of the instrument. In order to obtain a true 3D rendering of a region of interest, several scans—each from a different location within the area—may be required.

Figure 2 shows the same scene depicted in Figure 1, but from a different viewing angle. The darker region of Figure 2 in the lower left of the image indicates an area that was not visible by the scan position of the instrument and, therefore, had no returns. In order to represent this region in the 3D rendering, the LIDAR instrument would be moved to a point where this region is visible and another scan would provide points in this area. Software available with the LIDAR scanner allows the two scans (two point clouds) to be combined. Multiple scans can be digitally assembled to represent the entire scene. LIDAR also has the capability of monitoring reflectance and amplitude. Figure 3 is an example of the reflectance recorded for each point measured in the scan shown in the two previous figures. The reflectance parameter displays as white or lighter color indicating high reflectance and a darker color indicating low reflectance. Since security cameras are recording reflected light, areas of low reflectivity may result in poor image quality. Analysis of the LIDAR scan using the reflectivity return can identify

trated the scanned room through transparent surfaces such as windows or semi-transparent surfaces such as leaves and other collected points. The block in the lower right corner represents the room displayed in Figures 1 and 2. The software can easily “fence-in” a designated area by clicking around the desired location to keep (in this case the room) and clip out all other points (see Figure 6). What remains after the clean-up is consistent with the visuals from Figures 1 and 2. Clean-up is an easy way to reduce file size and ignore points that should not be part of the mesh creation for the following steps. The file is now ready to be exported for mesh creation. Typically, exporting as a *.pts file will yield the best results for the next program. There are a few programs, such as MeshLab and CloudCompare, which offer the necessary commands to create normals and tessellate a mesh. In this study, the authors examined the use of CloudCompare [5]. It should also be noted that CloudCompare has the capability to clean-up the original point cloud, which would negate the use of Autodesk Recap 360, but the ease of use in Autodesk Recap 360’s “fence-in” process facilitated a faster pipeline to the end product.



Figure 5. Demonstration of Full-Point Cloud from a Single Scan Prior to Point Cloud Cleanup

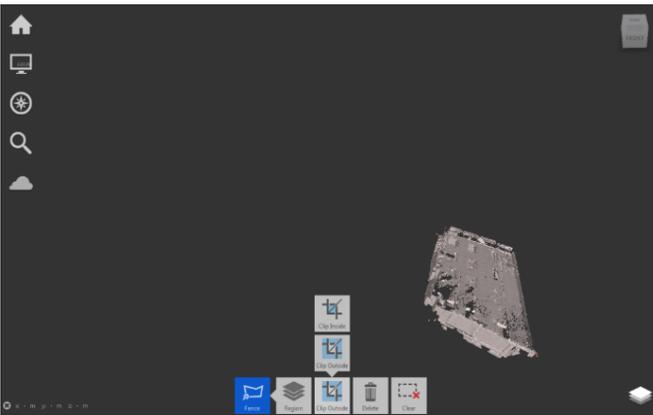


Figure 6. Fencing-in the Desired Location and Clipping All Other Points

CloudCompare offers an extensive list of compatible file extensions, including *.pts (see Figure 7). Upon opening a *.pts file, it is necessary to skip the header in the text file; this may be important for other file types as well.

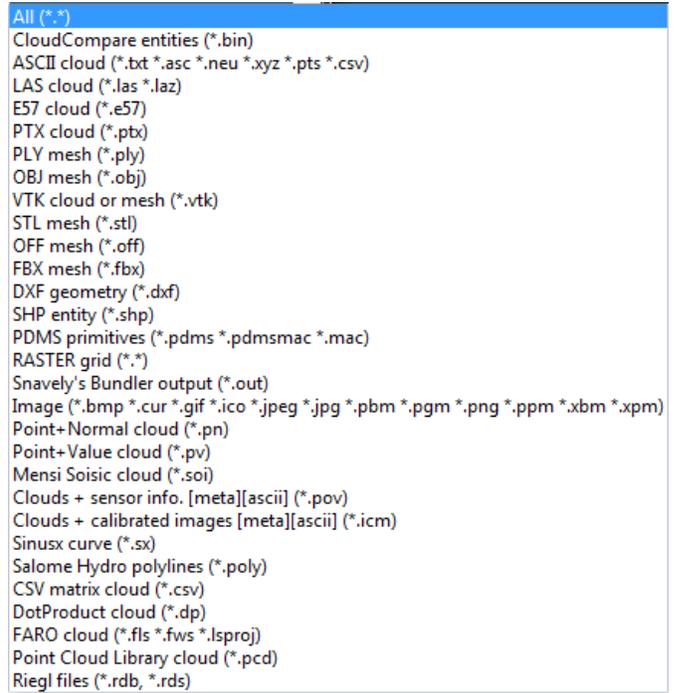
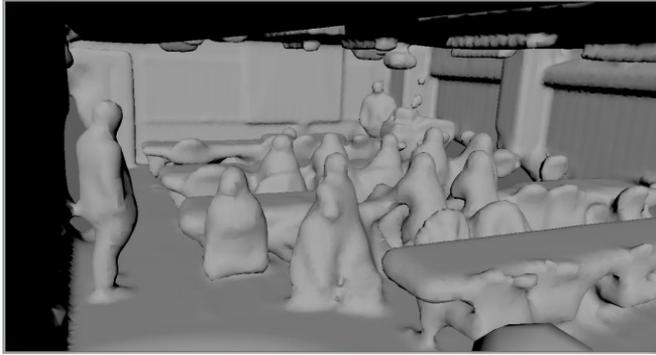


Figure 7. List of File Extensions Compatible with CloudCompare

Upon opening the point cloud in CloudCompare, it is necessary to prepare it for mesh creation. To do so, the point cloud must first generate normals. Normals assist the orientation of a face within a mesh. Since a face has no thickness, it is important for the face to determine which side of the face is inward-facing and which side is outward-facing. This is particularly important for the animation software to render the texture or color on the mesh and to create usable data in lighting situations. As such, this is a common and easy procedure in CloudCompare. With the point cloud selected, simply choose Edit>Normals>Compute. Depending on the settings, each point determines which way to orient a normal, based off of the points around it.

Once the normals are computed, the points are now ready to generate a mesh. The authors used the established Poisson Surface Reconstruction method for this study. Figure 8 shows the differences based on the Octree depth calculation; the higher the Octree depth, the better the detail of the mesh, but the file size can become unruly when reaching high numbers. When the mesh is generated, CloudCompare can export many common polygonal mesh types, such as *.fbx, *.obj, *.stl, *.dxf, *.ply, and *.ma.



(a) Octree Level 8—File Size 18.4 MB



(b) Octree Level 13—File Size 1.15 GB

Figure 8. Mesh within Autodesk Maya

Importing the mesh file into animation software such as Autodesk Maya, Autodesk 3DS Max, or Blender is easily completed. Next is a demonstration of the use of animation techniques in Autodesk Maya [6]. Note that CloudCompare can export the Maya file extension *.ma, which simply needs to be opened in Maya versus imported. With the mesh created and imported into the animation software, the animator can create cameras to view the mesh from any angle, lights to emulate visual conditions such as night visibility, street lighting, and shadows, and animation to simulate typical movement in a scene to demonstrate visual coverage.

Figure 9 shows how multiple cameras can be placed in one Maya scene, similar to visualizing the point cloud in the LIDAR software. Note that the mesh was based off of the point cloud, so any shortcoming in terms of line of site from the scan will also be reflected in the mesh. The camera benefit to the animation software is the customization of the camera's settings. In Maya, each camera can have independent settings for position and rotation, making it easy to set up multiple cameras and switch between them for comparison. Additionally, parameters such as resolution, focal length, center of interest, and depth of field (among others) can all be adjusted per camera. These settings can be adjusted to generate visuals in order to compare different camera

technical specifications. Any of the settings can also be animated within Maya to change over time. An example of a practical application for camera animation would be a camera that is continually sweeping or scanning an area.



(a) LIDAR Software



(b) Animation Software

Figure 9. Comparison of Point Cloud Visualization in LIDAR Software to Mesh Visualization in Animation Software

The animation software can also add numerous lights to the scene. In Figure 10, the circular icon close to the wall on the right side of the image represents a point light. The cone shape close to the ceiling on the left side of the image represents a spotlight. Point lights are light sources that emit in all directions (similar to light bulbs), while spotlights emit a cone of light in one direction (similar to a flashlight). Notice in Figure 10 that the spotlight is emitting a bright green light directly downward, while the point light is emitting a blue light that gradually fades as it gets further from the light source. Another notable light in Maya is a directional light, which could be used to emulate the sun. All three of these light types could be used in one scene to replicate different lighting conditions within an environment. The notable light attributes that can be customized and animated include color, cone angle (only for spotlights), intensity, falloff, and shadows.

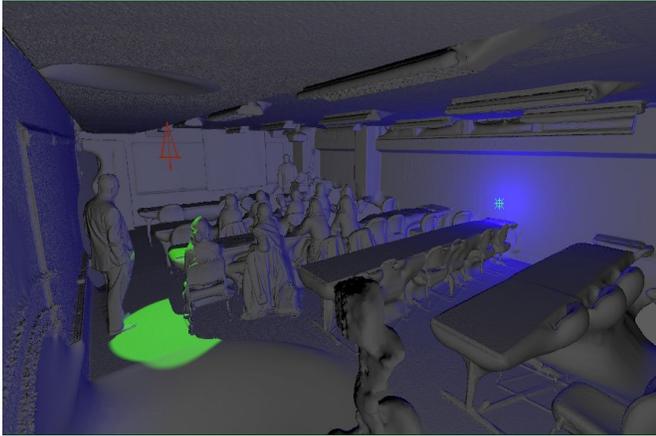


Figure 10. Visualization of Two Different Light Sources in Maya: Spotlight (green emission on left) and Point Light (purple emission on right)

The third advantage to incorporating a LIDAR scan in animation software that will be highlighted here is the animation of objects. Practically any object or attribute in Maya can be animated. This could be useful in visually demonstrating camera movement, as mentioned previously. Animation would also be useful in showing typical movement in a captured scene. For example, if a person typically walks into a room from a certain location and has limited locations to move to, based on the room setup, a digital character can be animated to walk in common paths. This can be useful in identifying blind spots for camera positions and optimizing key components of surveillance (see Figure 11). Animation can also be applied to any model, such as cars or any other object of note. Similarly, if a certain subject in the captured scene was the purpose for the surveillance, such as securing the safety of artwork, monitoring entrances and exits, or capturing movement at an intersection, this could all be simulated prior to the decision of surveillance camera location. Animations can be viewed within the software or can be exported as movie files such as *.avi and *.mov.

Establishing Surveillance Strategies

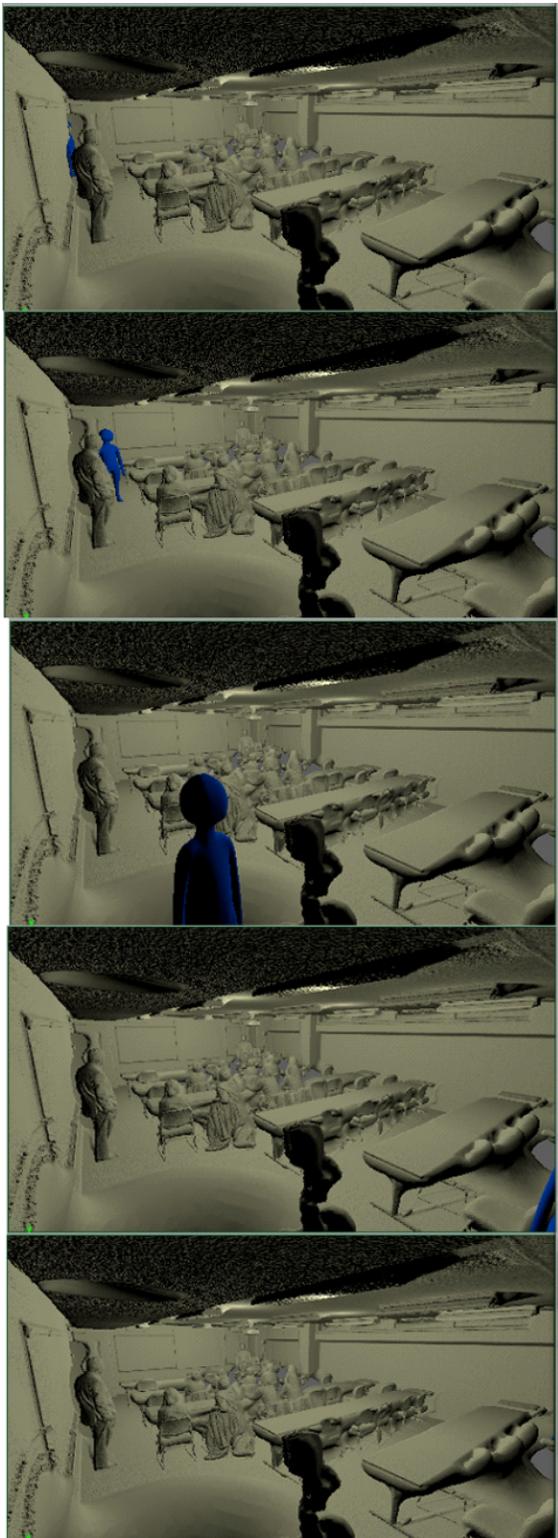
Combining the use of LIDAR scans and animation software provides extended capabilities of surveillance camera placement. This combination provides true 3D representation of the ROI and the capability of inserting a camera, with specific camera parameters, into the scene to evaluate the FOV. Figure 12 provides a 3D image showing the FOV of a camera (shown in dark pink) placed in the scene. In addition, the animation provides the capability of inserting light sources to simulate how the camera is functioning with variable lighting conditions and how moving objects into the scene may impact the field of view.

Developing the combined LIDAR and animation package for a specific location provides a true 3D tool to evaluate current surveillance camera locations. Once the 3D scene is developed in the animation software from the LIDAR scan, cameras with instrument-specific characteristics can be placed at the locations and tilt angles of existing cameras. The resulting image from each camera location can be compared to the video from the existing camera. The FOV of the camera can be modeled as a spotlight, and areas that are illuminated are those that are within the line of sight of the camera (similar to the spotlight and green area shown in Figure 10). Applying a spotlight to each camera location provides a scene showing the area of coverage of the surveillance system. Placement positions and camera mounting angles can be changed to simulate moving the camera to obtain better coverage. In addition, animated lighting and objects in the scene can be evaluated to determine the impact of each on the surveillance area.

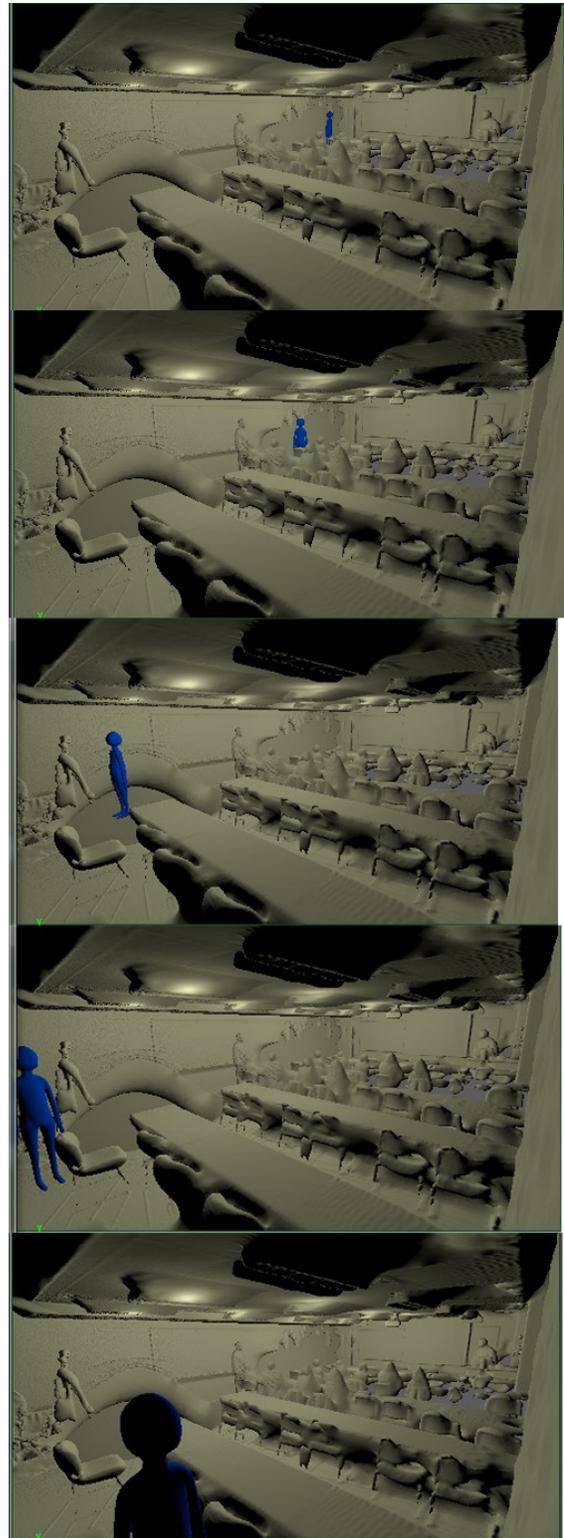
The design of a new surveillance system would initially identify areas that had the highest priority for surveillance and movement in the 3D image to locations that have the best vantage point of providing the necessary quality of video monitoring. Animated features (cars, trucks, and people) can be inserted into the image through animation to determine if the coverage is sufficient to meet security concerns. Developing a matrix of cameras within the scene provides the ability to monitor the scene from a network of camera locations during animation. Once a network of cameras is established, testing of the network as described above in the analysis of an existing system provides assessment of the camera placement.

Conclusions

This study shows that the use of LIDAR scans to establish a true 3D scene of a region of interest for a surveillance system and the integration of the 3D scene into animation software provides an effective tool to optimize the siting of security cameras. In addition, the capabilities of the animation software provide the opportunity to assess the lighting within the scene and the movement of objects that could have an impact on the line of site of the camera. This system was demonstrated to have potential in evaluating current surveillance systems and in the design of surveillance systems not yet deployed. The results of the application of this system can reduce the number of cameras within a region of interest, while increasing the coverage within the region. This results in lower costs of the system in both camera acquisition and installation, and in data storage and greater surveillance area.



(a) Appears to be Covering the Room



(b) Demonstration of the Blind Spot From the First Position

Figure 11. Camera Position



Figure 12. Camera Inserted into the Scene to Demonstrate Areas Covered in the FOV

KAREN RUGGLES is an instructor in the Mathematics and Computer Science department of DeSales University, specializing in digital art and 3D animation. She earned her BA degree from Lafayette College and MS in Digital Cinema: 3D Animation from DePaul University. Ms. Ruggles interests include 3D animation, graphic design, visualization projects, collaborative research, and 3D printing. Ms. Ruggles may be reached at Karen.ruggles@desales.edu

References

- [1] Moynihan, T. (Ed.). (n.d.). United States, Department of Homeland Security, U.S. Immigration and Customs Enforcement. Retrieved from <https://www.dhs.gov/sites/default/files/publications/privacy-pia-30-ice-smcctv.pdf>
- [2] Ghanem, B., Cao, Y., & Wonka, P. (2015). Designing Camera Networks by Convex Quadratic Programming. *Computer Graphics Forum*, 34(2), 69-80.
- [3] What is LIDAR. (2015, May 29). Retrieved from <http://oceanservice.noaa.gov/facts/lidar.html>
- [4] Reality Capture & 3D Scanning Software | ReCap 360 | Autodesk. (n.d.). Retrieved from <http://www.autodesk.com/products/recap-360/overview>
- [5] CloudCompare—Open Source project. (n.d.). Retrieved from www.danielgm.net/cc
- [6] Maya Computer Animation & Modeling Software. (n.d.). Retrieved from <http://www.autodesk.com/products/maya/overview>

Biographies

ROGER RUGGLES is an associate professor in the Civil and Environmental Engineering department of Lafayette College, specializing in surveying and water resources engineering. He is a recognized expert in flooding and has over 30 years of experience as an engineer and educator. He is a registered professional engineer in Pennsylvania and past president of the Lehigh Valley Section of the American Society of Civil Engineers. Dr. Ruggles may be reached at rugglesr@lafayette.edu

EVALUATING THE FIDELITY OF A SCALE MODEL HEXAPOD MOTION SIMULATOR

Christopher Lawrence, Schmidt Peterson Motorsports; Andrew Borme, Indiana University Purdue University Indianapolis;
Pete Hylton, Indiana University Purdue University Indianapolis

Abstract

Engineers have always endeavored to produce mathematical models of the systems on which they work. The advantages to having accurate mathematical models that represent real-life systems are numerous, and include reduced development time/cost, furthering system optimization, a deeper understanding of the system, predicting behavior in untestable circumstances, and understanding key performance indicators. In a similar vein, simulators attempt to replicate real-life events or experiences, based on underlying mathematical models that drive physical and/or visual stimuli. A key point is that simulators rely partly on human perception to determine the accuracy of the simulation experience. The use of a more complex simulator, as opposed to purely mathematical models, has some advantages specifically for use in areas where humans are integral to the overall functionality of the product. Prevalent examples of this are in the development of high-performance human-operated machines, such as airplanes and automobiles. In these cases, the performance of the machine is intertwined with the human, and the quality of the human-machine interface can be important. In particular, when the engineer is seeking maximum performance, such as in the case of jet fighter aircraft or racing cars, the use of a simulator can allow for performance gains and a deeper understanding of how the human and machine are interacting. In addition, simulators can provide benefits to the human in the form of low-risk training and observation.

A challenge with vehicle simulators that includes the driver is understanding how the platform stimuli correlate with the real stimuli experienced by the driver in racing conditions. If the platform motion stimuli are incorrect or unrealistic, the driver will either adapt to driving the vehicle based on false inputs, or be unable to interpret what the vehicle is doing, and can even lead to the driver becoming physically ill. Thus, it is important to understand how a human perceives the motion of the simulator. This study was supported by the driving simulator developed at Dallara Automobili. Dallara is an Italian company that designs and produces racing chassis for a variety of racing series, notably IndyCar, Indy Lights, and Formula 3. They also have a pair of identical driving simulators, one located in Varano de Melegari, Italy, and one located in Speedway, Indiana. IN this paper, the authors discuss the design, analysis, test-

ing, and evaluation of a scale model of a hexapod motion platform. The platform was instrumented such that the platform's motions could be measured and compared to the lap data provided from the Dallara simulator. These motions were evaluated with and without the use of human perception models that describe how humans perceive motion in quantitative terms.

Introduction

Lap time simulation offers a significant advantage to the motorsports industry. Simulation allows for estimation of vehicle performance around a circuit, which is of critical importance, since that is the singular measure of performance. Nearly unlimited optimization is available in a variety of areas, with very little expense or risk [1]. Simulators also offer the advantage of quick results when compared to a track test. Pais et al. [2] and Mulder noted that, "The advent of car simulators has improved time and cost-effectiveness, while allowing better control and repeatability of the experimental conditions. Furthermore, simulators offer a myriad of possible scenarios while guaranteeing the driver's safety."

A simulator offers the ability to drastically increase track testing time, which can be restricted by budget or rule limitations. A simulator offers a much lower risk to the driver, and requires a smaller crew. It may negate travel costs entirely, in addition to eliminating wear and tear on the car, and running costs such as fuel and tires. It can allow for more risky car setups, or the experimentation with virtual models of components that do not yet exist. It allows for near instantaneous car setup or track changes. Also, Toso and Moroni [3] found that, "Professional driving simulators can be successfully exploited to shorten the traditional design-prototype testing-production process relative to a new racecar." This is beneficial to reducing lead time, development costs, and allowing for preliminary evaluation before car construction begins. In particular, fundamental parameters can be easily changed during simulation.

However, simulators are not perfect. A simulator will always be an approximation of reality. The model inputs to the simulator determine the accuracy, allowing the possibility of false conclusions. For a useful simulation, a wealth of data about the vehicle must be known. Even with a very

capable motion platform, the motion and information fed to the driver will not be as good as the physical car and, in some cases, can be misleading or even make the driver sick. Some drivers adapt to the simulator experience better than others, whereas some adapt to driving based on unrealistic stimuli. A typical racing team utilizing a simulator must have a depth of knowledge of the vehicle being simulated and a realistic expectation of the quality results. However, they may not understand how the simulator dynamics and cueing algorithm are stimulating the driver, and this may directly affect the results. By instrumenting the motion platform and comparing the accelerations produced by the simulated vehicle model, the limitations become apparent. If the human perception models are also integrated, driver feedback can be aligned with platform and cueing algorithm strengths and weaknesses. This knowledge allows for a better expectation of the quality of the simulator results, and reduces the chances of false conclusions.

Including the driver in simulations has significant potential advantages, but also significant challenges. The potential advantages consist of increasing simulator accuracy, identification of problems, and understanding car stability issues. Also, by including the driver, subjective feedback from the driver can be gathered, the driver may improve his/her skill, and the performance of the combined driver and car can be ascertained. Some challenges to including the driver are: 1) complicated physical hardware is required; 2) it is significantly more expensive to produce a realistic simulation; and, 3) it is necessary to have a motion platform and cueing algorithm that creates a good illusion of self-motion. When a human driver is included in the simulation, driver inputs are measured. These inputs often do not result in the vehicle being at the limit behavior, so there is no optimization involved here to find the fastest lap time, or best braking point, as the driver is choosing those. This also demonstrates the need for a professional driver when motorsport simulations are being performed. As in real life, the driver must be able to extract performance out of the vehicle to provide value to the simulation effort. In simulation, a single or many real components may be tested. In this case, one real component being tested is the driver. Rather unique to driving simulators, the vehicle mathematical model is also a test subject. In another light, the simulator itself, as a collection of motion stimuli, visual stimuli, and hardware, is a test subject. As mentioned before, the driver makes decisions based on the quality and quantity of visual and motion inputs provided. The interaction between the driver and simulator is of critical importance [4].

In the development and use of driving simulators, there will always be error between the stimuli of the actual vehicle and the stimuli of the platform. If the stimuli are incorrect, the simulator is not perceived as an exact copy of reali-

ty to the driver, and can allow for false conclusions. To understand this error, a scale model of a hexapod simulator was designed, built, and tested in this current study to compare the driver stimuli. The platform stimuli were limited by many factors, which included the platform motion envelope, actuator speed and force, and the cueing algorithm. The particular stimuli that were considered were the platform and vehicle motions, specifically the linear accelerations and angular velocities in X, Y, and Z.

Simulation Fidelity

Simulation involves both objective and perceptive fidelity. Objective fidelity is achievable in terms of the mathematical model of the vehicle. Fields such as suspension kinematics, aerodynamics, tires, and vehicle dynamics are well researched, and the car can be modeled within a computer with relative ease [5-9]. Other objective fidelity requirements are an accurate track model. A track laser scan is preferable, if not required, to give correct road profile and roughness [3]. A challenge for creating objective fidelity is replicating the motions of the vehicle utilizing a motion platform. The rapid onset of lateral and longitudinal acceleration, and the rapid change in acceleration directions during successive maneuvers may present a problem. There exists a major compromise in motion platform selection. Smaller, lighter platforms have better response, but larger platforms have higher sustained acceleration and displacement [2]. In any case, it is somewhat unrealistic for a simulator to exactly replicate all accelerations of the vehicle. This necessitates a motion cueing algorithm to determine what motions are desired and possible to replicate.

Perceptual fidelity is achievable in terms of the visual, auditory, and physical vehicle controls. There are advanced graphics packages available, which support driving simulation. Depending on the simulator, the physical user interface can be as accurate as needed. The Dallara simulator features an actual carbon tub within the simulator, including the correct steering wheel, display system, and various controls, such as anti-roll and brake bias adjustments. Automotive and flight simulators also often have an actual cockpit or interior of the vehicle. The sound generation quality does not have a large impact on perceptual fidelity, though its absence is not suggested. A challenge for creating perceptual fidelity is to create an accurate illusion of self-motion. The inability to create a consistent, convincing illusion of self-motion is the primary factor for simulator sickness, where the driver may become disoriented or physically ill. In a study by Pais et al. [2], drivers preferred a motion cueing algorithm that produced no false cues, followed by no motion cueing, followed by a motion cueing algorithm that occasionally had false cues.

Human Perception Transfer Functions

The methods of perception considered in this study were the linear accelerations and angular velocities. The physical human body organs that sense these parameters are the otoliths and the semi-circular canals, respectively. Their transfer functions and motion thresholds were examined by Telban and Cardullo [10]. The transfer functions of the otolith organs are needed to create a transfer function between actual lateral/longitudinal acceleration and sensed lateral/longitudinal acceleration. The transfer functions and perception thresholds were found using Equations (1) and (2):

$$\frac{f_{sensed}}{f_{actual}} = \frac{G_{oto}(K_{oto})(s + A_0)}{(s + B_0)(s + B_1)} \quad (1)$$

$$K_{oto} = K\tau_1\tau_2/\tau_L \quad (2)$$

where,

$$G_{oto} = 0.0625 \text{ m/sec}^2 \text{ for X motion (longitudinal)}$$

$$G_{oto} = 0.0569 \text{ m/sec}^2 \text{ for Y motion (lateral)}$$

$$K = 0.4$$

$$\tau_1 = 5$$

$$\tau_2 = 0.016$$

$$\tau_L = 10$$

$$A_0 = 1/\tau_L$$

$$B_0 = 1/\tau_1$$

$$B_1 = 1/\tau_2$$

The transfer functions for the semi-circular canals were needed to create a transfer function between the actual yaw, pitch, and roll and the sensed yaw, pitch and roll. The transfer function used by Telban and Cardullo [10] is given by Equation (3):

$$\frac{\omega_{sensed}}{\omega_{actual}} = \frac{5.73(80)s}{(1 + 80s)(1.573s)} \quad (3)$$

The motion thresholds presented by Telban and Cardullo were roll = 3.0 deg/sec, pitch = 3.6 deg/sec, and yaw = 2.6 deg/sec.

Methodology

To fully understand objective and perceptual fidelity, a motion platform with configurable control systems, cueing algorithm, and measurement system was needed. Figure 1 shows a desktop-sized motion platform that was modeled in MATLAB and designed in SOLIDWORKS. Figure 2 shows the manufacture and testing of the platform.

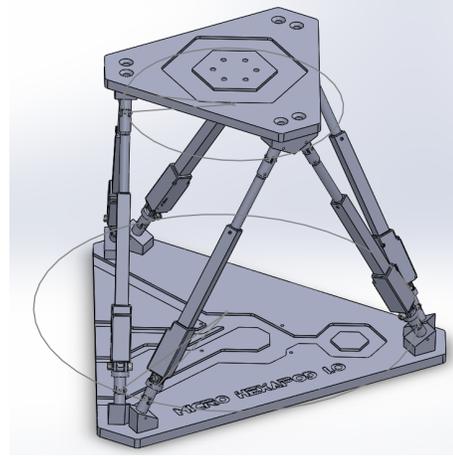


Figure 1. Model of the Mini-Simulator

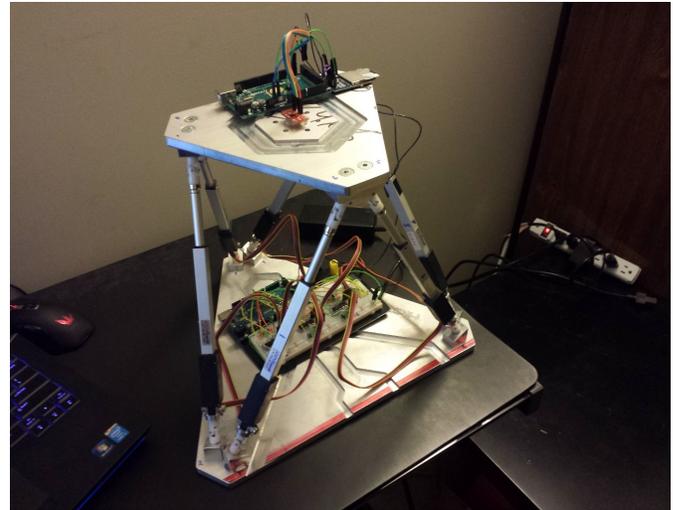


Figure 2. Constructed Mini-Simulator

The motion platform was selected to be a Stewart platform design. A Stewart platform, also known as a hexapod, is a motion platform consisting of two plates connected by six legs with adjustable lengths (actuators). One of the plates is the base, the other is the motion platform. This results in a platform with six degrees of freedom (DOF): 3 DOF in translation and 3 DOF in rotation. Stewart platforms were originally designed to be motion platforms for aircraft simulators, but are now used for driving simulators, machine tool applications, pick-and-place robots, and other applications. For the application of a driving simulator, total motion envelope and maximum platform accelerations are prioritized over platform stiffness and platform positional accuracy [11]. Stewart platforms are classified as parallel robots, where the platform motion in all six DOF is determined by all six actuators in all cases. For example, all six actuators need to move to generate a yaw angle. The benefit

of a parallel robot is that the inverse kinematics are relatively simple. The downside is that the motion envelope is highly interconnected between the DOFs. Linear displacement in one direction limits possible roll, pitch, yaw, and linear displacement in the other two directions [12-14]. The Dalara simulator uses a Stewart platform concept, as does the mini-simulator developed for this current project.

This hexapod configuration was selected due to its freedom for selecting motion cueing algorithms, simplistic design, simple inverse kinematics, and common use in the simulator industry. The scale of the platform was limited by actuator availability and budget. The linear actuators selected had the following specifications:

Stroke: 100 mm
Maximum Speed: 32 mm/s
Measurement method: 18 k Ω linear potentiometer
Positional accuracy: 0.4 mm
Backlash: 0.2 mm

Additional mechanical components selected were nylon U joints. Also, 6061 Aluminum was chosen for construction of the base and top plates of the platform, due to its low cost and easy machinability. Before a detailed CAD design was started, the kinematics of the platform were studied using MATLAB Simulink, specifically with SimScape SimMechanics. With the given actuator constraints of minimum and maximum lengths, and the additional length of the universal joints considered, the platform top and bottom radiuses were chosen to balance high platform linear travel in X and Y with high angular displacements in X and Y. This was to allow for the possibility of evaluating tilt coordination, where high angular displacements in X and Y (pitch and roll) were needed when compared to a more typical algorithm, where high linear displacements in X and Y (longitudinal and lateral acceleration) would be needed. The design was loosely based upon a scaled version of the Dalara driving simulator.

The platform also needed an adequate control system to produce the desired motions in the platform. The control system chosen was a standard closed-loop PID controller, implemented using an Arduino Mega 2560 and a series of Texas Instruments SN754410 H-bridge motor drivers. Initially, the PID control was implemented using the external driving capability of MATLAB Simulink. This allowed the model to be created and deployed to the Arduino Mega 2560 without the need to directly interact with the code. This allowed for easy troubleshooting, tuning, and model updates, as it allowed for live updates of the input signals and control system gains. PID gains were tuned by comparing responses to step and sine inputs, and resulted in the following controller:

- Proportional gain: 5
- Integral gain: 0.1
- Derivative gain: 0

Marginally faster responses were noted at higher proportional gain values, but the lower gain value of five was chosen to minimize platform oscillation with a constant signal (no motion). After the errors were worked out and the control system tuned, MATLAB Simulink would not allow for parameters to be directly input using the USB serial connection or specify that certain variables to be stored anywhere other than RAM. This prevented anything other than simple or periodic functions to be implemented on the platform, due to program size that would exceed the storage capability of the Arduino.

The cueing algorithm was chosen to be a hybrid of several of the cueing algorithms commonly used in simulators. Tilt coordination was used for X and Y (longitudinal and lateral acceleration). Due to the lack of Z acceleration data, there was no cueing input for the Z direction. For yaw, the BSS cueing algorithm was implemented; the pitch and roll angles of the car were added to the pitch and roll cues that were created from the tilt coordination algorithm. These algorithms were chosen for their demonstrated effectiveness and simplicity from the studies previously discussed. The gains for the tilt coordination algorithm were varied in order to demonstrate the effect of the cueing algorithm on the motion of the platform. Additionally, the platform inverse kinematics were implemented [11, 14]. The leg lengths were then converted into a 10-bit format for use in the Arduino PID control, where a value of 0 indicated minimum leg length, and a value of 1023 indicated maximum leg length (100 mm).

The platform dynamics were measured to allow for evaluation of the platform motion. The motions of particular interest were the platform accelerations in the X, Y, and Z directions, as well as the yaw, pitch, and roll rates. Because the input data were limited to accelerations in X and Y, as well as yaw rate, those channels in particular were the most important data to acquire for comparison purposes. To measure the platform motion, an integrated 3-axis accelerometer and a 3-axis gyro were selected. The data from these sensors were logged by another Arduino Mega 2560 and stored on an external 16 GB micro-SD card, which was attached to the Arduino. The accelerometer and gyro combination interface used a digital serial I2C interface, and the SD card reader interfaces used a digital serial SPI interface. The measurements were calibrated and converted into m/s^2 and deg/s , respectively, and then written on the SD card in CSV format for later offload. The system logged all six of the parameters and wrote to the SD card at 100 Hz.

There were several unexpected limitations or weaknesses that developed during the course of the study.

- No Z (vertical) acceleration data were available.
- Actuator speed limited the platform's linear and angular acceleration.
- Arduino memory limited the cueing algorithm from running at 50 Hz for full a lap of data.
- Platform measurement gyro used gains from a data sheet, though offsets were measured and accounted for.

Methodology

The motion platform was tested using a simulated racecar data set from the Dallara simulator. This data set was a lap of the Indianapolis Motor Speedway Road Course, and driven by an experienced IndyCar driver using a mathematical model of the 2015 Indy Lights racecar. The data set included all of the necessary parameters to implement the cueing algorithms studied here, including vehicle accelerations, chassis yaw, pitch, and roll, as well as body side slip angle and yaw rate. Note that due to limitations of the actuators, accelerations were scaled. The human perception models show that humans are most sensitive to accelerations and angular velocity and acceleration. Thus, the motion platform accelerations and angular velocities and accelerations were directly compared to the mathematical vehicle model accelerations and angular velocities and accelerations. Additionally, the perceptual fidelity of the motion platform was studied using the transfer functions presented by Telban and Cardullo [10]. The accelerations and angular velocities were compared, as sensed by the human perception models. For example, if the physical racecar on the track produced a motion that the driver cannot sense, then it was not a problem if the platform did not reproduce this motion, despite the obvious discrepancy in objective fidelity.

The tilt translation cueing algorithm included simulator pitch and roll, based on lateral and longitudinal G demands, summed with raw pitch and roll values, based on a one-to-one representation of chassis pitch and roll. The X and Y translation cues were driven by lateral acceleration of the vehicle, and limited by the maximum displacement of the platform. The gains used were:

- Pitch gain: 1/100
- Roll gain: 1/100
- Lateral gain: 1/750
- Longitudinal gain: 1/750

Figure 3 shows a comparison of the raw data. The red lines indicate the motion perception thresholds. That is, any values recorded between the red lines would not be sensed

by the average human. An interesting feature of the results is how well the acceleration traces matched, with only a few exceptions. From the lateral acceleration graph of Figure 4, there were several places where the motion platform lagged behind the actual car in generating lateral acceleration. This was due to the limitations of the platform, where at these points one or more of the actuators were at maximum speed. This would be a good issue to fix by increasing actuator speed, however the platform lag was not causing any false cues because the accelerations were still in the correct directions. Another possible way to correct this error would be to reduce the overall magnitude of motion, so that maximum platform acceleration would be less.

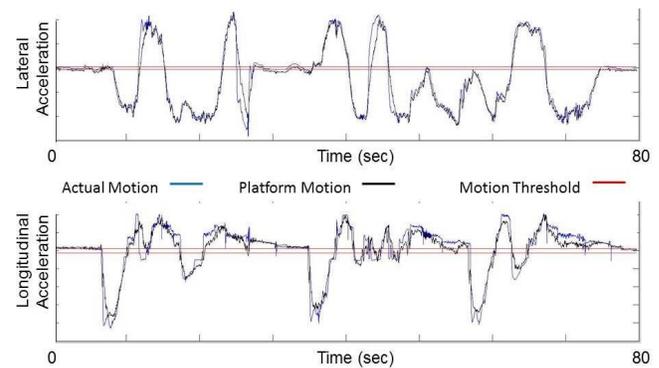


Figure 3. Lateral and Longitudinal Acceleration (m/sec^2) versus Time (sec)

[Note: Acceleration amplitudes are restricted data and were omitted.]

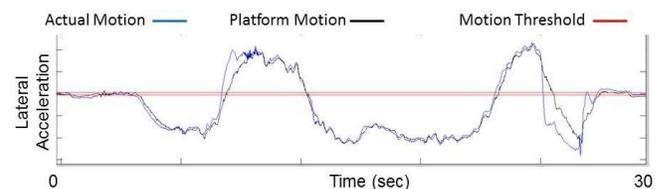


Figure 4. Detail View of Lateral Acceleration (m/sec^2) Error versus Time (sec)

[Note: Acceleration amplitudes are restricted data and were omitted.]

Figure 5 shows the comparison of platform yaw rate with actual vehicle yaw rate, and body side slip rate (BSS rate) versus actual. The cueing algorithm was chosen to follow the BSS angle precisely, thus the BSS rate and the yaw rate of the platform should be close. The platform cannot achieve the yaw rates of the vehicle due to platform limitations, but it can be seen that any change in yaw rate slope (yaw acceleration) was mimicked using the BSS algorithm. Physically, the BSS angle was the slip angle of the chassis of the car, or the angle between where the chassis was pointing and where the chassis was heading. A spike in BSS

rate indicates a spin or massive understeer, depending on the sign and the direction of the corner. There were several spots where the platform yaw rate was opposing the actual BSS rate; Figure 6 represents this as a false cue.

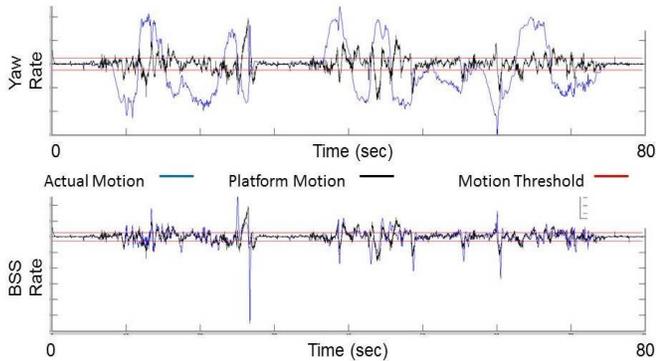


Figure 5. Yaw and BSS Velocity (deg/sec) versus Time (sec)
[Note: Amplitudes are restricted data and were omitted.]

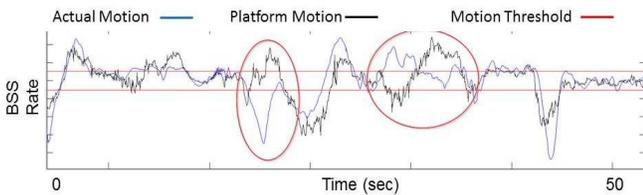


Figure 6. BSS Rate (deg/sec) versus Time (sec). Detailed View of False Cues
[Note: Amplitudes are restricted data and were omitted.]

It can be observed that nearly all of the pitch and roll inputs from the actual car were not perceived by the average human (i.e., they fall between the red lines representing the threshold). However, the pitch and roll rates of the platform were much higher, due to the tilt coordination algorithm rolling and pitching the platform to create the lateral and longitudinal acceleration cues. Clearly, this algorithm did not pitch or roll the platform slow enough to stay under the perception thresholds of the driver. This showed the basic compromise in the tilt coordination algorithm. The pitch and roll rates can be reduced to eliminate the false pitching and rolling cues, but then the magnitude of the acceleration cues would decrease and have delayed onset.

The graphs in Figures 7 and 8 add the human perception transfer functions discussed thus far to show how well humans perceive the motion of the racing car and platform, respectively. The figure was a direct comparison of both the lateral acceleration and yaw rate of the platform, before and after the human perception transfer functions, to illustrate the differences between reality and perception. It can be seen that the transfer functions do not show drastic differences between the actual and perceived motions. The lateral

acceleration motions were mostly perceived to be the same, with a reduction in magnitude for lower acceleration values, and an increase in magnitude for high changes in acceleration (high jerk). This might indicate that an algorithm for lateral and longitudinal motion based on matching jerk of the platform and vehicle may produce good simulator results. The yaw rate compared to the perceived yaw rate showed attenuation of the signal for high yaw rates. It also showed a high response to yaw acceleration, but a decreasing response to sustained constant yaw rate. This also indicated that an algorithm based on yaw acceleration might produce good simulator results. It also showed why the BSS algorithm had good success, due to BSS angle representing the changes in yaw angle. Based on the perceived yaw rate, the platform may be able to achieve a one-to-one algorithm.

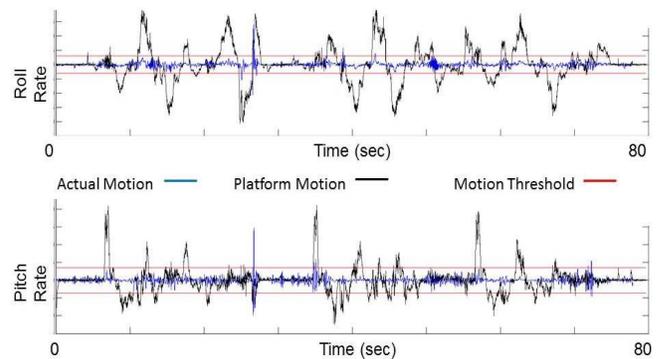


Figure 7. Pitch and Roll Velocity (deg/sec) versus Time (sec) with Human Perception Transfer Functions
[Amplitudes are restricted data and were omitted.]

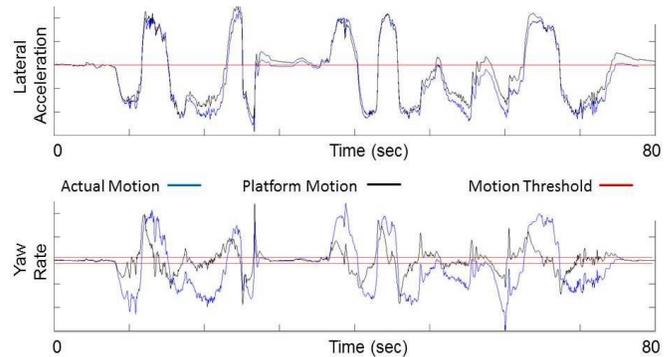


Figure 8. Lateral Acceleration (m/sec²) and Yaw Rate (deg/sec) versus Time (sec). Comparison between Measured and Perceived Values from the Actual Vehicle
[Amplitudes are restricted data and were omitted.]

The graphs of Figures 9-11 echo Figures 3, 5, and 7, where accelerations and angular velocities from the vehicle and simulator were compared, but with all of the data processed using the given human perception transfer functions to illustrate the human perception of the motion.

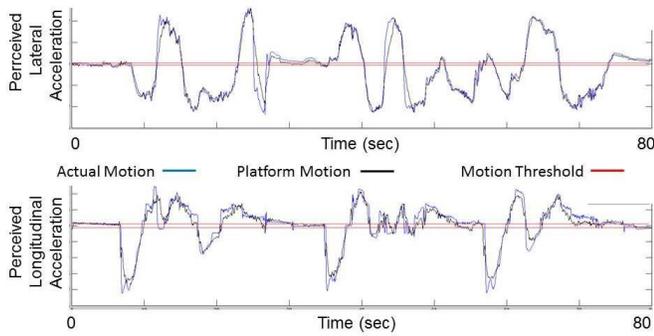


Figure 9. Sensed Lateral and Longitudinal Acceleration (m/sec²) versus Time (sec)
[Amplitudes are restricted data and were omitted.]

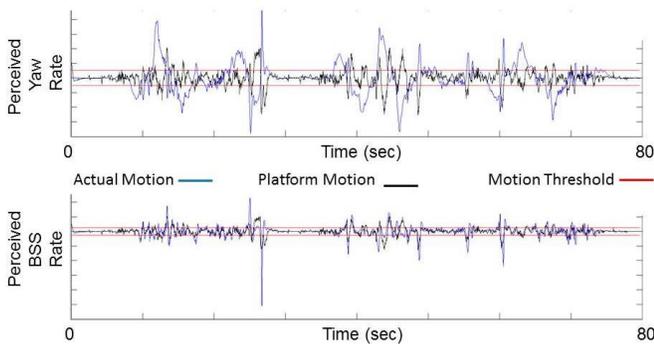


Figure 10. Sensed Yaw and BSS Velocity (deg/sec) versus Time (sec)
[Amplitudes are restricted data and were omitted.]

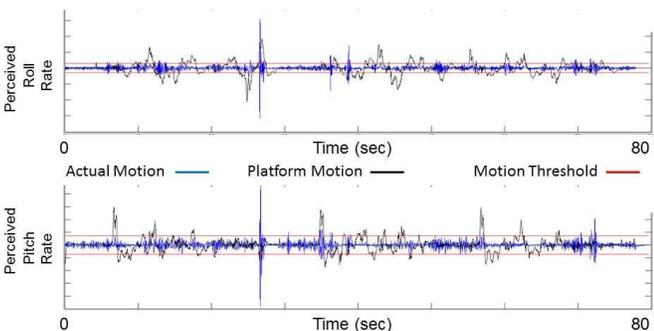


Figure 11. Sensed Pitch and Roll Velocity (deg/sec) versus Time (sec)
[Amplitudes are restricted data and were omitted.]

The perceived acceleration traces were very similar to the actual acceleration traces, and the conclusions about the algorithm effectiveness were similar. There were still a few areas on the lateral acceleration trace where the platform speed limitation impacted the platform acceleration preventing better cueing. The reduction in sensitivity to yaw has benefits in reducing the perceived magnitude of the false cues noted above. Figure 12 shows that there was only one

major false cue and one minor false cue observed in the BSS rate graph. Overall, more of the cues were below the perception threshold, and could be filtered out. The perception modeling had the biggest impact on the large false pitch and roll cues produced by the simulator, and it can be seen how much less magnitude of the pitch and roll motions were perceived. However, there were still 18 false perceived roll cues and 18 false perceived pitch cues.

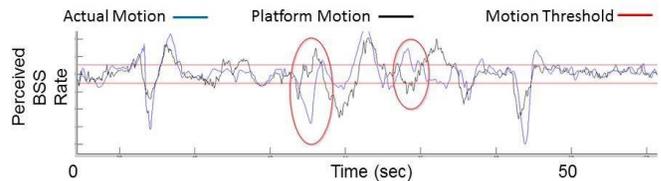


Figure 12. BSS Rate (deg/s) versus Time (s). Detailed View of False Cues
[Amplitudes are restricted data and were omitted.]

Conclusions

A scaled down model of a hexapod driving simulator was designed, built, tested, and evaluated using lap data. The platform kinematics were studied, actuators and electronics selected, key dimensions chosen, and the platform was constructed. A tilt coordination cueing algorithm was created in MATLAB and implemented using Arduino code. The platform linear accelerations and angular velocities were measured. Based on the platform accelerations and angular velocities, the cueing algorithm demonstrated very good acceleration results and moderately good yaw results. The pitch and roll results showed significant false cues, which may be a detriment to driver performance and realism in the simulator. When these results were viewed after being transformed by the human perception models, the acceleration results were very good. The false cues in the yaw direction were minimized, but still existed. To a much higher degree, the pitch and roll false cues were minimized, but there were still pitch and roll motions above the driver perception threshold that were not present in the lap data. This cueing algorithm could be improved by reducing the gain for the tilting of the platform and increasing the gain for the translation of the platform. This would reduce the pitch and roll false cues, at the expense of reducing the linear acceleration magnitudes.

The platform developed in this study was limited by the maximum speed of the actuators used. If faster actuators could be developed or purchased, a more translation-biased cueing algorithm could produce high linear accelerations based on platform motion alone. Additionally, based on the human perception results from this simulation, a translation based algorithm that correlates vehicle jerk with platform

jerk may produce good results, especially in the case of a fast platform with a limited-motion envelope. Also, based on the perceived yaw, an algorithm that attempts to do a one-to-one perceived yaw seems possible, and would provide the driver with more information about car balance, which is a traditional simulator shortcoming.

Acknowledgements

The research conducted in this study was funded by a two-year grant from the Indiana Economic Development Corporation (IEDC), in partnership with the Purdue School of Engineering and Technology at Indiana University-Purdue University, Indianapolis (IUPUI) and Dallara. Without the support of the IEDC and Dallara's collaborative efforts with IUPUI, this project would not have been possible.

References

- [1] Siegler, B., Deakin, A., and Crolla, D. (2000). Lap Time Simulation: Comparison of Steady State, Quasi - Static and Transient Racing Car Cornering Strategies. *SAE Technical Paper 2000-01-3563*, doi:10.4271/2000-01-3563.
- [2] Pais, A., Wentink, M., Paassen, M., & Mulder, M. (2009). Comparison of Three Motion Cueing Algorithms for Curve Driving in an Urban Environment. *Presence: Teleoperators and Virtual Environments*, 18(3), 200-221.
- [3] Toso, A., & Moroni, A. (2014). Professional Driving Simulator to Design First-Time-Right Racecars. *SAE Technical Paper 2014-01-0099*, doi:10.4271/2014-01-0099.
- [4] Waeltermann, P., Michalsky, T., and Held, J. (2004). Hardware-in-the-Loop Testing in Racing Applications. *SAE Technical Paper 2004-01-3502*, doi:10.4271/2004-01-3502.
- [5] Milliken, W., & Milliken, D. (1995). *Racecar Vehicle Dynamics*. Warrendale, PA: Society of Automotive Engineers
- [6] Pacejka, H. (2012). *Tire and Vehicle Dynamics (3rd ed.)*. Waltham, MA: Butterworth-Heinemann.
- [7] Gillespie, T. (1992). *Fundamentals of Vehicle Dynamics*. Warrendale, PA: Society of Automotive Engineers.
- [8] Guiggiani, M. (2014). *The Science of Vehicle Dynamics: Handling, Braking, and Ride of Road and Racecars*. New York, NY: Springer.
- [9] Sergers, J. (2014). *Analysis Techniques for Racecar Data Acquisition (2nd ed.)*. Warrendale, PA: Society of Automotive Engineers.
- [10] Telban, R., & Cardullo, F. (2005). Motion Cueing Algorithm Development: Human-Centered Linear and Nonlinear Approaches. *NASA Technical Paper NASA/CR-2005-213747*.
- [11] Gong, Y. (1992). *Design Analysis of a Stewart Platform for Vehicle Emulator Systems*. (Master's thesis). Retrieved from <http://dspace.mit.edu/bitstream/handle/1721.1/34314/26408968-MIT.pdf?sequence=2>
- [12] Garrett, N., & Best, M. (2013). Model Predictive Driving Simulator Motion Cueing Algorithm with Actuator-Based Constraints. *International Journal of Vehicle Mechanics and Mobility*, 51(8), 1151-1172.
- [13] Korobeynikov, A., & Turlapov, V. (2005). Modeling and Evaluating of the Stewart Platform. *Proceedings of the International Conference Graphicon 2005*, doi:10.1.1.189.5199.
- [14] Jakobovic, D., & Jelenkovic, L. (2002). *The Forward and Inverse Kinematic Problems for Stewart Parallel Mechanisms*. Retrieved from <http://alvarestech.com/temp/RoboAseaIRB6S2-Fiat/CinematicaHexapode.pdf>

Biographies

CHRIS LAWRENCE is a simulation engineer for Schmidt Peterson Motorsports. He received his bachelor's degree in Mechanical Engineering from Colorado State University and his master's degree from Indiana University Purdue University Indianapolis. Mr. Lawrence may be reached at chrisl@spmindyear.com

ANDREW BORME is a Senior Lecturer of Motorsports Engineering at Indiana University Purdue University Indianapolis. He received his BS in engineering from Rensselaer Polytechnic Institute and MS in engineering from California State University Long Beach. Mr. Borme may be reached at aborme@iupui.edu

PETE HYLTON was a Professor of Motorsports Engineering at Indiana University Purdue University Indianapolis. He earned his BS degree in engineering from Rose-Hulman Institute, MS degrees in engineering and math from Purdue University, and an EdD from Grand Canyon University. Dr. Hylton may be reached at phylton@iupui.edu

PUSHOVER ANALYSIS OF THE NEW SELF-ANCHORED SUSPENSION BAY BRIDGE TOWER

Abolhassan Astaneh-Asl, The University of California, Berkeley; Xin Qian, The University of California, Berkeley

Abstract

The new eastern span of the San Francisco-Oakland Bay Bridge, which opened to traffic in 2013, is a self-anchored suspension bridge with a single tower. The main cables of an SAS bridge are connected to the bridge deck instead of being connected to the anchor blocks, as is the case in traditional suspension bridges. The bridge, which was designed with only one tower with four shafts, is located in a highly seismic area in northern California. Steel shear links connect the four shafts to each other along the height. A cable saddle connects the top of the four tower shafts to each other. High-strength anchor rods connect the base of the tower to the pile cap. In this paper, the authors present the results of a series of realistic, nonlinear pushover analyses of the single tower of the SAS Bay Bridge. A detailed, nonlinear finite-element model of the main tower was constructed using shell elements available in the finite-element analysis software ANSYS. The analysis consisted of pushing the top saddle horizontally in five different directions (at 0-, 30-, 45-, 60-, and 90-degree angles with respect to the longitudinal direction of the bridge) until it collapsed. The authors focused on the behavior in the transverse direction (90°). Gravity loads were included in the pushover analysis. The results showed that local buckling of the tower shaft plates may occur relatively early in the pushover analysis, resulting in a drop in strength and a reduction in ductility. The original designers of the bridge did not take this behavior into account. Their pushover analysis of the tower, which used a model with only beam elements instead of shell elements, concluded that the only nonlinearity in the tower would be yielding of the shear links connecting the tower shafts to each other; the tower would remain essentially elastic with no local buckling. The realistic modeling and accurate analysis presented here show that this conclusion was incorrect.

Introduction

Figure 1 shows the new eastern span of San Francisco-Oakland Bay Bridge is a self-anchored suspension (SAS) bridge with a single tower. The bridge opened to traffic in 2013 and is located between two active seismic faults: the Hayward and San Andreas faults. There is a 63% probability of occurrence of an earthquake with a magnitude 6.7 or greater in the greater San Francisco Bay Area during the 30-year period from 2007 to 2032 [1].

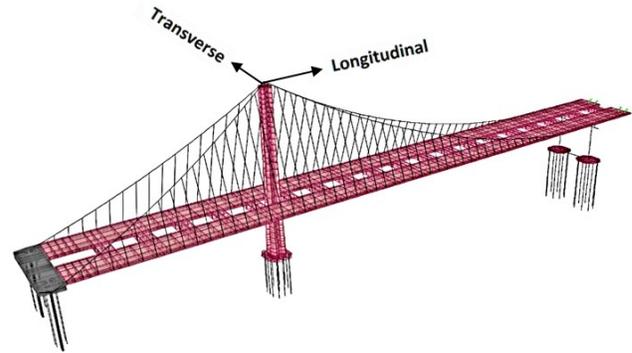


Figure 1. The SAS Bay Bridge Longitudinal and Transverse Directions

Figure 2 shows the overall structure of the new SAS Bay Bridge, and Figure 3 shows the elevation and plan of the main tower. Figure 4 shows a typical deck segment of the bridge. As can be seen in Figure 3, the 512 ft (156m) tower (at the cable intersecting point) consists of four shafts, each shaft being a pentagonal steel hollow box with vertical stiffeners and horizontal diaphragms. The steel used in the tower shaft was ASTM A709 Gr. 50 steel with a yield stress of 50 ksi (345 MPa) and an ultimate strength of 65 ksi (448 MPa). As is standard, the shafts have horizontal stiffeners every 9.9 ft (3m). I-shaped shear links connect the shafts to each other along the height of the tower and by a saddle at the top. The main cable is a single cable connecting the top of the tower to the orthotropic steel deck and which passes through the saddle.

Figure 3 shows how high-strength A354 BD zinc-coated (hot-dip), galvanized anchor rods with a yield stress of 115~130 ksi (793~896 MPa), and an ultimate strength of 140~150 ksi (965~1034 MPa), connect the base of the tower to the pile cap. The use of A354 BD zinc-coated, hot-dip galvanized anchor rods in this key traffic corridor exhibited undesirable behavior a few months before the opening of the bridge; thirty-two of the anchor rods connecting seismic shear keys to the top of the Pier E2 on the east end of the SAS Bay Bridge fractured when tightened. In this study, the authors focused on the pushover behavior of tower itself and assumed that the base plate was rigidly connected to the pile cap, or pile substructure. Previous work by Astaneh-Asl et al. [2] focused on the pushover behavior of the tower in a scenario where anchor rods fractured and could no longer resist the tension created by the bending moment at the base of the tower.

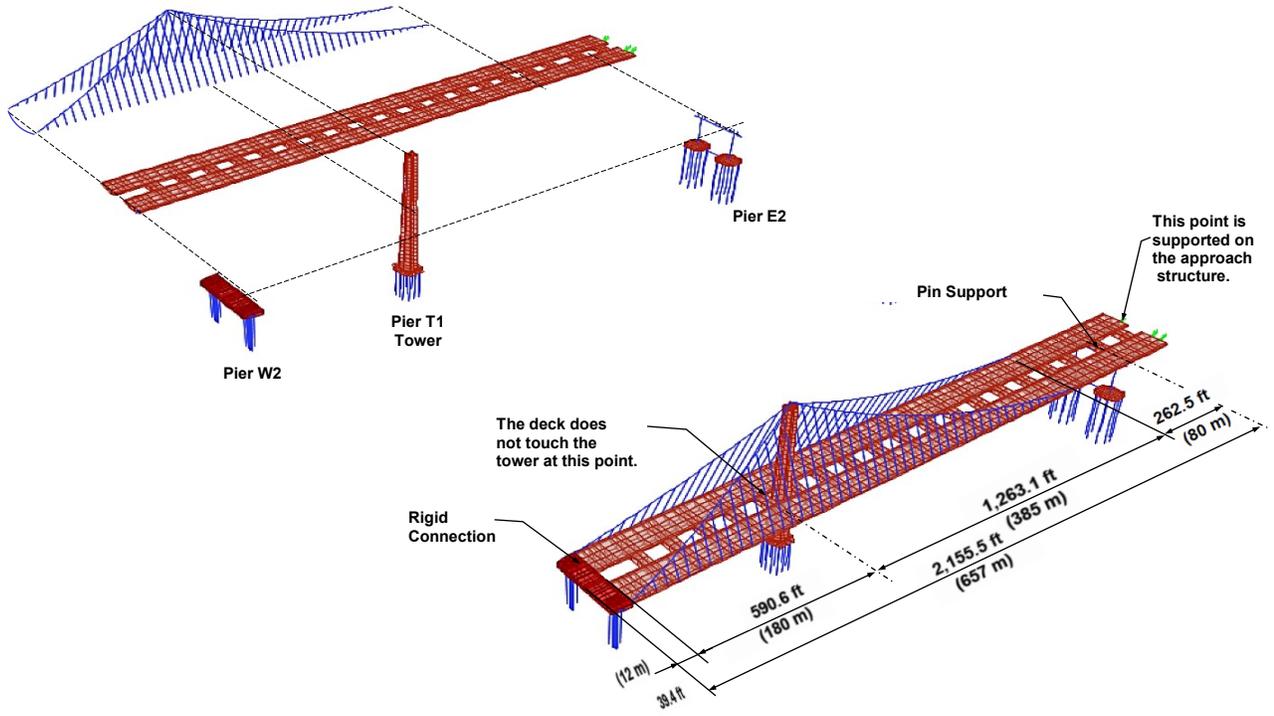


Figure 2. Disassembled and Assembled Views of the SAS Bay Bridge

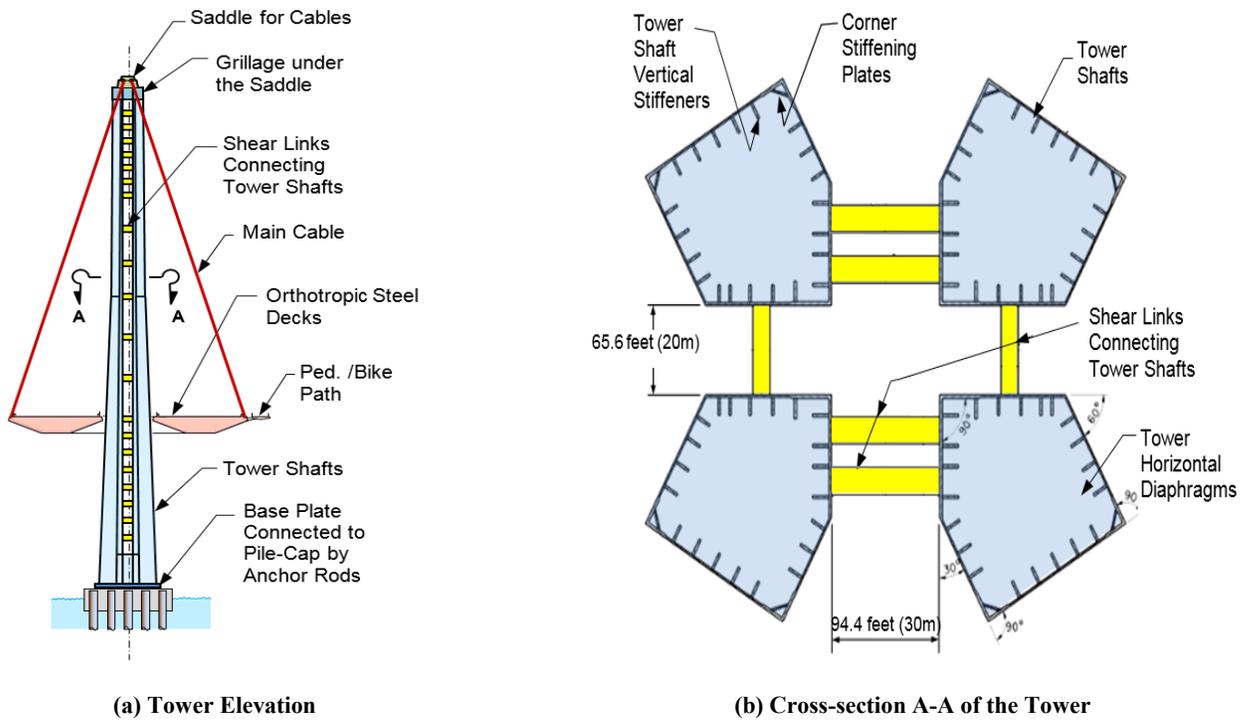


Figure 3. Elevation and Cross-Section of the Tower of the SAS Bay Bridge

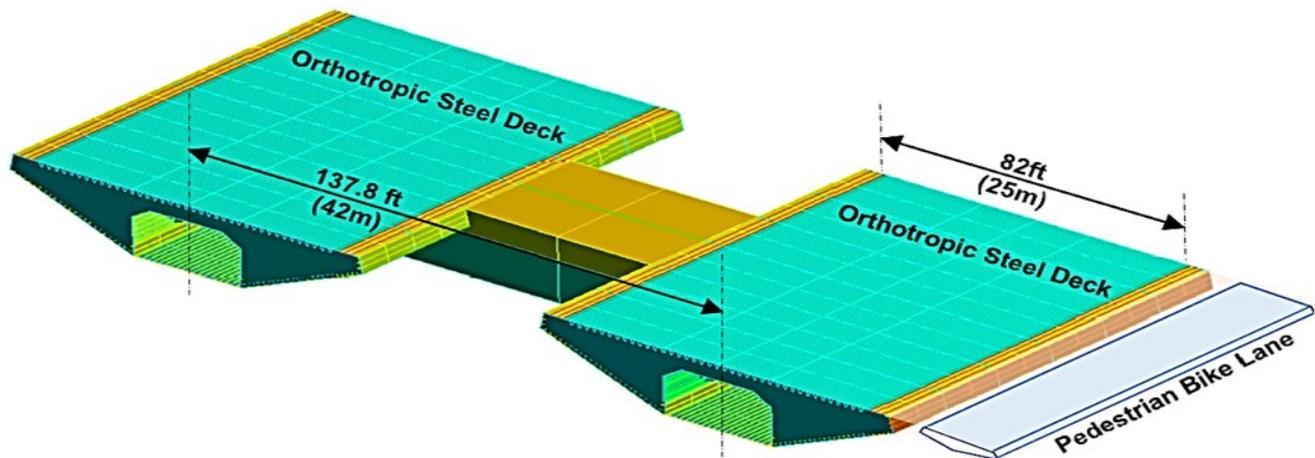


Figure 4. Typical Segment of the SAS Bay Bridge Showing Two Separate Orthotropic Decks Connected to Each Other by Transverse Steel Box Girders

Objectives

The main objective of this current study, summarized in this paper, was to establish the stiffness, strength, buckling behavior, and ductility of the as-built main tower of the SAS Bay Bridge, when pushed by a horizontal force at the top.

Background

Previous studies by the chief designers of the SAS Bay Bridge [3-6] provide information on the analysis and design aspects of the bridge. In this paper, the authors cite excerpts from those previous papers on performance criteria and expected behavior during seismic events. What follows is a discussion of the validity of the original assumptions and the accuracy of the results that were based on those assumptions. Dr. Marwan Nader, a co-author of these papers, was the Chief Engineer of Record for the SAS Bay Bridge designed by T.Y. Lin International of San Francisco. Dr. Brian Maroney was the Chief Engineer for the SAS Bay Bridge for the California Department of Transportation (Caltrans), the state agency that owns the SAS Bay Bridge. The other co-authors involved in the analysis and design of the SAS Bay Bridge, engineers from T.Y. Lin International, hereafter will be referred to as the bridge design team. The co-authors of those publications played a critical role in the analysis, design, construction, and inspection of the bridge, and were directly responsible for its design. The information contained in those publications represents the official record of the bridge design. In this current study, the authors focused on the results presented in those publications, the methodology used, and the assumptions made in the analysis and design, with an emphasis on the pushover analysis of the tower.

Performance Criteria Presented by the Bridge Design Team

Following are excerpts from the work by Nader et al. [3], the designers of the SAS Bay Bridge.

SEISMIC PERFORMANCE CRITERIA – The Bridge is designed to provide a high level of seismic performance. It is designed to resist two levels of earthquake, a functional evaluation earthquake (FEE) and a safety evaluation earthquake (SEE). After a functional evaluation earthquake, the bridge will provide full service almost immediately and there will be minimal damage to the structure. Minimal damage implies essentially elastic performance and is characterized by minor inelastic response, narrow cracking in concrete, no apparent permanent deformations, and damage to expansion joints. After a safety evaluation earthquake, the bridge will provide full service almost immediately and will sustain repairable damage to the structure. Repairable damage is damage that can be repaired with minimum risk of losing functionality; it is characterized by yielding of reinforcement, spalling of concrete cover and limited yielding of structural steel. (pp.4-5)

Expected Behavior during Seismic Events

The design criteria for the SAS Bay Bridge required that the bridge be operational almost immediately after a major earthquake [4]. According to the bridge design team, “Seismic analysis was performed using the ADINA general-

purpose finite element program. Three forms of analysis were employed: time history analysis (global model), pushover analysis and local detailed analysis” [4].

Bridge Model and Pushover Analysis used by the Bridge Design Team

In a pushover analysis used by the bridge design team, they stated that, “Pushover analysis was primarily used to evaluate ductility of critical elements and to establish failure mode sequence” [3]. The ADINA global analysis model used by the bridge design team in their pushover analysis consisted of only linear and some selected nonlinear truss and beam elements, and did not contain any shell elements [3]. Nader et al. [3] went on to say that, “The shear links between the shafts were also modeled with inelastic moment-curvature beam elements.” The base of the tower, where the four shafts are connected to each other by steel plates to form a single multi-cell tower, was modeled as a single shaft with elastic beam-column “stick” elements [4].

“The shear links, connecting the tower shafts to each other, ..., were modeled with inelastic “moment-curvature” beam elements, calibrated using the shear displacement relationship from a detailed local model” [3]. According to the bridge design team [3], the basis of design was:

... the bridge is designed based on a limited ductility design in which plastic deformations are clearly defined and predetermined. ... the bridge is designed to remain largely elastic with the exception of the east and west piers which are designed to form plastic hinges. ... The shear links between the tower shafts are also designed to yield in shear during the SEE earthquake. (p.5)

They further stated that during the seismic safety evaluation earthquake, which was used as the design earthquake, the top of the tower will move maximum horizontally 4.27 ft (1.3m) and 3.28 ft (1.0m) in the transverse and longitudinal directions, respectively. The bridge design team stated that the only inelastic areas will include: plastic hinge formation at the top and bottom of R/C Pier W2, plastic hinge formation at the bottom of R/C Pier E2, and yielding of shear links connecting four shafts of the main tower to each other. And, the pushover analysis of the main tower was performed.

... to evaluate the base shear versus top of tower displacement relationship, to optimize the design of the tower shear link and shaft, to evaluate the lateral ductility of the tower before collapse and to evaluate the ductility demands on the shear links

and tower shafts at various levels of displacement demand during an earthquake. (p.8)

In this current study, the authors demonstrated that the pushover performance of the main tower of the SAS Bay Bridge does not represent the actual behavior of the tower, when subjected to ground shaking. The main inaccuracy is that the ADINA global analysis model of the tower uses only beam elements to represent the actual steel plate members, which are shell elements. These beam elements are unable to predict the local buckling phenomenon, which is the main cause of instability in structures designed utilizing steel plates, such as the tower of the SAS Bay Bridge.

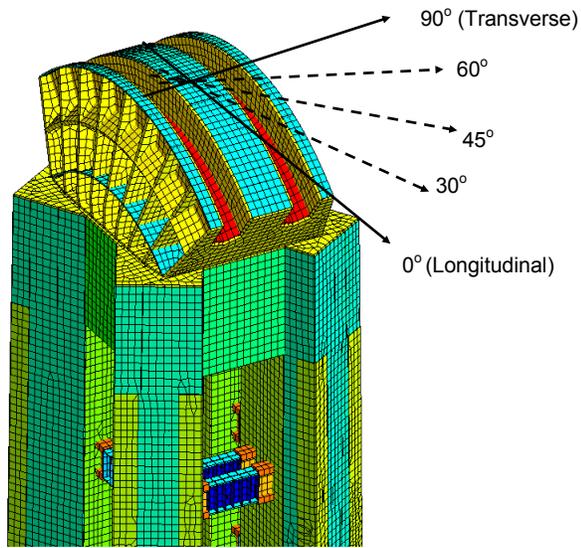
Realistic Pushover Analysis of the Tower of the SAS Bay Bridge

The analysis used inelastic shell elements capable of yielding and local buckling to represent all steel plates in the tower. The only exception was the vertical stiffeners in the tower shafts, which were modeled as beam elements.

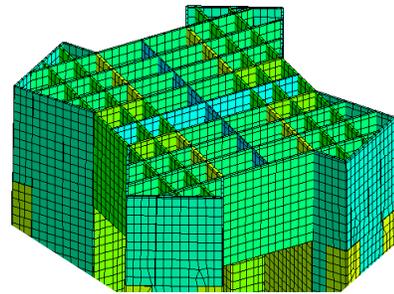
Finite Element Modeling of Tower

The general-purpose finite element software ANSYS R15.0 was used to determine the pushover behavior of the main tower. Figure 5 shows how all of the components of the main tower, except the vertical stiffeners of the tower shaft, were modeled with the SHELL181 element. This is a 4-node shell element that is suitable for linear, large deflection, and large-strain nonlinear applications. The BEAM188 element of ANSYS, which is a 2-noded linear, quadratic, or cubic 3D beam element based on Timoshenko beam theory, was used to model the vertical stiffeners of the tower shafts. The tower shaft stiffeners were attached to the tower shaft plates with bonded contact [7]. Figure 5 shows an ANSYS finite-element model of the tower. The geometric features of the tower were modeled in detail, based on the construction drawings [8].

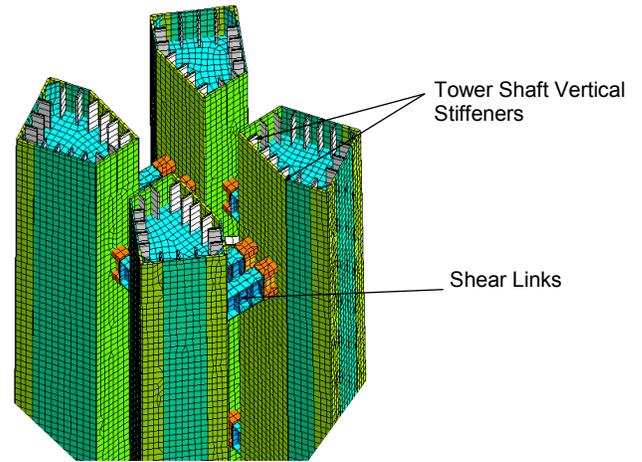
All critical structural components of the main tower, including all three types of shear links, vertical stiffeners of the tower shafts, and horizontal diaphragms inside the tower shafts, were modeled. Non-structural and architectural features such as the tower skirt at the base of the tower were not included in the model. In this current study, anchor rods were assumed to be fully functional, so the tower base was assumed to be fixed in order to prevent both displacement and rotation. Figure 2 shows that the impact of the decks on the tower shafts was not included, considering the gap between the main tower and the bridge decks.



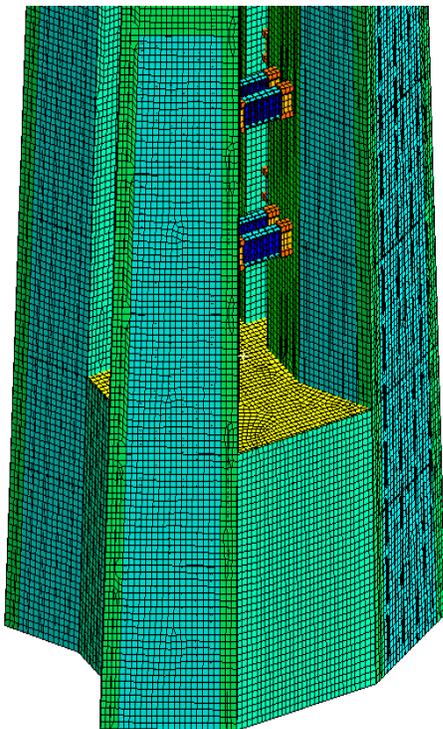
(a) Tower Saddle and Grillage



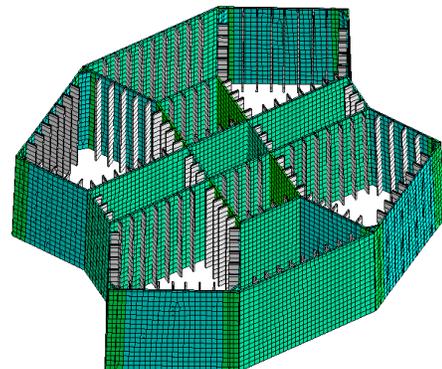
(c) Cross-section of Tower Grillage



(d) Cross-section of Tower with Four Shafts



(b) Tower Base



(e) Cross-section of Tower Base

Figure 5. Finite Element Model of the SAS Bay Bridge Tower with Details of the Cross-Sections

Pushover of the tower was achieved by a nonlinear, large-deflection, multi-step static analysis, including both material and geometric nonlinearities. In addition to the self-weight of the tower itself, concentrated forces were applied at the tower tip saddle to represent the vertical and horizontal components of the cable forces acting on the saddle. Figure 1 shows the cable forces were obtained from a gravity analysis of the entire bridge model in SAP2000. With the pre-stress from gravity effects in place, incremental horizontal displacements were then applied using the displacement controlled iteration algorithm. The horizontal displacements were applied at the cable saddle-groove location.

Figure 5(a) shows that pushover analyses were conducted in five different directions: longitudinal (0°), transverse (90°), and 30, 45, and 60° from the longitudinal axis of the bridge, with the focus on the behavior in the transverse direction, which is normally the most critical direction. The material for all components of the steel tower was Gr.50 steel with a yield stress of 50 ksi (345 MPa). The only ex-

ception was the rigid connection plates of the shear links to the tower shafts, which were Gr.70 steel with a yield stress of 70 ksi (485 MPa). The steel was modeled using a bilinear kinematic hardening material model with an initial elastic modulus equal to 29,000 ksi (200 GPa), a Poisson ratio of 0.3, and a strain-hardening ratio of 1%.

Results of Transverse Direction Pushover Analysis

Pushover behavior of the tower in its transverse direction is normally more critical than the other directions. Figure 6 shows the pushover curves in the transverse direction, which indicate that the tower yielded gradually and then the lateral load resistance dropped relatively quickly after the applied pushover force reached its maximum value. There was no pronounced yield plateau on the pushover curve. Based on the results of the analysis, several phases of behavior were observed, and three important points were identified as points Y, U, and D in Figure 6.

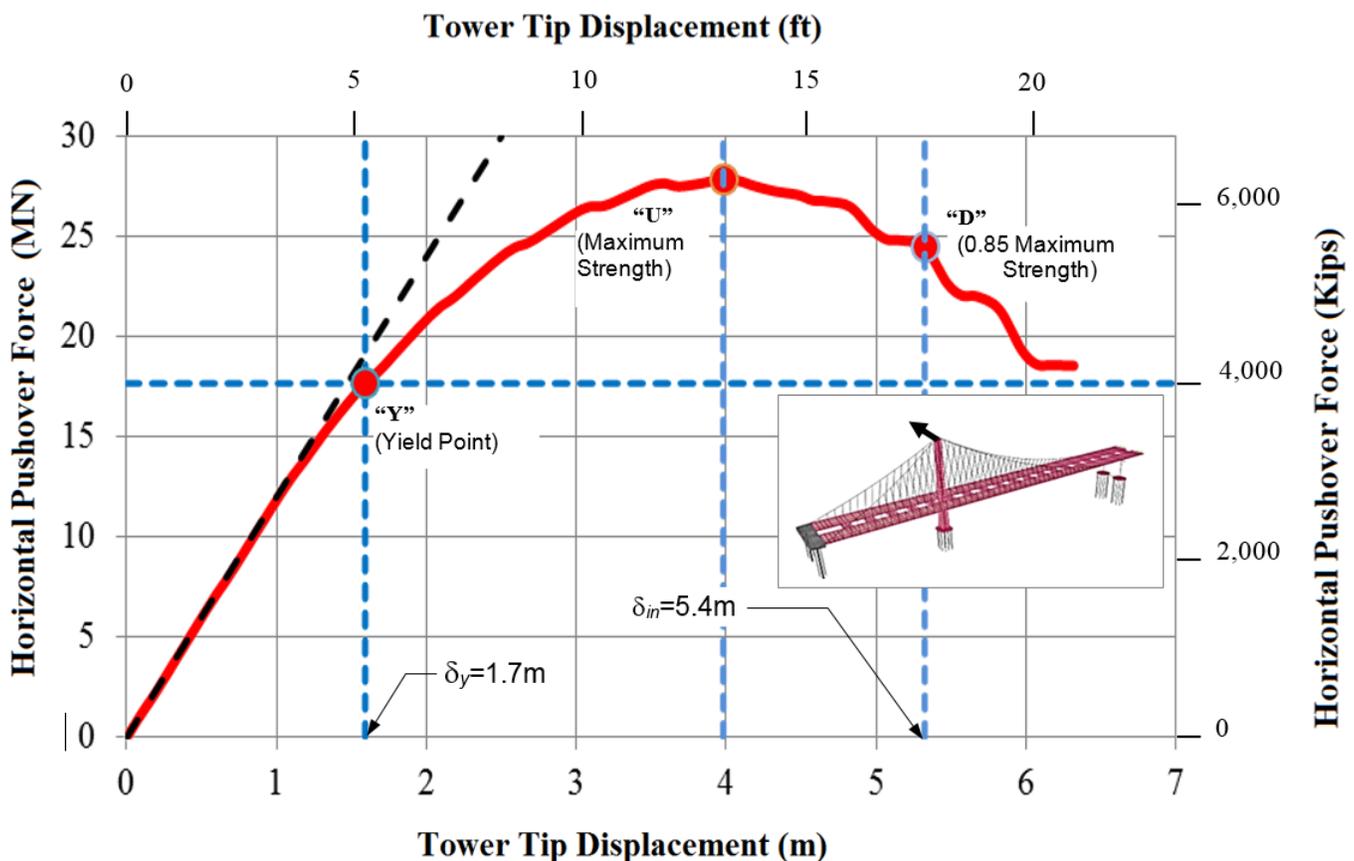


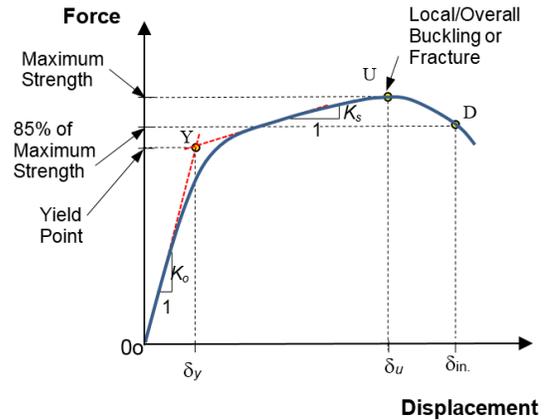
Figure 6. Transverse Pushover Curve of the Tower

Point Y on the pushover curve represents the yield point of the tower. For large and complex structures, such as the main tower of the SAS Bay Bridge, local yielding occurs at relatively small displacements, due primarily to stress concentrations; such small local yielding cannot be considered as the yield point of the tower. Therefore, a yield point must be defined for such structures. Figure 7 shows two definitions. Figure 7(a), point Y, shows the definition of yield point when there is relatively clear initial elastic linear behavior, as well as a linear second branch of the force-displacement curve. In this case, the yield point can be defined as the point of intersection of the initial stiffness line and the secondary stiffness line. However, in many cases, the initial behavior and the secondary branch of the force-displacement curve is not a straight line, which was the case for the pushover curve of the tower shown in Figure 6. In these cases, the yield point can be defined as the point where the displacement of the structure deviates from the initial stiffness line [line “ob” in Figure 7(b)], with an amount equal to 10% of the elastic displacement. In other words, the yield point Y is a point where distance bY in Figure 7(b) is equal to 10% of distance ab. In this paper, the authors followed the definition of the yield point using the 10% deviation rule, as shown in Figure 7(b).

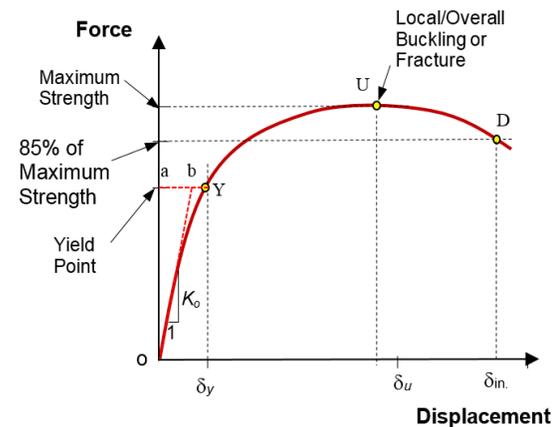
Point U on the pushover curve of Figure 7 corresponds to the point where maximum push-over strength was reached. Point D corresponds to a point where the pushover strength drops to 85% of the maximum strength at point U. Point D is worth noting as it can be used to calculate the ductility of a system. The displacement δ_m in Figure 7 is considered the maximum inelastic displacement. The ductility of a system is defined as this displacement δ_m at 85% maximum strength divided by the displacement at yield point δ_y in Figure 7. The pushover curve of Figure 6 shows the ductility of the tower in the transverse direction to be $\delta_m/\delta_y = 5.4\text{m}/1.7\text{m} = 3.2$, where 5.4m and 1.7m are δ_m and δ_y , respectively.

Figure 8 shows the equivalent stress plot of the tower at three important points (i.e., yield, maximum strength, and 0.85 maximum strength points) for pushover in the transverse direction. Here, points Y, U, and D correspond to the same points on Figure 6. At the yield point (point Y), the top seven and middle three pairs of shear links yielded first, while all other parts of the tower remained essentially elastic. Then, as the tower was pushed beyond the yield point, all of the shear links gradually yielded—with the exception the four at the bottom of the tower—as the system strain hardened before reaching the peak strength at point U. From points Y to U, some yielding of the tower shafts also occurred. Such yielding of tower shaft plates occurred at the mid-height portion of the tower, where there was a slight

change in the slope of the tower shaft. The connection interface between the tower grillage directly below the saddle also yielded. After passing point U, the strength of the tower dropped relatively quickly, and local buckling of the yielded mid-height portions of the tower shafts became more pronounced. The local buckling of the tower structure was essentially the overall buckling of the stiffened vertical shaft plates between the horizontal diaphragms.



(a) Structure with Clear Linear Hardening Slope



(b) Structure with No Clear Hardening Slope

Figure 7. Definition of Yield Point

During this stage, from point U to point D, larger regions of the tower’s grillage-to-tower shaft interface yielded. The yielding was likely due to the difference in rigidity of the grillage and the four separate tower shafts: the four tower shafts tended to deform independently (the mechanism by which the shear links work), while the rigid grillage was trying to hold them together and remain flat. Figure 9 shows shear link rotation versus tower tip displacement for two specific transverse direction shear links, which were the shear links at an elevation of 173 ft 10 in. (53m) and 357 ft

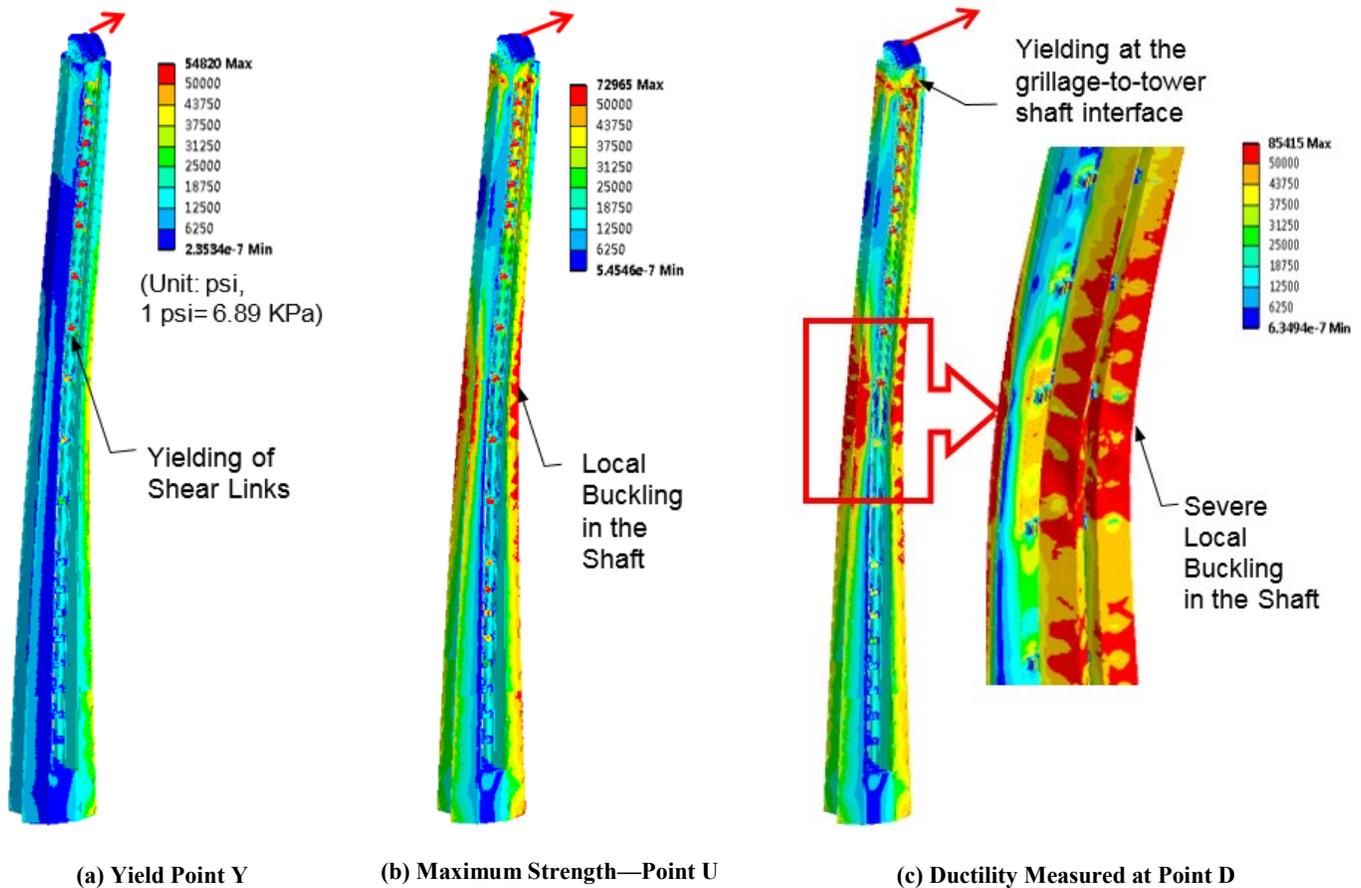


Figure 8. Equivalent (von Mises) Stresses at the Yield Point, Maximum Strength, and 85% Maximum Strength Points for Transverse Pushover of Tower

7 in. (109m). The shear link rotations plotted in Figure 9 were calculated using the same method described by McDaniel et al. [9], who performed tests of actual shear links representing the shear links in the SAS Bay Bridge.

Due to the assumed fixed tower base boundary condition, the shear link at 53m experienced less than 0.01-radian rotation during the entire pushover analysis. However, the shear links located at 109m experienced much higher rotation demand; the rotation reached 0.03 at a tower tip displacement of about 8.2 ft (2.5m), which was halfway between yield point Y and ultimate point U in Figure 6. The rotation reached the ultimate rotation capacity of 0.09 radians at the displacement of 15.75 ft (4.8m) on the pushover curve. Since the fracture of the material was not included in the model, and considering that the top group of shear links reached their ultimate rotation capacity, the degradation of base shear capacity could be even worse than that shown in Figure 6, in addition to the local buckling of the tower shafts.

Conclusion

In this paper, the authors resented a pushover analysis of the single tower of the SAS Bay Bridge performed by the bridge design team [3–6] and compared these results with a pushover analysis performed by the current authors. The analysis model of the tower used by the bridge design team in their pushover analysis consisted of members—such as the tower shafts—represented by beam elements with no shell elements. The beam elements are not able to capture the most important failure mode of steel plates in compression, local buckling of the plates, unless a more accurate nonlinear force-displacement relationship is incorporated. As presented here, all plates except the vertical stiffeners were modeled, including the vertical shaft plates, horizontal diaphragms, and the shear links using nonlinear shell elements capable of developing local buckling. Thus, a more robust prediction of the actual behavior of the tower was achieved. Figure 10 shows a comparison of the two ap-

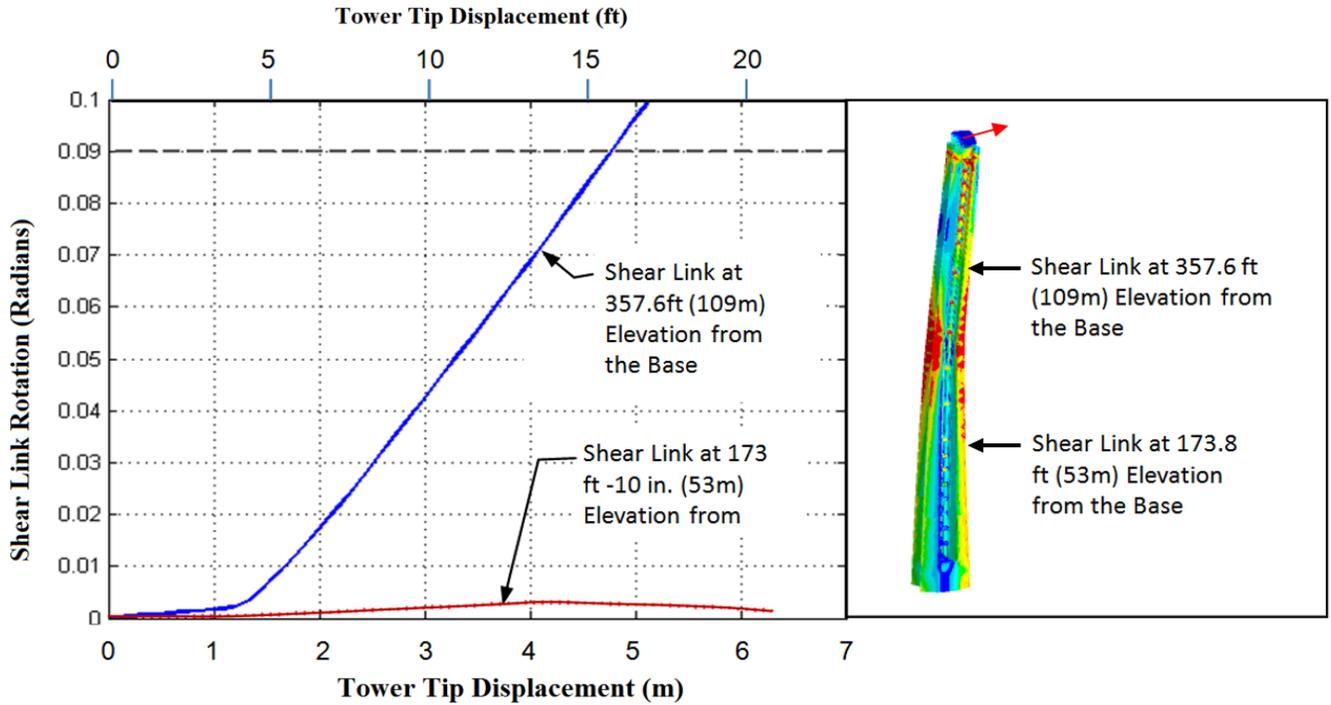


Figure 9. Shear Link Rotation versus Tower Tip Displacement for Transverse Pushover Analysis

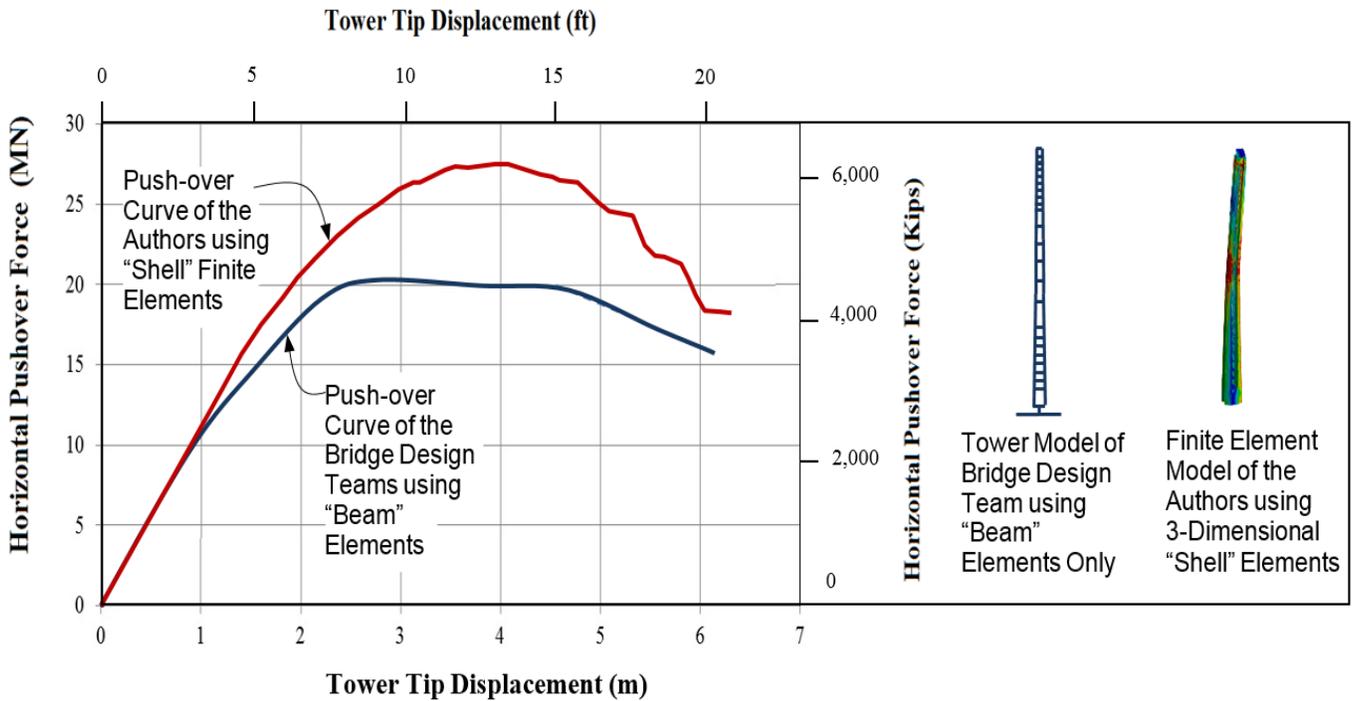


Figure 10. Comparison of Pushover Curves by the Original Bridge Design Team Using Simplistic Beam Elements to the Curves by the Current Authors Using Realistic Shell Elements

proaches: the tower pushover curves obtained using a very simplistic model of the tower and one that used a realistic, detailed finite element model. The most important finding of this current study was that the bridge design team adopted a simplistic modeling approach of the tower shafts by using only beam elements; premature local buckling of the tower shafts was not predicted in their pushover analysis and was not considered in the design of the tower. The main reason for premature local buckling of the tower plates is that the vertical stiffeners used in the tower were flat plates instead of geometries, T or U, which can be more effective in stiffening steel plates and, thus, prevent local buckling. Such stiffeners are used in most steel bridges, including in the orthotropic deck of the SAS Bay Bridge.

Qian and Astanah-Asl [10] studied the effects of various geometries and locations of the vertical stiffeners in steel bridge towers and piers. Their research found that equally spaced flat-plate stiffeners are the least effective stiffeners in preventing local buckling of plates. Based on these findings, the tower of the SAS Bay Bridge should be retrofitted to prevent premature local buckling during a major earthquake. Figure 11 shows the recommended retrofit measures. Here, bolted T sections, or welded pipes and channels, were added to the vertical stiffeners over about three-fourths of the height of the tower, where local buckling of the vertical plates of the tower is likely to occur. If a welded option is

selected, the traffic on the bridge needs to be reduced or halted during welding. However, to avoid welding in the field, the pipe and the channel in options 2 and 3, can be shop-welded to a plate, and the plate field-bolted to the vertical stiffeners.

The following are the authors' observations and conclusions, based on the results of realistic pushover analysis of the SAS Bay Bridge tower performed using shell elements for the steel plates:

1. This study showed that using beam elements in modeling large and complex structures instead of realistic shell elements may result in an incorrect prediction of the behavior of the structure. In this case, a local buckling failure mode of the steel plates, a critical failure mode of the bridge tower, was not captured correctly by beam elements used by the bridge design team.
2. The bridge design team's model used beam elements for the shafts, which resulted in underestimating the stiffness and ultimate strength before significant yielding. The consequence of underestimating stiffness and strength is that the inertia forces generated in the structure during a seismic event will be significantly larger than the time-history analysis of the structure predicted by the designers. Thus, the bridge

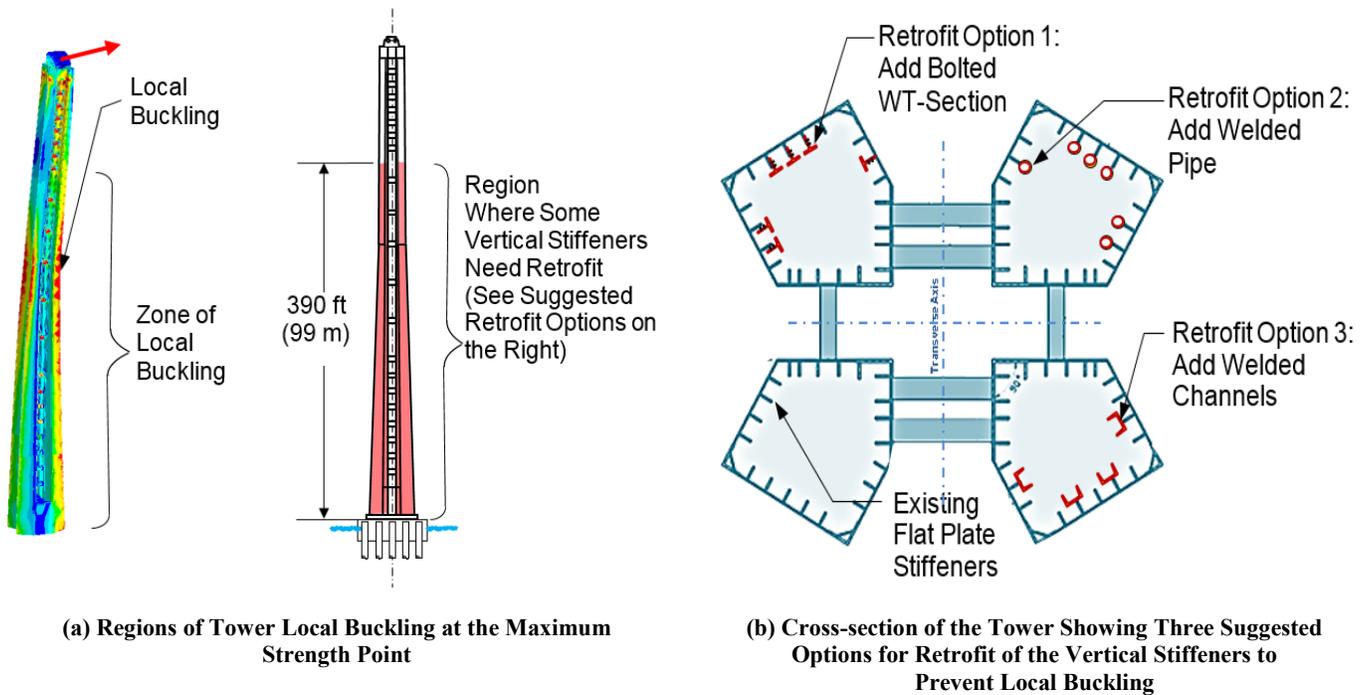


Figure 11. Suggested Retrofit for Vertical Stiffeners of the SAS Bay Bridge to Prevent Local Buckling of the Tower

was designed for smaller forces than it would experience when subjected to the seven earthquake records [11] that the bridge design team used in their design of the bridge.

3. As shown in Figure 10, because the bridge design team used beam elements in their pushover analysis, yielding of the tower occurs under much smaller forces than it would if the tower were realistically modeled using shell elements. This unrealistic early yielding resulted in a ductility of about 5.5 for the tower by the bridge design team, which is incorrect, compared to the realistic value of 3.2, resulting from a pushover analysis of the tower modeled using shell elements. The lower, but more realistic ductility, can result in less than desirable performance of the tower during major earthquakes.
4. The statement by the bridge design team, that “The shear links between the tower shafts will be the only inelastic elements in the tower and will act as fuses to protect the tower shafts from yielding,” is inaccurate. Pushover analysis of the tower incorporating steel plates modeled as shell elements demonstrates that, in fact, yielding and local buckling of the tower shafts occurs relatively early in the pushover analysis.
5. The analysis results summarized here also pose a question on the effectiveness of using a shear-link coupling system as a seismic fuse with a rigid restraint at the top in the form of the saddle and its supporting grillage. The shear link yielding depends on the relative displacement of the tower shafts in the vertical direction. However, with a rigid saddle restraining the top of the tower shafts, such a yielding mechanism is disrupted because of the saddle and its supporting grillage. The results also showed that a change of the slope of the tower shafts at about mid-height may result in stress concentration at that location and may cause local yielding and local buckling to initiate at that location. A constant slope for the tower shafts could prevent such stress concentrations.
6. Local buckling of the tower shafts occurred relatively early in the pushover analysis. Since the tower will be pushed beyond the design level earthquake during major seismic events, it is critical to prevent local buckling of the tower shafts.
7. Based on the findings of Qian and Astaneh-Asl [10], which detail the behavior of stiffeners in steel bridge towers and piers, efficient and economical retrofit measures, as shown in Figure 11, are proposed for vertical stiffeners of the tower to prevent local buckling of tower shafts.

Acknowledgments

This study was part of a larger project on Investigation of Seismic Performance of the New Self-Anchored Suspension (SAS) Bay Bridge East Spans at the University of California Berkeley, with Professor Abolhassan Astaneh-Asl as the principal investigator. The authors would like to express their sincere appreciation for the tremendous technical support provided by Dr. Metin Ozen, President, Casey Heydari, and the analysts at Ozen Engineering, Inc. (<https://www.ozeninc.com>) on the use of the powerful ANSYS non-linear structural analysis software featured in this project. The authors would also like to express their sincere appreciation to Claire Johnson for her excellent editing of this manuscript.

References

- [1] USGS (2016). 2008 Bay Area Earthquake Probabilities. Website of the U.S.G.S. Retrieved from <http://earthquake.usgs.gov/regional/nca/ucerf>
- [2] Astaneh-Asl, A., Tabbakhha, M., & Qian, X. (2016). Performance of Anchor Rods of the Single Tower of the New Self-Anchored Suspension Bay Bridge. *Proceedings of the 2016 IAJC-ISAM International Conference*. Orlando, Florida.
- [3] Nader, M., Manzanarez, R., & Maroney, B. (2000). Seismic Design Strategy of the New East Bay Bridge Suspension Span. *Proceedings of the 12th World Conference on Earthquake Engineering*. Paper No. 0911, Auckland, New Zealand.
- [4] Nader, M., Lopez-Jara, J., & Mibelli, C. (2002). Seismic Design of the New San Francisco-Oakland Bay Bridge Self-Anchored Suspension Span. *Proceedings of the 3rd Nat. Seismic Conf. and Workshop on Bridges and Highways*, Portland, Oregon.
- [5] Manzanarez, R., Nader, M., Abbas, S., & Baker, G. (2000). Design of the New San Francisco - Oakland Bay Bridge. *Advanced Technology in Structural Engineering*, 67, 1-12.
- [6] Nader, M., & Maroney, B. (2007). One-of-a-Kind Design: The New San Francisco-Oakland Bay Bridge Self-Anchored Suspension Span. *Structure Magazine*, October.
- [7] ANSYS, Inc. (2013). ANSYS Mechanical APDL Theory Reference, Release 15. ANSYS Inc., Canonsburg, PA.
- [8] Caltrans. (1999). Project Plans for Construction on State Highway in San Francisco County in San Francisco from 0.6km to 1.3km East of the Yerba Buena Tunnel East Portal. Engineering Drawings, The State of California, Department of Transportation, Sacramento.

-
- [9] McDaniel, C. C., Uang, C. M., & Seible, F. (2003). Cyclic testing of built-up steel shear links for the new bay bridge. *Journal of Structural Engineering*, 129(6), 801-809.
- [10] Qian, X., & Astaneh-Asl, A. (2016). Behavior and Seismic Design of Stiffeners for Steel Bridge Tower Legs and Piers. *Proceedings of the World Congress on Civil, Structural, and Environmental Engineering (CSEE'16)*, Prague, Czech Republic.
- [11] Fugro-Earth Mechanics. (2001). Ground Motion Report, San Francisco-Oakland Bay Bridge East Span Seismic Safety Project, A Report Prepared for the California, Department of Transportation, Earth Mechanics, Inc., and Fugro, a Joint Venture.

Biographies

ABOLHASSAN ASTANEH-ASL is a professor in the Department of Civil and Environmental Engineering at the University of California, Berkeley, and has over 48 years of experience in design, research, and teaching courses in structural and earthquake engineering, design of buildings, bridges, and other structures, as well as blast protection of structures. He has testified as an expert before policy-making and judicial bodies, and has completed more than 50 major research projects and has more than 300 technical publications. Dr. Astaneh-Asl may be reached at As-taneh@berkeley.edu

XIN QIAN is currently (2017) a PhD candidate at the University of California, Berkeley. She earned her BS degree at the Hong Kong Polytechnic University and MS in Civil Engineering from the University of California, Berkeley. Her research interests include steel and composite structures and earthquake engineering. She is the recipient of the first Professor T. Y. Lin Fellowship in Structural Engineering at UC Berkeley and is active in student activities in earthquake engineering as the president of the EERI Berkeley student chapter. Ms. Qian may be reached at cin-dyqian@berkeley.edu

PERFORMANCE ANALYSIS OF DISTRIBUTED SYSTEMS INVOLVING LOOPS

Toqeer Israr, Eastern Illinois University

Abstract

In this study, the author evaluated the execution and performance of a distributed global system (global activity) comprised of sub-services sequenced with strict and weak loops. For this global activity, performance was determined based on the performance and sequence types of the constituent sub-activities. This was achieved by modeling each -activity as a partially ordered specification (POS), such that each sub-activity was identified by independent input and output events and the minimum delays between these events. This technique allowed hierarchical composition of two or more sub-activities.

Introduction

Many commercial systems, especially cloud-based applications, are the result of an aggregation of various types of services (fine-grained/composite services), and will be referred to as activities in this paper. Quite often, these composite activities may involve multiple interacting components located on different processors. In the rapidly changing world of technology, no system is ever constant. These requests for changes, due to various reasons, such as evolving customer requirements, have significant effects on the performance of such systems. For example, a system for analyzing the data of an online retail application could be implemented on a multi-tiered system, where the data are broken down for several servers to analyze each sub-analysis. The results of these sub-analyses are then combined to give an overview of a customer's shopping pattern in order to enable the company to predict leading trends and customer habits, prepare for demands, and optimize pricing and promotions.

For this current study, the performance of a distributed global system was examined, in particular for when either a sub-service or an involved component changes. A global system could be a system implemented using cloud technology working with big data architecture. This becomes even more complex as the assumption is made that execution of all components does not start or end at the same instant.

Focus of this Research

This study focused on the following:

- First, the performance of a composite activity was determined, given the performance delays of its sub-systems; this becomes quite interesting when a particular assumption is made such that not all involved components may start or end their executions at the same time.
- Second, how long a particular component is involved in a service was calculated, such that an inference can be made about its availability for other services. Note that a component may complete all of its executions long before all of the executions of a service, and hence an involved component completion time, will always be less than or equal to the completion time of a service in which the component is involved.
- Third, the effects of the global service on performance were examined by modification of the existing global service: either a sub-service involved in the global service is replaced by another sub-service or a component involved in the global service is replaced with another component with different performance parameters.

In previous studies by Israr [1] and Israr and Bochmann [2], the authors focused on a global service composed of multiple sub-services that were sequenced with strict and weak sequencing, concurrent and alternative sequencing operators. This current study was focused on analyzing performance of the global services, where the global service is composed of sub-services sequenced with strict and weak while loops. In addition, the analyses in this study can be extended to analyze performance of parts of a big data framework.

Big Data Analysis

An online retailer with thousands of online customers can serve as an example of a big data analysis. As the number of transactions increase (sometimes up to 500,000 transactions per second) [3], so does the data being generated, causing companies occasionally to implement big data strategies in a cloud-based environment. This typically involves tools and technologies such as Hadoop, Pig, and Hive to collect, interpret, and analyze data from many angles with the goal of discovering a previously hidden insight, which, in turn, can provide a competitive advantage or address a pressing business problem.

Figure 1 illustrates a MapReduce infrastructure of a Hadoop eco system and the generated data from an online retail transaction labeled as “big data.” MapReduce was used to process parallelizable problems across enormous datasets using a great number of servers (components). According to the tutorial, “A MapReduce job typically splits the input data-set into independent sub-data-sets, which are analyzed and processed by the map tasks in a completely parallel manner. This framework then sorts the outputs of the maps” [4] for use as inputs to reduce the number of tasks. The outputs from the reduce function yield the desired end results.

In this case, the data processing done by the *map* functions could be determining trends for the products purchased. Each *map()* function involves two servers and analyzes data from different geographical locations, such that *map()*₁, *map()*₂, and *map()*₃ are analyzing data from transactions originating from North America, Europe, and rest of the world, respectively. The reduce functions bring all of the analyses back together to enable the company to predict things such as leading trends, customer habits, prepare for demands, optimize pricing, and promotions. These map and reduce functions can be modeled as activities (sub-services) that can be sequenced using a sequence, a choice, and/or loop operators.

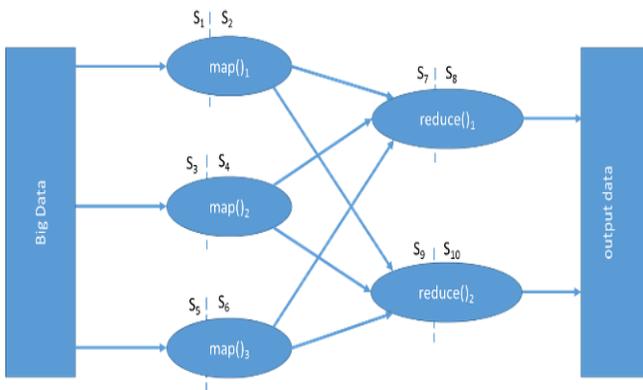


Figure 1. MapReduce Architecture

As mentioned previously, no state of a system is ever constant and is always evolving. From a performance point of view, one questions the effect on the performance of the complete data analysis (end-to-end), if the nature of map/reduce jobs change (service change), or the number of servers or the servers (S_1, S_2, S_3, \dots) (component change) processing these jobs changes. Also, when do these servers become available for other services? As each of these *map()* and *reduce()* functions involves multiple servers, not all servers will always be ready to process the request and could cause a time delay.

Modeling

Several methodologies have been used to describe the modeling of a distributed composite system. Typically, several of these models can possibly be refined into sub-activities and additionally into sub-sub-activities, where the end result is often identified with a single component [5]. However, that is not necessarily practical in many distributed systems, as even the final decomposition involves multiple components. This is also known as the “crosscutting” nature of distributed services identified with multiple collaborating components, where each component contributes to more than one service. The behavior of the components can be specified precisely how they are modeled by their roles. However, the behavior of the service becomes fragmented. Another view is required, where the behavior is focused on modeling the global behavior of a given service explicitly [5]. In this view, there would be an essential need to demonstrate the relationships and the dependencies among the roles (implemented on various components) that are involved. This becomes essential when the roles do not necessarily start and end their execution at the same time.

Modeling Distributed Activities

To satisfy such needs, partial order specification (POS), a new modeling paradigm, was introduced. Figure 2 shows how this new modeling paradigm, POS, models a partially ordered set of inputs and outputs [1, 2], which enables the events occurring at different times to be modeled. These events are typically starting and ending events, which can be independent. Using these events, POS illustrates the dependencies between different actions of a role (such as ending of a role and starting of another role) and allows one to model a service as an activity and analyze its performance.

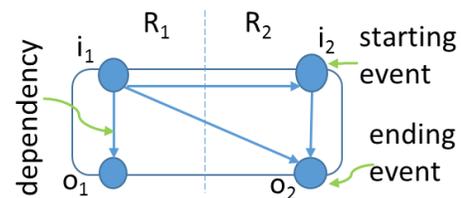


Figure 2. Partial Order Specification

For a given role, a POS models a starting and an ending event, as a filled-in circle in Figure 2 [1]. As can be seen, a partially ordered set can be formed of these events, giving rise to causal relationships between some of the events—these relationships are depicted by arrows in Figure 2. A starting event marks the beginning of the execution of the first action for a given role, while an ending event marks the

ending of the execution of the last event for a given role in a given activity. Even though the ending events may not be ordered relative to each other, there exists a causal relationship between each starting event and an ending event of the same role, also known as local sequencing. Figure 2 depicts an activity having two starting events (i_1 and i_2) and two ending events (o_1 and o_2). For events of the same role, the ending event must occur after the starting event of the same role, due to local sequencing (i.e., o_1 occurs after i_1 and o_2 occurs after i_2). Additionally, it can be seen that i_1 causes i_2 to occur. This, in turn, along with local sequencing, postulates that all of the events in the activity must occur after i_1 . Due to the relationships $i_1 \rightarrow i_2$ and $i_2 \rightarrow o_2$, there is another dependency from i_1 to o_2 .

Strict and Weak Sequencing

Strong sequencing between two activities, for example where A2 occurs after A1, implies that all of the actions of all of the roles in A1 must have executed to completion before the action of any role in A2 may begin. In contrast, weak sequencing between the same two activities, A1 and A2, implies that there exists a local sequencing between A1 and A2 such that a role may start executing actions of A2 as soon as that role has completed execution of all its actions in A1. Whenever strong sequencing exists, weak sequencing exists as well, but not inversely. Furthermore, in weak sequencing, a role may start execution of its actions in A2 if it is not involved in A1 and, hence, execution of A2 may start before execution of A1 even begins.

Performance

Performance of such systems on a larger scale can be very difficult to analyze. Huang et al. [6] discussed a stochastic Petri-net workflow model to propose various performance equations for basic routing pattern of a workflow system. Li et al. [7] extended the Workflow net (WF-net) with timing information to provide a formal framework for modeling and performance analysis. In their work, they proposed a method for computing the lower bound of the average turnaround time of transaction instances in a given workflow. Similar work has been done by Wang et al. [6], Hao and Pei-an [8], and Lazowska et al. [9]. McNeile [10] modeled and analyzed end-to-end workflow delays, but that analysis assumed that all of the components were available at the beginning of the workflow, thereby yielding a single workflow delay.

All of the aforementioned work assumed that a workflow implemented as an execution of activities was implemented by a single role. With a workflow involving multiple roles, the very first concern are the starting and ending times of

each role involved in the given workflow—times of the starting and ending events of each role. Secondly, within a given collaboration, dependencies need to be identified, which may exist between various events of different roles. Based on these dependencies, performance can be analyzed for not only events involved in local sequencing but also between events of one role and events of another or similar role. As such, well-structured collaborations with multiple roles, sequenced with standard UML operators to yield a global collaboration describing an abstract service, were explored.

Delay for a Given Activity

Israr [1] introduced an approach for determining the dependencies among the input and output events within a given sub-activity. For a given sub-activity, according to this approach, the delay is measured between the time instance of the occurrence of input event i and dependent output event o , provided all of the other events on which o depends have occurred a long time before. This delay is called nominal execution time delay (NETD), written as Δ_{io}^i . This led to Equation (1), which yields the performance of a collaboration, D , based on dependencies between input and output events:

$${}_D T_o = \max_{i \in I(D)} ({}_D T_i + {}_{(cp)D} \Delta_{io}^i) \quad (1)$$

where, T_o is the time of output event o ; T_i is the time of input event i ; Δ_{io}^i is the NETD from input event i to output event o ; and, $I(D)$ is the set of input events. Subscript “D” indicates that all of the notations are for the abstract activity D , and (cp) depicts a control flow path, a single execution of a given system depicting a single control flow (SCF), as compared to multiple executions of a system (i.e., multiple control flows).

If the events are independent, then the NETD between input event i and non-dependent output event o is assumed to be determined by Equation (2):

$${}_D \Delta_{io}^i = -\infty \quad (2)$$

hence, Equation (1) is not limited to dependent events, but rather is applicable for all involved events—dependent and independent events alike.

Israr [1] assumed that the NETD is attained during a control flow path, which may not be realistic, if shared resources are involved in the processing of several inputs on which a single output depends. Hence, the assumption was made that there were no shared resources, and each role or all concurrent activities of a given role would be implemented by an independent processor.

Types of Delays

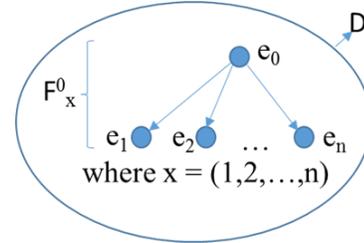
For the modeling of the performance of collaborations, the NETD was considered, such that capture of the performance of a collaboration could be one of three types: fixed delays, range of delays, and delays defined in terms of probability distributions—stochastic delays. For stochastic delays (SD), written as ${}_{(\text{stoc})}^{(\text{cp})}D\Delta_o^i$, it was assumed that the execution delays (measured or specified) were of a stochastic nature and were defined by a probability distribution that could be measured by performing a large number of delay measurements. Stochastic delays could follow any kind of distribution. Hence, general equations are provided that can be applied to any kind of distribution. A special case of stochastic delay is the fixed delay (FD). If a given NETD has a stochastic delay, where the distribution is a Dirac Delta function, then it can be said that the NETD has a determinist duration, or a fixed delay (FD), of ${}_{(\text{fixed})}^{(\text{cp})}D\Delta_o^i$. This means that the delay for a given control flow path, cp, always results in the same value, provided the starting conditions (or initial states) are the same for all participating components. Fixed delays are commonly used when there is a need to specify performance for a single control flow path for hard, real-time control systems.

For example, a performance requirement could be that a traffic light takes exactly 45 seconds to go from green to red, exactly another 20 seconds to go from green to yellow, and exactly another 2 seconds to go from yellow to red. An NETD ${}^{(\text{cp})}D\Delta_o^i$ can have a range of delays (RD), where the delay specified for a control flow cp may provide different values, and it is important to specify minimum and maximum values, written as ${}_{(\text{max})}^{(\text{cp})}D\Delta_o^i$ and ${}_{(\text{min})}^{(\text{cp})}D\Delta_o^i$, respectively. In actuality, the delays may be of a stochastic nature, but one is normally not interested in the precise form of the probability distribution. Range of delays is used to describe the behavior of system models for hard, real-time systems using, for instance, the formalization of Timed Automata. If the range is infinitesimally small, where ${}_{(\text{max})}^{(\text{cp})}D\Delta_o^i - {}_{(\text{min})}^{(\text{cp})}D\Delta_o^i \approx 0$, then this results in a fixed delay, as detailed above.

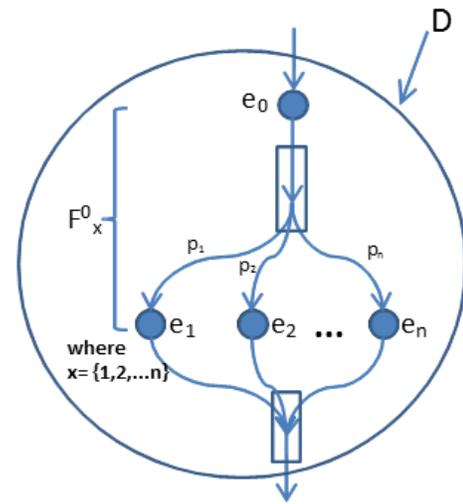
Basic Performance Characteristics of an Activity

Let us consider the various orderings of events e_0, e_1, \dots, e_n in Figure 3, and abstract the sequence of events e_0, e_1, \dots, e_n by a composite activity, D, where composite activity D abstracts concurrent events in Figure 3. It was assumed that there existed a stochastic delay, $F_x^w(t)$, between events e_w and e_x , as shown in Figure 3 and characterized by a cumulative distribution function (CDF), $F_x^w(t)$. It was further assumed that these delays were statistically independent. If the

distributions of the delays between the events were considered to be a delta distribution, then these delays would have a deterministic duration, leading to the situation of “Fixed Delays.”



(a) Concurrency



(b) Alternative

Figure 3. Sequencing Operators

Concurrency: Stochastic Delays

To determine the time it would take for the earliest event among e_1, e_2, \dots, e_n after event e_0 to occur, as shown in Figure 3, then the earliest or “minimum” CDF of global activity D, composed of parallel events e_x , can be calculated using Equation (3), which represents the completion of the earliest event [11]:

$$\min F_D(t) = 1 - \prod_{x=1}^n (1 - F_x^0(t)) \quad (3)$$

Next, the completion of global activity D requires all of the events e_x to occur and, hence, the delay for the occurrence of the last event e_x needs to be calculated. This can be accomplished by calculating the delay between events with

the maximum time delay [11], as defined in Equation (4), which represents the completion of the last event:

$$\max F_D(t) = \prod_{x=1}^n F_x^0(t) \quad (4)$$

Concurrency: Fixed and Range of Delays

If the delays between the events are assumed to be a delta distribution, and if a range of delays are considered for the deterministic duration of the abstract activity D, then the range of delays can be defined by Equations (5) and (6), which represent the completion of the earliest and last events, respectively:

$$\min F_D = \min_{x=1..n} (F_x^0) \quad (5)$$

$$\max F_D = \max_{x=1..n} (F_x^0) \quad (6)$$

These equations can be used to determine the distribution of a global scenario, given the distribution function between individual events.

Alternatives: Stochastic and Fixed Delays

Alternative execution occurs when there is a decision to be made among multiple successive events, and only one of the successor events occurs, as shown in Figure 3(b). To obtain a distribution delay, each path is assigned a probability value p_i , leading to alternative paths i . If the alternative case is examined, such as in Figure 3(b), then the overall distribution of D is defined by Equation (7) [11]:

$$F_D(t) = \sum_{i=1}^n p_i F_i^0(t) \quad (7)$$

where, p_i is the probability for event e_i to occur.

The fixed delay of the abstract activity D is the delay of the particular single control flow path selected to be executed as defined in Equation (8):

$$F_D = F_i^0 \quad (8)$$

where, e_i is the event occurring and where $1 \leq i \leq n$.

If the range of delays is considered, then all of the control flow paths need to be considered, which would lead to the greatest and smallest delays, as calculated by Equations (9) and (10) for the minimum and maximum delays, respectively:

$$\min F_D = \min_{i=1..n} (F_i^0) \quad (9)$$

$$\max F_D = \max_{i=1..n} (F_i^0) \quad (10)$$

These equations can be used to determine the distribution of a global scenario, given the distribution function between individual events.

Deriving General Equations for Various Operators

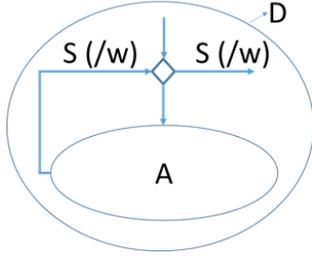
For this current study, NETDs were considered to be fixed, range, or stochastic delays. This was annotated by ${}_{(mzz)D}\Delta_z^w$ for the delay from the starting event, $w \in I(D)$, to the ending event $z \in O(D)$, where mzz was the delay type (i.e., $mzz = \text{fixed, max, min or stoc}$). The delay ${}_{(mzz)D}\Delta_z^w$ of the composite collaboration D depends on the participation of the roles w and z in the sub-collaborations A and/or B, w represents the role of the input, and z represents the role of the output in collaboration D. As such, for each operator, the equations were classified according to this participation. $R(X)$ describes the roles involved in collaboration X, while $I(X)$ and $O(X)$ describe all of the input and output events of collaboration X, respectively.

Israr [1] proposed definitions of Δ^i , for strict, weak, concurrent, and alternative sequencing operators. In the following section, these were extended by analyzing performance of sub-activity A and B sequenced with **weak** and **strict while loop** operators with **independent input** and **output** events. With these definitions, one can calculate the completion time of the global activity and the total time a component is involved in a sub-activity as well as a global activity [1]. These compositions are abstracted by activity D.

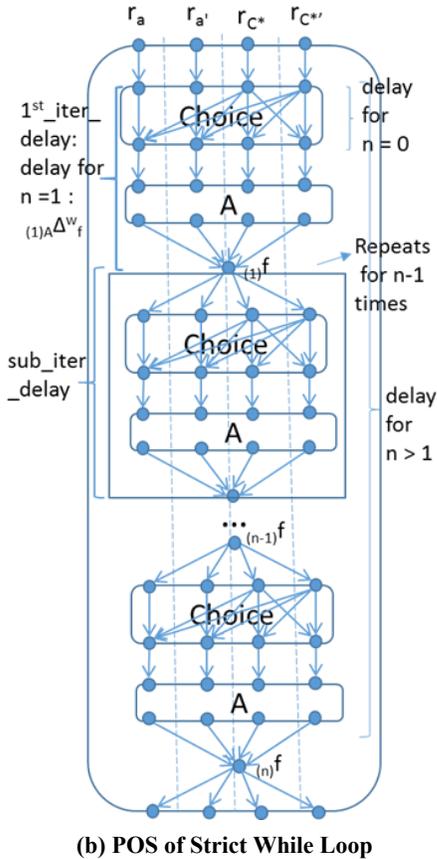
Strict While Loop

Figure 4(a) shows a control flow diagram of a strict while loop repeating sub-collaboration A. The follow-up collaboration "B" is not modeled (or analyzed), as the analysis for a sub-collaboration strictly sequenced with B was previously analyzed [1]. Similar to definitions of strict and weak sequencing, Bochmann [5] defines a strict while loop, where collaboration C_1 is repeated and then followed by C_2 , as "C₁ is executed zero, one or more times and then C₂ will be executed; more precisely, the behavior starts with a choice between C_1 and C_2 ; if C_1 is executed, there is **strict** sequencing between the end of C_1 and the choice of executing C_1 again or terminating the loop with C_2 , written as $C_1 *s C_2$." Before each iteration of C_1 , a set of roles C^* [represented by r_{c^*} ... to r_{c^*} in Figure 4(b)], where $C^* \in R(D)$ makes a choice for the execution of C_1 repeatedly or to terminate the loop and execute C_2 , where all of the roles in C^* make the same choice. No action in either sub-collaboration A or B may

start to execute until this choice is made. This is modeled by adding a sub-collaboration “Choice,” preceding the sub-collaborations C_1 . Choice sub-collaboration has roles $R(\text{Choice}) = C^* \cup R(A) \cup R(B)$, where dependencies are introduced from the starting events of C^* to the ending events of $R(\text{Choice})$, which ensures none of the roles start their execution in C_1 or C_2 until the choice is complete. It is assumed that the act of making the choice and the propagation of this choice to $R(\text{Choice})$ [(shown by dependency arcs in Choice sub-collaboration in Fig 4(b)] is done instantly, and does not add any delay.



(a) Strict (/weak) While Loop



(b) POS of Strict While Loop

Figure 4. While Loops

Figure 4(b) shows a partial order diagram that defines the dynamic behavior of the strict while loop for the control flow path, where sub-collaboration A is executed n times. The special case of $n=1$ is also indicated. In case of $n=0$, the delays are zero. As this is a strict while loop, all of the ending events of each iteration of collaboration A synchronize at a synchronization event, written as $(j)f$ for iteration j . To keep track of the iteration number, the notation $(j)\alpha$ is introduced, where α represents any of the behavioral properties of a collaboration and j is an integer representing the number of the iteration. The index j is shown for only those item that relate to different iterations. However, there is no explicit synchronization before the starting events of the first iteration of collaboration Choice and A ; that is, no assumption is made about the kind of sequence that precedes the composite collaboration D .

Consideration of a Single Control Flow Path

If a single control flow path of a strict while loop is considered, where sub-collaboration A repeats n times, the NETD of the composite collaboration $D \stackrel{(cp)}{D} \Delta_z^w$ for $w \in I(D)$ and $z \in O(D)$ is given by Equations (11)-(13) in Table 1:

Table 1. Strict While Loop Operator (single control flow)

n		Fixed Delays / Range of Delays	
0	if $w \in C^*$	0	(11a)
	Otherwise	$-\infty$	(11b)
1	if $w \in C^*$	$\max_{y \in I(A), x \in O(A)} ((l) \stackrel{(cp)}{(mzz)A} \Delta_x^y)$	(12a)
	Otherwise	$\max_{x \in O(A)} ((l) \stackrel{(cp)}{(mzz)A} \Delta_x^w)$	(12b)
> 1	if $w \in C^*$	$\sum_{j=1}^n (\max_{y \in I(A), x \in O(A)} ((j) \stackrel{(cp)}{(mzz)A} \Delta_x^y))$	(13a)
	Otherwise	$\max_{x \in O(A)} ((l) \stackrel{(cp)}{(mzz)A} \Delta_x^w) + \sum_{j=2}^n (\max_{y \in I(A), x \in O(A)} ((j) \stackrel{(cp)}{A} \Delta_x^y))$	(13b)

For Equations (11)-(13), it was assumed that all of the dependencies shown in Figure 4(b) by an arrow are associated with a zero delay.

Case $n = 0$ – A executes 0 times, but Choice executes once.

If $w \in C^*$, none of the involved roles can start their execution until the choice is made by roles $r_{c^*} \dots r_{c^{**}}$ of collaboration Choice, causing a dependency from role $r_{c^*} \dots r_{c^{**}}$ to all of the involved roles, $R(D)$. It is assumed that the delay to make the choice and for choice propagation to be zero, and, hence, the delay from the roles $r_{c^*} \dots r_{c^{**}}$ to all of the involved roles is zero. Otherwise, as there is no execution of A , there

is no dependency and, hence, no delay from any starting event to any ending event other than those mentioned previously. This is represented as $-\infty$, as per Equation (2).

Case $n = 1$ – A executes once, but Choice executes twice.

If $w \notin C^*$, collaboration A executes once, which is equivalent to sub-collaboration A strictly sequenced with a following collaboration. As all of the paths of execution merge at $(1)f$, using Equation (6), this delay was calculated as the maximum delay over all of the outputs of A from input w, as defined by Equation (14):

$$1^{st_iter_delay} = \max_{x \in O(A)} ({}_{(1)}^{(cp)} ({}_{(mzz)A} \Delta_x^w)) \quad (14)$$

If $w \in C^*$, and if the starting event of role w belongs to a role making the choice, then no execution in collaboration A could have started until this starting event from role w occurs. Furthermore, the ending event of role z cannot occur until all of the dependencies have been satisfied from all of the starting events. Hence, using Equation (6), the collaboration's delay is the maximum over all of the starting and ending events, as stated in Equation (12a).

Case $n > 1$ – sub-collaboration A executes n times, and Choice executes n+1 times.

From Equations (12a) and (12b), the delay for A's first iteration is known. As seen in Figure 4(b), the *subsequent iter delay* looks identical to the $1^{st_iter_delay}$ with the addition of the synchronization event, f, at the beginning of each iteration, with the dependencies. When the delay for $1^{st_iter_delay}$ is calculated, it is assumed that starting events except starting event of role w have occurred some time ago. Hence, the delays from remaining starting events to the synchronization event, $(1)f$, are not considered. This is not the case for the 2nd iteration and onwards, as the starting events for those iterations cannot occur until $(n-1)f$ has occurred. Only then can the execution of the nth iteration start. Since the delay from all starting events needs to be considered, applying Equations (6)-(14) yields the subsequent iteration delay, as defined by Equation (15):

$$subsequent_iter_delay = \max_{w \in I(A)} ({}_{(j)}^{(cp)} ({}_{A} \Delta_x^w)) \quad (15)$$

For each of these iterations, the NETD depends on the execution path of the body of A being executed. Hence, this delay could be different, f, or each iteration, depending on the control flow path in A. The total delay can be calculated using Equation (16):

$$2_to_n_delay = \sum_{j=2}^n ({}_{(j)}^{(cp)} ({}_{A} \Delta_x^w)) \quad (16)$$

From Figure 4(b), it is quite evident that the NETD of the composite collaboration is the sum of $1^{st_iter_delay}$ and $2_to_n_delay$, which yields Equation (13b).

Consideration of Multiple Control Flow Paths: Range of Delays

If all of the possible control flow paths are considered then, based on the equations from Table 1, it is quite evident that the minimum delay occurs for $n = 0$ and the maximum delay occurs when $n = \infty$, which gives rise to Equations (17a/b) and (18), minimum and maximum delay, respectively:

$$\text{if } w \in C^*, \quad ({}_{(min)D}^{(cp)} \Delta_z^w = 0) \quad (17a)$$

$$\text{otherwise,} \quad ({}_{(min)D}^{(cp)} \Delta_z^w = -\infty) \quad (17b)$$

$$({}_{(max)D}^{(cp)} \Delta_z^w = \infty) \quad (18)$$

Consideration of Multiple Control Flow Paths: Stochastic Delays

For stochastic delays, it is assumed that each time the Choice is performed, there is a probability p that A is executed and a probability $q=1-p$ to stop the iterations. Then the NETD for collaboration D, $({}_{(stoc)D}^{(cp)} \Delta_z^w)$, is given by Equations (19) and (20):

$$\text{if } w \in C^*, \quad q * \sum_{i=1}^n p^i \otimes_{j=1}^i ({}_{(stoc)A}^{(cp)} \Delta_z^w) \quad (19)$$

otherwise,

$$pq * \prod_{x \in O(A)} ({}_{(1)}^{(cp)} ({}_{(stoc)A} \Delta_x^w(t)) \otimes) \quad (20)$$

$$\left(\delta + \sum_{i=1}^{n-1} p^i \otimes_{j=1}^i \prod_{y \in I(A), x \in O(A)} ({}_{(j)}^{(cp)} ({}_{(stoc)A} \Delta_x^w(t)) \right)$$

And there is the probability q that there is no dependency from w to z, when $w \notin C^*$. This happens when $n = 0$.

The above delay is calculated by applying Equation (7) to the equations in Table 1. The probability p can be applied for each time A executes and probability q once for A to stop the iteration and delays are summed. If $w \in C^*$, then the starting event belongs to the roles involved in making the Choice, then applying Equations (7)-(11a), (12a), and (13a) and summing the delays yields Equation (19).

Otherwise, the starting event does not belong to a role involved in making the Choice, then applying Equations (7)–(11b), (12b), and (13b) and summing the delays yields Equation (20). If the sub-collaboration does not iterate at all (i.e., $n = 0$), then there obviously is no delay from w to z and, hence, no dependency from w to z .

Weak While Loop

Israr [1] defines a weak while loop with body C_1 followed by collaboration C_2 similar to that of a strict while loop “...except that **weak** sequencing is used between the end of C_1 and the choice of executing C_1 again or terminating the loop with C_2 .” This is annotated as $C_1 *w C_2$ [5]. Figure 5 shows a POS definition of such a weak while loop.

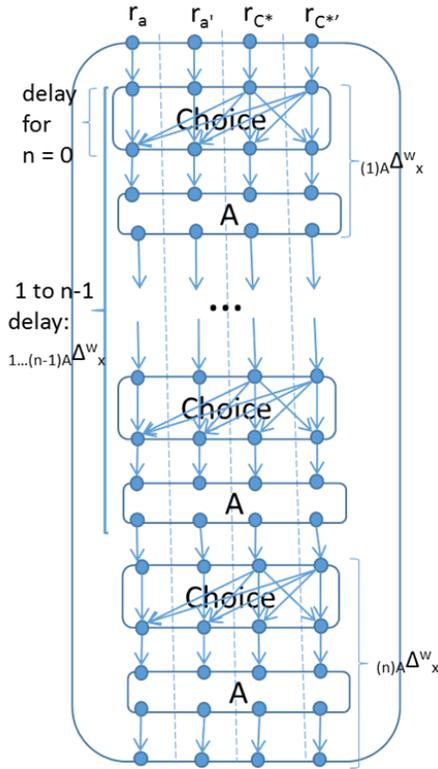


Figure 5. POS of the Weak While Loop

Consideration of a Single Control Flow: Fixed Delays

If a single control flow path is considered of a weak while loop, where sub-collaboration A repeats n times, equations to calculate NETDs for $w \in I(A)$ and $z \in O(A)$ for the composite collaboration D , ${}^{(cp)}_D \Delta_z^w$, are given by Equations (21)–(24) in Table 2.

Table 2. Weak While Loop

n		Fixed Delays / Range of Delays
0	if $w \in C^*$	0 (21a)
	Otherwise	$-\infty$ (22b)
1	if $w \in C^*$	$\max_{y \in I(A)} ({}^{(cp)}_{(1)}(mzz) A \Delta_z^y)$ (23a)
	Otherwise	${}^{(1)}(cp)_{(mzz)A} \Delta_z^w$ (23b)
> 1		$\max_{x \in O(A)} (1 \dots (n-1) {}^{(cp)}_{(mzz)D} \Delta_x^w + {}^{(cp)}_{(mzz)A} \Delta_z^x)$ (24)

For case $n = 0$, justification is the same as for case $n = 0$ for the strict while loop. For case $n = 1$, the proof for Equations (23a) and (23b) is similar to the proof of Equations (12a) and (12b) of Table 1. However, none of the roles may start their execution in A until the choice is made by the roles in C^* . For the ending event of z to occur, dependencies from all of the starting events of A to the ending event of z must occur, and no action in A may start until the choice is made by all roles in Choice collaboration. If w belongs to C^* , then dependencies from the starting events of all the roles to the ending event of z must be satisfied and, therefore, delays from the starting events of all the roles to the ending event of the desired role in A are considered. If w does not belong to C^* , the only dependency remaining is from the starting event of w to the ending event of z and, hence, only the delays between these two events are considered.

Case $n > 1$ is calculated using a recursive equation, where ${}^{(n)}(cp)_D \Delta_z^w$ denotes the NETD for the n^{th} iteration of the composite collaboration D . Figure 5 shows the 1 to $(n-1)$ iterations of collaboration A can be abstracted by $1 \dots (n-1) {}^{(cp)}_D \Delta_x^w$. To calculate the n^{th} iteration of collaboration A , first the previous $(n-1)$ iterations with the delay $1 \dots (n-1) {}^{(cp)}_D \Delta_x^w$ are considered, which are weakly sequenced with the n^{th} iteration with the delays ${}^{(n)}(cp)_D \Delta_x^w$. The equation for weak sequencing of these delays is discussed in the study by Israr [1].

Consideration of Multiple Control Flow Paths: Stochastic Delays

In this case, all of the possible control flow paths are considered and assume that each leads with probability p to the execution of A and probability $q = (1-p)$ to the termination of the loop. Based on the equations for fixed delays, the NETD for collaboration D , ${}^{(cp)}_{(stoc)D} \Delta_z^w$ can be calculated by Equation (25):

$$q \sum_{i=1}^n p^i * {}^{(cp)}_{(i)} * {}^{(cp)}_{(stoc)D} \Delta_z^w \quad (25)$$

where, there is the probability q that there is no dependency from w to z , when $w \notin C^*$.

This happens when $n = 0$. The proof is very similar to the proof of Equations (19) and (20). If the role w does not belong to C^* , there exists a probability q for no dependency from w to z . This can happen when $n=0$.

Conclusions

Israr's [1] method of representation was used to model collaborations and analyze various scenarios. The delays of composite activities sequenced with strict and weak while loops were considered, and the delays were calculated for such composite activities as well as for the individual components involved. This approach to the performance modeling of distributed system designs can be useful in many fields of application, including performance analysis of cloud computing, big data as well as distributed workflow management systems, e-commerce applications, and/or Web services. Also, a tool was implemented that takes as input an Activity Diagram with defined performance characteristics and provides outputs as the NETDs of the global collaboration for fixed delays. Even though this work was quite mathematical and proofs were provided throughout the paper, it would be beneficial to illustrate this research with an industrial case study such as the MapReduce example discussed earlier.

References

- [1] Israr, T. (2011) Performance modeling of distributed activity services. *Proceedings of the 2nd ACM/SPEC International Conference on Performance Engineering*, (pp. 475-480). Karlsruhe, Germany.
- [2] Israr, T., & Bochmann, G. V. (2013). Stochastic Performance Analysis of Distributed Activities. *Proceedings of the 5th International Workshop on Non-Functional Properties in Modeling: Analysis, Languages and Processes*, (24-31). Miami, FL.
- [3] Garcia, A. (2015, July 16). Amazon "Prime Day" shattered global sales records. *CNNMoney*. Retrieved from <http://money.cnn.com/2015/07/15/news/amazon-walmart-sales/>
- [4] MapReduce tutorial. (2013, August 4). Retrieved from https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html
- [5] Bochmann, G. V. (2008). Deriving component designs from global requirements. *Proceedings of International Workshop on Model Based Architecting and Construction of Embedded Systems*, (pp. 55-69). Toulouse, France.
- [6] Wang, Y., Lin, C., Ungsunan, P., & Huang, X. (2011). Modeling and survivability analysis of service composition using stochastic Petri nets. *The Journal of Supercomputing*, 56(1), 79-105.
- [7] Li, J., Fan, Y., & Zhou, M. (2004). Performance modeling and analysis of workflow. *IEEE Transactions on Systems, Man and Cybernetics—Part A*, 34(2), 229-242.
- [8] Hao, J., & Pei-an W. (2007). An approach for workflow performance evaluation based on discrete stochastic Petri net. *Proceedings of the IEEE International Conference on e-Business Engineering*. Hong Kong, China.
- [9] Lazowska, E., Zahorjan, J., Graham G., & Sevcik, K. (1984). *Quantitative system performance: computer system analysis using queuing network models*. Upper Saddle River, NJ: Prentice Hall.
- [10] McNeile, A. (2013). Using Motivation and Choreography to Model Distributed Workflow. *Proceedings of the 5th ACM SIGCHI Annual International Workshop on Behavior Modelling—Foundations and Applications*, (1-11). Montpellier, France.
- [11] Sahner, R., & Trivedi, K. (1987). Performance and reliability analysis using directed acyclic graphs. *IEEE Transactions on Software Engineering*, (10), 1105-1114.

Biography

TOQEER ISRAR is an assistant professor at Eastern Illinois University. He earned his BEng and MASc in Electrical Engineering degrees from Carleton University, Ontario, Canada, and a PhD in Electrical and Computer Engineering, 2014, from the University of Ottawa, Ontario, Canada. He is an internationally recognized expert in the areas of performance and software engineering and has over five years of experience in industry. He is a registered professional engineer in Ontario, Canada. Dr. Israr may be reached at taisrar@eiu.edu

DESIGNING STANDALONE MICROGRID AND GRID-CONNECTED SMARTGRID HYBRID SOLAR/WIND ENERGY SYSTEMS

Dominick Lauria, University of Hartford; Sarah Lamb, University of Hartford; Akram Abu-Aisheh, University of Hartford

Abstract

The design and implementation of microgrid technology is a leading trend in modern energy management. A microgrid is a multi-input, standalone energy system that is capable of operating in parallel with or independent from the main electrical grid. The prefix “micro” can be misleading, however, as most current technology can be scaled with reasonable precision to meet the energy demands of the end consumer. Modern microgrids take in energy from a variety of sources—in this case, sunlight, wind, and a main electrical grid – and facilitate not only its conversion into electrical energy, but also the demand management, storage, and generation associated with the system’s output. This research project focused on the design and simulation of a 48V standalone microgrid that was supplied primarily by photovoltaic (PV) panels and a wind turbine, but which also had the capability to tie in to a main electrical grid. A system of this size should be able to supply power for up to two average-size homes or office buildings. The most important objectives of this project were the selections of an appropriate PV array and wind turbine, the selection or design of a charge controller, and the design of the system’s renewable energy converter.

Introduction

Figure 1 shows a basic overview (block diagram) of the intended microgrid system. For the purposes of this project, the most significant parts of the system were the PV array, wind turbine, and renewable energy converter blocks. The project focused on the design considerations and specifications for a grid-connected PV-wind hybrid system with battery back-up. PV panels are photoelectric devices that take in light energy from the sun and convert it into a DC voltage output.

This output voltage can be regulated by applying it to a DC-DC converter, which will adjust the magnitude of the voltage. In the case of a PV panel specifically, any DC voltage output can be “dumped” into a battery bank and stored for future use. The direct or DC-DC converted output voltage can also be applied to an inverter, which would yield an overall AC output voltage.

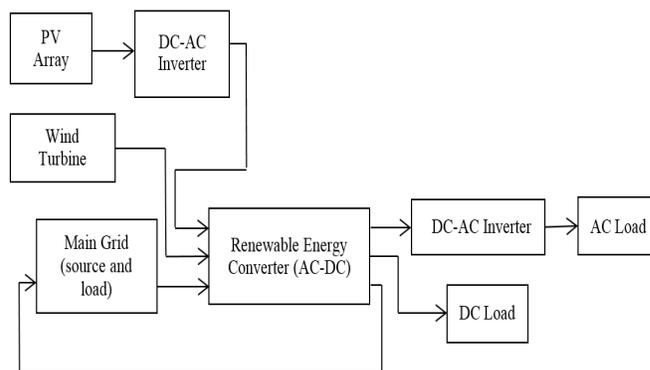


Figure 1. Block Diagram of a Smart Microgrid

Wind turbines are electromechanical devices that transfer the mechanical energy of a spinning rotor into an AC voltage output. Since the generated voltage is AC, a transformer can be used to step-up or step-down the voltage in order to obtain the desired voltage level to power the load. PV panels and one or more wind turbines were selected from commercially available models in order to meet the design specifications of the project. However, the renewable energy converter had to be uniquely designed and configured to meet the exact parameters of all other system components. For a multi-source renewable energy system, such as the one studied here, the renewable energy converter is often designed to include the DC-AC inverter and the DC-DC converter—which are separated in the block diagram—into one coherent circuit.

There are two likely configurations for this system’s renewable energy converter: a Cuk converter paired with a SEPIC converter or a boost converter paired with an inverter. The advantages, disadvantages, and standard performance of each type of converter were analyzed, and the results of these analyses applied to the design of a converter tailored to the exact needs of the microgrid. Once the components of the microgrid system have been selected (PV array and wind turbine) and designed (renewable energy converter), the system was analyzed—both as a whole and as the sum of its parts—using Matlab and Simulink software. Simulations were focused on input and output power, such that the overall efficiency could be calculated in each case.

Flat-Panel Photovoltaic System Considerations

A flat panel photovoltaic (PV) system will generate DC electricity in direct proportion to the amount of surface area that is exposed to sunlight. Modules are designed to supply this electricity at a certain voltage; however, the current produced is directly dependent on how much light is absorbed by the system. Any desired combination of voltage and current can be produced by connecting an array of panels in a series or parallel topology. The PV panels themselves are comprised of modules, which are, in turn, comprised of individual PV cells. Each PV cell is actually a very thin semiconductor wafer of photosensitive silicon or selenium that has been “doped” with boron (positively-charged, p-type material) and phosphorous (negatively-charged, n-type material) to increase its electrical conductivity to a level that is sufficient for the cell to distribute charge and induce an electric field. By themselves, PV modules or arrays do not represent an entire PV system. Systems also include structures that point them toward the sun and components that take the DC electricity produced by the modules and “condition” that electricity, usually by converting it to AC electricity. PV systems may also include batteries and/or back-up generators. These items are referred to as the balance of system (BOS) components. Combining PV modules with BOS components creates an entire PV system. A system is usually everything needed to meet a particular energy demand, such as an industrial appliance, the lights in a home, or—if the system is large enough—the electrical demand of an entire community. A BOS may also include any or all of the following: a renewable energy credit revenue-grade meter, a maximum power point tracker (MPPT), a battery system and charger, a GPS solar tracker, energy management software, solar irradiance sensors, and an anemometer.

Flat-panel systems account for the majority of renewable energy installations in the U.S. At any given location on a clear day, the amount of sunlight striking the earth’s surface is equivalent to approximately 1000 watts of power per square meter [1]. A single, flat panel is composed of at least 600 PV cells and can produce between 5V and 300V of electrical power depending on sunlight exposure. One panel alone is rarely sufficient for home or industrial applications, so multiple panels are wired together in series, parallel, or series-parallel topologies to create the large PV arrays that most people associate with solar power. Most commercially available PV systems can deliver voltage magnitudes in multiples of 12V, and current magnitudes in multiples of 3 amps. Roof installations of PV arrays—facilitated by simple and versatile mechanical brackets—are most common among home and business consumers, although an increas-

ing number of business/industrial consumers are installing solar panels on large tracts of open land, such as land that has been contaminated and cannot be developed.

The performance of a PV system module is measured as a function of its solar cell electrical performance, degradation factors associated with array design and assembly, environmental versus operating temperatures, and array power output capability. The industry standard is to report a PV module’s performance as a peak watt rating. The peak watt (Wp) rating is determined by measuring the maximum power of a PV module under laboratory conditions of relatively high light, favorable air mass, and low cell temperature. But these conditions are not typical in the real world. Therefore, researchers may use a different procedure, known as the normal operating cell temperature (NOCT) rating. In this procedure, the module first equilibrates with a specified ambient temperature so that maximum power is measured at a nominal operating cell temperature. This NOCT rating results in a lower watt value than the peak-watt rating, but is probably more realistic. However, neither of these methods is designed to indicate the performance of a solar module under realistic operating conditions. Another technique, the AMPM Standard, involves considering the entire day rather than peak sunshine hours. This standard, which is intended to address the practical user’s needs, is based on the description of a standard solar global-average day (or a practical global average) in terms of light levels, ambient temperature, and air mass.

The feeding of electricity from a PV system back into the grid requires the transformation of DC power into AC by a synchronizing grid-tied inverter. Modern inverters used in this capacity are quite effective. They allow the PV array to operate at the maximum power point (MPP) under all conditions; they generate AC output current in phase with the AC utility grid voltage; and, they achieve a power conversion efficiency of nearly 100%. In some cases, the inverter even provides energy storage to balance the power difference between the PV array (DC power) and the AC (time-domain) power of the grid.

Wind Turbine System Considerations

A wind power system relies on the fluid flow of air to apply a force on its rotor blades, causing the turbine to rotate; the system will then convert the rotational kinetic energy of the turbine into DC electricity via an electric generator. The two critical factors for power generation are wind speed and the quality of wind. Environmental (buildings) and atmospheric (turbulence) factors can interfere with the available wind; thus, wind turbines are most efficient when constructed in elevated, open areas.

The primary difference between wind and solar systems is that wind systems convert pure mechanical (kinetic) energy into electrical power, whereas solar systems rely on chemical reactions and thermal properties to generate electrical power. Consequentially, the physical design of a wind turbine is far simpler than something like a PV array. A wind turbine's most visible components are its blades, which are aerodynamically designed to capture the maximum amount of the wind's kinetic energy. The blades turn a rotor, which, in turn, rotates a shaft. Ultimately, the most important part of any wind turbine is its generator, which is driven by the shaft and functions similarly to an electric motor. The generator consists of a rotor and a stator; the wind turbine's shaft is connected to the rotor such that it causes the rotor to spin, creating (inducing) a rotating magnetic field within the stator (stationary portion of the motor). This induced magnetic field (B) effectively rotates the north-south poles of the stator, which "pulls along" the loops on the armature (rotor) windings, ultimately causing the armature to "follow" the rotation of the field and create an electromotive force (E) that is harnessed as electrical power.

An AC generator will produce AC electricity that can be directly transferred into an electrical grid for consumption. A DC generator will produce DC electricity that must be inverted into AC before it can be transported or consumed. Not only is AC power the primary source for home and business/industrial consumption, but it is also much easier and safer to transport over power lines with a higher efficiency than DC power. Depending on the needs of the consumer, the final AC output of a wind turbine system will pass through an inverter to stabilize its voltage, current, and frequency to the local electrical grid's standard, and will then be connected to the grid in one of three ways: 1) grid connection only, 2) grid connection with battery back-up, or 3) grid connection with generator back-up.

When a wind turbine is directly connected to the grid (option 1), the grid will become the consumer's primary source of electricity on days with very little wind. The addition of a battery back-up (option 2) allows a certain amount of the turbine's excess energy to be stored in the batteries as DC potential energy, but requires the inclusion of a controller unit between the turbine's output and the batteries to regulate the flow of current to and from the batteries. A back-up generator (option 3) can be used in place of or in conjunction with a battery back-up; the generator's main purpose is to keep the battery pack charged during periods of main grid failure (e.g., utility power outage) and no wind. Depending on the generator's configuration, it could also be used as the home's or business' chief source of power during an emergency.

There are two major challenges to using wind power systems: the supply of wind and the logistics of wind-generated electricity transportation. Wind is a naturally intermittent resource, and thus does not always blow when electricity is needed. Wind cannot be stored (although wind-generated electricity can be stored, if batteries are used) and not all winds can be harnessed to meet the timing of electricity demands. Some of the most consistent wind sites are often in remote locations, far from areas of high-demand electricity areas, so once power is generated, it must be transported over large distances and difficult terrain.

Wind turbines of any scale have relatively little impact on the environment, especially when compared to fossil fuel power plants. Consumers living or working near wind farms occasionally express concern over the noise produced by the rotor blades of a large wind turbine, the degradation of their view, and the potential for flying wildlife mortality (birds, bats, etc.). However, it is important to note that most of these issues have been resolved (or are in the process of being resolved) through technological development and improved site planning protocols for large-scale wind power operations. The major advantage of wind power applications is that they meet the basic power needs of remote areas that are not yet grid-connected. Increasingly, wind systems are being paired with PV systems to create a hybrid topology that is nearly 100% grid-independent. The benefits of hybrid systems for remote areas and/or developing regions of the world are immense.

Hybrid Energy System Considerations

When a reliable grid can be accessed from the location at which renewable energy sources are being used, it is common for excess power generated by those renewable sources to be fed into the main power grid. This allows consumers to save money on their electric bills, because they are generating power for the electric company. Feeding some amount of renewable-sourced power back into the grid is also a common practice, because batteries are one of the most expensive components in renewable systems designs. Batteries take up space, need to be properly stored, require extra circuitry for control purposes, and even after all of that, they will still eventually need to be replaced. Although grid connection is a more common practice than battery storage, there are some challenges and considerations to take into account when connecting a renewable system to the main grid. The first challenge is making sure that the hybrid system will be able to reliably output the same voltage and frequency, so as to input that voltage and/or frequency into the grid on a continuous cycle. This is especially important for the grid side of the system, because if the voltage and frequency are not what they need to be, then there will be a

loss of power quality within the grid. Another consideration is determining at which point in the grid to insert the renewable-sourced power. If a large amount of power is being generated, then transmission line insertion is the best location, because the voltage in those lines averages 500 kV, depending on where the transmission lines are located. If a small amount of power is being generated, then the distribution lines are the best place for insertion [2, 3].

It is important to note that any energy generated through small renewable energy generators cannot be directly connected to the grid. An interface is required between the generation system and the utility distribution grid. For PV systems, an inverter is required; for wind and hydro systems, an induction generator is required. Power electronic converters are also used to interface most of the distribution grids. Used alone, these converters will inject harmonics into the system, often resulting in poor power quality for the end consumers. Thus, harmonic filters must be added to overcome any harmonics. There is one notable exception to this rule: if a pure sinewave inverter is being used, then a harmonic filter will not be required, because it is already built into the inverter. In contrast, squarewave inverters and modified sinewave inverters both require harmonic filters. For induction generators, voltage flickering may occur, but can often be reduced by utilizing soft-starting mechanisms for the generator itself.

A common circuit to utilize in grid connections is a phase locked loop (PLL). PLLs are circuits that can quickly and accurately detect the phase angle of the grid voltage. By use of a PLL, proper regulation of the power flow between a renewable source and the grid can be achieved. Four industrially standardized examples of PLL circuits include: synchronous reference frame (dq PLL), stationary reference frame ($\alpha\beta$ PLL), decoupled synchronous reference frame (DSRF PLL), and decoupled stationary reference frame [3]. A major design consideration with any inverter-based distribution grid interface is to configure the interface in such a way that, if the main power grid falls offline, the inverter also shuts down. If the inverter does not shut down, it will pose a significant high-current hazard to any electrical maintenance worker attempting to restore grid power. Therefore, all sources that feed power into the main grid must be shut down without question any time the main power grid goes down, both for safety purposes and to comply with the National Electric Code's (NEC's) requirements.

Hybrid System with PV Array and Wind Turbine

In a PV-wind hybrid system, the power generated from both wind and solar components is stored in a battery bank

for later use, thus increasing the reliability of the system. In some cases, the size of the storage battery may be slightly reduced, compared to a pure-solar or pure-wind system, because the system is capable of generating power from more than one source. Wind speeds are often low in periods when the sun's resources are at their best (summer). The wind is often stronger in seasons when there are fewer sun resources (winter). Even during the same day, in many regions worldwide or in some periods of the year, there are different and opposite patterns in terms of wind and solar resources. And those different patterns can make the hybrid systems the best of both worlds for power generation. One potential drawback of hybrid PV-wind systems is that they carry a significantly higher up-front cost than pure-solar or pure-wind systems. However, the PV-wind hybrid system offers the greatest return on investment (ROI) in terms of output and performance achieved per dollar invested.

Renewable Energy Converter for PV-Wind Hybrid Systems

There are two renewable energy converter topologies that are overwhelmingly prevalent in existing industrial applications: a Cuk converter paired with a SEPIC converter and a boost converter paired with an inverter. For a Cuk-SEPIC hybrid topology, the pairing of the two converters is possible because the existing diodes are reconfigured such that the SEPIC converter shares the Cuk converter's output inductor [3]. This topology is advantageous because it does not require any low-pass filters between the DC inputs and the hybrid converter, which were utilized in previous designs to eliminate high-frequency harmonics. A boost-inverter hybrid is conceptually simpler—amplify the input DC voltage with the boost converter, and then invert it to an AC signal to drive the load—but mechanically more complex. The converter and inverter must be constructed in parallel with each other, which decreases the system's overall efficiency (measured in terms of power transfer, P_{out} / P_{in}).

An existing industrial system, similar to the block diagram shown in Figure 1, was recently patented by the U.S. Department of Energy via a research team at the University of Arkansas [4]. This system is capable of consolidating multiple levels of DC input into a single, stable AC signal. The system is revolutionary, because it uses a multiple squarewave input design and, thus, can be contained entirely in a small, high-frequency transformer. However, it has yet to be introduced to the U.S. commercial market. With consideration given to the practical advantages and disadvantages of each converter topology, as well as to the commercial availability of components in the U.S., it is advisable to use the Cuk-SEPIC converter for this and any standalone microgrid application.

Within the Cuk-SEPIC converter, it is necessary to include integrated circuits for both AC-DC rectifiers and DC-AC inverters. Depending on the characteristics of a system's input voltages (AC or DC, magnitude, and phase), rectifiers will be needed to step-up or step-down the voltage so that it is suitable to drive the load. Similarly, inverters will be needed to transform any DC signal into an output that is compatible with an AC load. There are two widely accepted rectifiers for renewable energy applications: a switching power supply and integrated microcontrollers. The latter option increases the cost of the rectifier considerably and does not yield noticeably better results, so the use of a switching power supply is recommended for most renewable energy applications.

In terms of any DC-AC conversions, the use of a full-bridge inverter is overwhelmingly favored for renewable energy applications, although a half-bridge inverter may also be used, depending on the specifications of the system and the availability of components. Any renewable energy system that includes a PV array must also include an inverter, due to the array's DC output and the strong likelihood of an AC load or main grid connection. Figure 2 shows a full-bridge inverter. Figure 3 shows how this inverter takes in a constant DC signal and generates an oscillating AC signal. For commercial applications, an AC oscillation frequency of 60 Hz is preferred. The quality of an inverter is determined by the smoothness of its 60 Hz signal, with better quality inverters yielding smoother signals. The more noise that is present in an inverter's output signal, the less power transfer that inverter can achieve. In general, a full-bridge inverter will output a higher quality signal than a half-bridge inverter, which is why full-bridge inverters are preferred for renewable energy and power electronics applications.

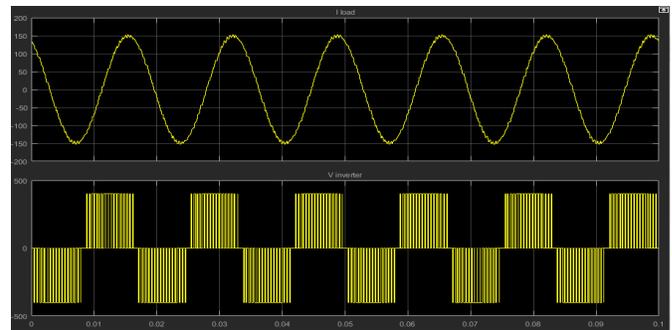


Figure 3. Simulink Scope for Full-Bridge Inverter Output

Battery Back-Up Considerations

When operating a PV, wind, or hybrid system, there are times when more power will be generated than what is needed to drive the load. In such cases, the extra power must either be fed into the main grid or captured and stored in a battery bank. Even in the case of a grid-connected system, a battery back-up is often preferred (in addition to the grid connection), because it allows the system to remain self-sufficient during non-operational hours. From a physical perspective, a battery is an electrochemical device that stores DC power and undergoes chemical reactions in order to add to (charge) or release (discharge) its initial level of electric charge. The batteries used in renewable systems are referred to as “secondary batteries,” because they are used as a secondary energy source and can easily be charged when extra power is present. It is important to distinguish these from “primary batteries,” which cannot be used for renewable systems, because they must start with a full charge and can only be discharged while the system is in operation.

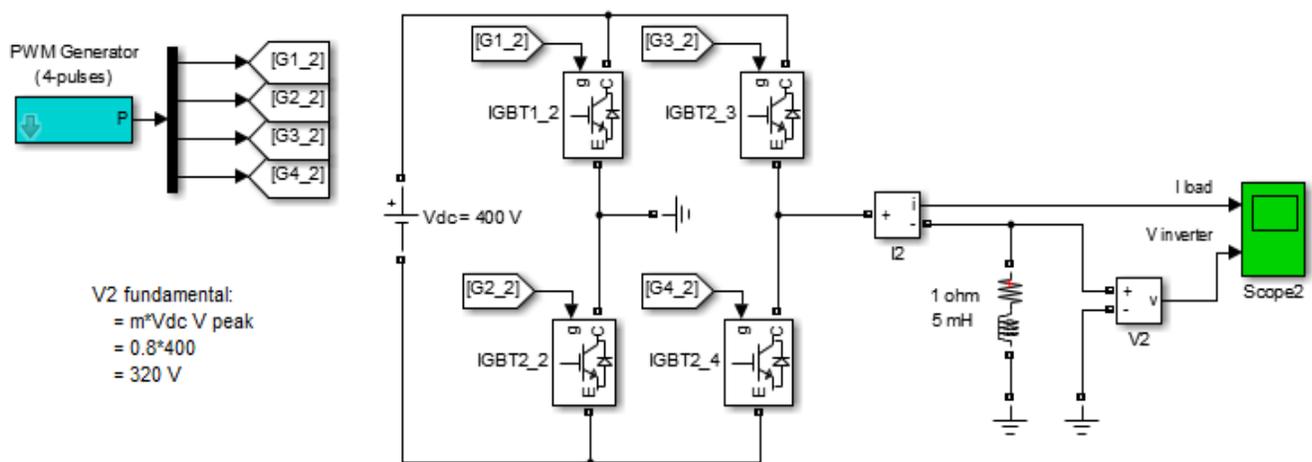


Figure 2. Simulink Model of Full-Bridge Inverter

One very important consideration for the selection of a renewable system's secondary battery is how much current the load will draw from the battery bank. The maximum current drawn cannot exceed what the battery is rated for, otherwise the chemical reactions that occur inside the battery will not be able to keep up with the current draw and the battery life will be greatly reduced [4, 5]. Other battery specifications to consider when designing a PV system include the charge/discharge cycle history, ambient temperature, and battery age. Lead-acid batteries must not be overcharged in order to avoid hydrogen particles separating from the oxygen. Overcharging will cause the battery to start gassing, which results in water loss. Water loss will decrease the battery's charging efficiency and reduce the battery's operating life. Similarly, Lead-acid batteries should not be undercharged, because undercharging makes them susceptible to freezing, which also shortens their operating life.

Any battery must be kept inside a certain temperature range to maximize its life, so long-term storage is another consideration that must be accounted for. Large batteries should be kept in a storage area where they can be heated or cooled, depending on the climate and/or season. Many PV, wind, or hybrid systems are "sized" or rated in terms of their battery capacity. Ideally, a battery bank should be sized to provide power to a load for up to five days during inclement weather conditions [4]. If the battery bank has less than a three-day capacity, the battery will be deep-cycling on a regular basis, which will shorten its operating life.

Lead-acid batteries are used most frequently in PV systems, and various types include starting batteries, RV or marine "deep cycle" batteries, lead-calcium (Pb-Ca) batteries, and true deep-cycle batteries. A true deep-cycle battery is a battery that delivers on average a few amperes of current to the load for hundreds of hours between charges. In contrast, shallow-cycle batteries deliver hundreds of amperes to a load over a very short period of time and then the battery is recharged, making them ideal for automobile applications. The true deep-cycle battery is most reliable for PV systems because, when fully charged, it is recommended to use 50% of the battery's energy to power the load before recharging [4]. This timing makes it ideal for PV applications, which rely on solar energy during the day and battery back-up at night. In case of an emergency—such as prolonged inclement weather—the battery charge may be reduced to 30%, but going below 20% too frequently will greatly reduce battery life [4].

Many systems also incorporate a charge controller to assist in controlling the charging and discharging states of a

battery. For example, a charge controller can be designed where the low-voltage disconnect (LVD) will go into effect when the battery is going to fall below a 20% charge. When the LVD is engaged, the battery will be disconnected from the load until the battery has been charged to an appropriate level. A high-voltage disconnect (HVD) also exists in charge controllers and is used to detect when the battery is fully charged. When a battery is fully charged, the HVD will go into effect and limit the amount of current that flows to the battery. True deep-cycle batteries are ideal for PV systems, because they can go through hundreds or even thousands of cycles if the battery is properly cared for.

Batteries are the key element for the storage of extra power generated from renewable energy systems, and they are critically useful in remote and rural locations, where a grid connection is not available. The inclusion of a battery bank in any system's design will ultimately save the consumer money as a long-term investment, because a fuel-consuming gas- or diesel-powered generator need not be relied upon as often in the event of bad weather. It is important to note that when selecting batteries for any PV, wind, or hybrid system, the list of possible lead-acid, true deep-cycle batteries that can be used will be further differentiated by the relative size of the system and the approximate number of kilowatt-hours (kWh) that a consumer plans to use the system for. In addition to battery storage capacity, PV system designers must be sensitive to both cost and physical space available, so it is likely that trade-offs will be made and compromises reached during the design phase of the system.

Flat-Panel PV Array Simulation

When simulating the flat-panel PV system, it is important to first consider the theory: in particular, the PV modules' equivalent circuit design, open-circuit voltage, and power and I-V curves under specified standard test conditions (STC). STC for a typical PV module is generally taken as 1000 W/m² irradiance, 25°C temperature, and 1.5 AM air mass [6]. The equivalent circuit for a standard PV module shown in Figure 4 is derived from the physics of current generation but, for the purposes of this simulation, it was most important to consider the electrical calculations relating to the actual application in real systems. The PV module's characteristic equation—which solves for the module's open-circuit voltage (V_{OC})—can be derived from the equivalent circuit shown in Figure 4. Ultimately, this derivation yields Equation (1):

$$V_{OC} = \left(\frac{nkT}{q} \right) \left(\ln \left(\frac{I_{SC}}{I_0} \right) + 1 \right) \quad (1)$$

where, n is electron density; k is the Boltzmann constant; T is temperature; q is electric charge; I_{SC} is short-circuit current; and, I_O is the output current and is determined by current density, J , times PV cell area.

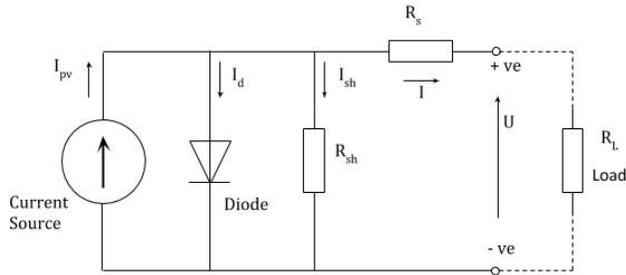


Figure 4. PV Module Equivalent Circuit

A PV system's MPP is a variable quantity that depends on solar irradiance and cell temperature. It is best demonstrated by the unique power and I-V curves of any PV system. A system's I-V curve is obtained by varying an external resistance from zero (short-circuit) to infinity (open-circuit) at STC. Power delivered by the PV cell is the product of current and voltage, so zero power is delivered at short-circuit and open-circuit points. This means that the MPP must fall between the extremes.

The Matlab simulation performed in this study utilized a pre-published, detailed example file for a grid-connected, flat-panel PV system [6]. The example file was scripted for a 100-kW PV array connected to a 25-kV grid via a DC-DC boost converter and a three-phase (3 ϕ), three-level (3L) VSC. When run in Simulink, the example file specified 66 parallel strings of five series-connected PV cells, with each cell capable of delivering a power output of 305.2W at STC. Specifications also included: $V_{OC} = 64.2V$; $I_{SC} = 5.96A$; $V_{MPP} = 54.7V$; and, $I_{MPP} = 5.58A$. For the purpose of this project, the example file was modified such that it could simulate a 40-kW array with a 25-kV grid connection. Equation (2) shows how these modifications were accomplished by varying the number of parallel strings of series-connected PV cells.

$$P_{out} = (305.2W)(\# \text{ strings})(\# \text{ series connected cells}) \quad (2)$$

$$40,000W = (305.2W)(\# \text{ strings})(5 \text{ cells}) \rightarrow 26.21 \text{ strings}$$

Following this formula, the final simulation modeled 26 parallel strings, each of which contained five series-connected PV cells. Once parameters were established at STC, it was possible to ascertain the key values for the 26 \times 5 array, using the model given in Figure 5.

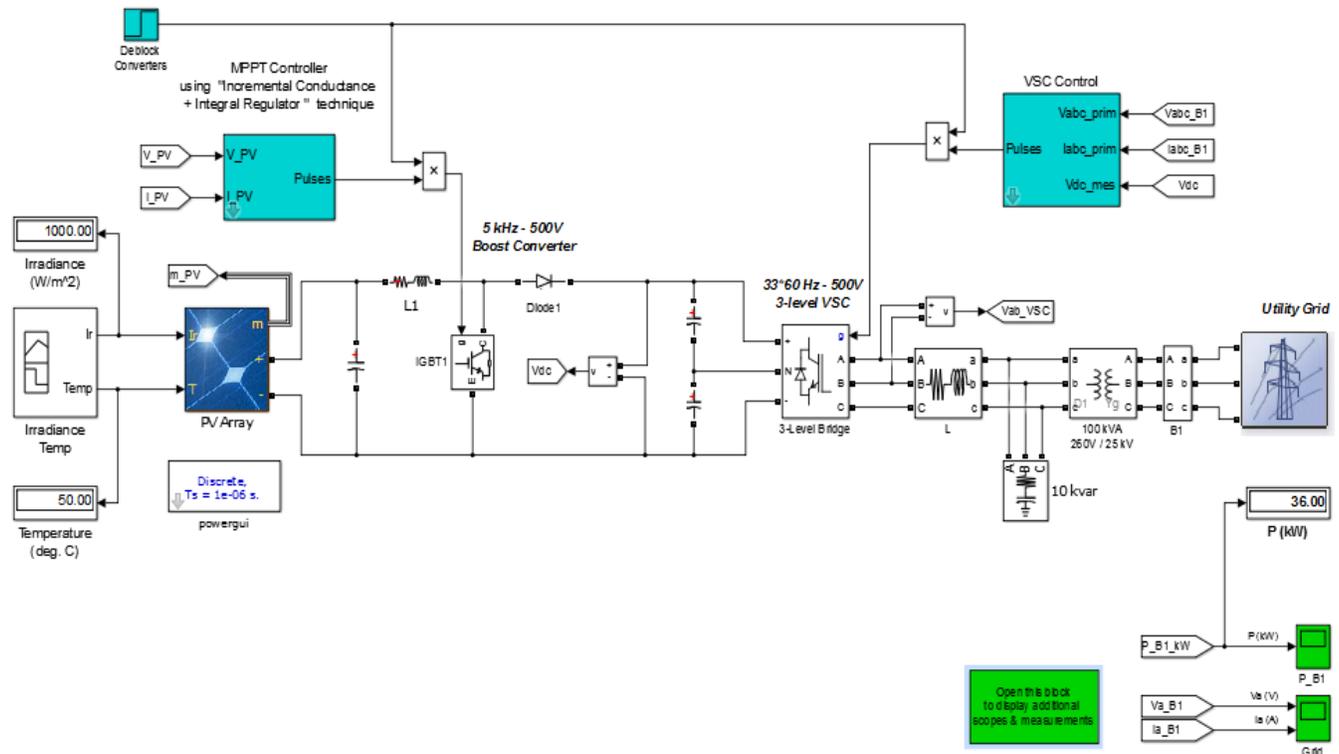


Figure 5. Final 40 kW PV Array with 25 kV Grid Tie-In

Figure 6 shows that, over a period, T , of 2.5 seconds, the average output power generated by the system was approximately 36.0 kW. This makes sense, as the measurement was taken between the boost converter and the grid; the grid can absorb 25.0 kW of power from the system, so the excess 11.0 kW in this example was wasted. To harness this excess power, the system could be expanded to include a battery back-up.

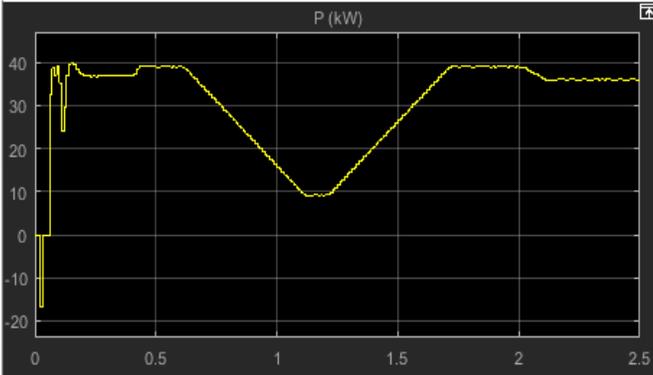


Figure 6. Simulink Scope for P_{out} at $T = 2.5s$

Furthermore, the simulation demonstrated a periodic (sinusoidal) current and a pulsating voltage at the VSC controller. This also makes sense, because the VSC block regulates the MPP of the system via pulse width modulation techniques that are applied as the three phases of voltage come off of the three-level bridge. The experimental (traced) I_{MPP} value is easily verified by the system's characteristic equation, given as Equation (3), and a basic percentage difference calculation.

$$\text{original } I_{MPP} = 5.58A = (\text{current density } J)(66 \times 5) \quad (3)$$

$$\therefore \text{current density } J = 0.0169$$

$$\text{new } I_{MPP} = (0.0169)(26 \times 5) = 2.198A$$

$$\% \text{ difference} = \frac{|2.198 - 2.199|}{2.198} \times 100 = 0.0455\%$$

Having verified the value of I_{MPP} given by the simulation, it was possible to further verify the values of V_{MPP} and P_{in} using Equations (4) and (5):

$$V_{MPP} = \frac{P_{MPP}}{I_{MPP}} = \frac{36kW}{2.199A} = 16.4kV \quad (4)$$

$$\eta = \frac{P_{MPP}}{P_{in}} \rightarrow P_{in} = \frac{P_{MPP}}{\eta} = \frac{36kW}{0.25} = 144.0kW \quad (5)$$

For these calculations, the simulation value of $P_{MPP} = 36.0$ kW is used, which differed from the desired system's $P_{MPP} = 40.0$ kW by 4.0 kW, due to the rounding-down of the number of PV cells in the array. Initial calculations indicated a need for 26.21 strings, but an integer number of 26 strings was used for the final model. Additionally, these calculations assumed the industry standard PV cell efficiency of $\eta = 25\%$, which was effectively determined for 2015 by the National Renewable Energy Laboratory [4].

Wind Turbine Simulation

An ideal wind turbine simulation begins with the theory—in this case, for a horizontal-axis turbine with generator back-up that is connected to a utility grid. The turbine's most important characteristic is the output power that it can generate, given by Equation (6):

$$P_{out}[kW] = \frac{1}{2}(kC_p)(\rho AV^3) \quad (6)$$

where, k is a constant (0.000133); C_p is the maximum power coefficient ($0.25 < C_p < 0.45$, theoretical maximum at 0.59); ρ is air density in lbs/ft^3 ; A is the area swept out by rotor blades in ft^2 ; and, V is wind speed in mph.

The portion that is available as input power to the turbine is expressed as Equation (7):

$$P_{in}[kW] = \frac{1}{2}(\rho AV^3) \quad (7)$$

This, in turn, influences the maximum possible efficiency of the turbine, which is calculated as a ratio of the output and input power, or as a ratio of the air speed on either side of the turbine blades. Assuming the general case of a horizontal-axis machine of unknown height, with three blades spaced at 120° intervals, all with negative blade tilt angles, the efficiency is calculated using Equations (4), (5), and (8):

$$\eta = \frac{v_{air, out of turbine}}{v_{air, into turbine}} = \frac{P_{out}}{P_{in}} = \frac{\frac{1}{2}(kC_p)(\rho AV^3)}{\frac{1}{2}(\rho AV^3)} = kC_p \quad (8)$$

$$kC_p \approx 0.33 = 33\%$$

Thus, the maximum possible efficiency of a generalized horizontal-axis wind turbine is significant, because it does not depend directly on the wind's velocity and cannot exceed 33%. In specific cases when the exact dimensions and specifications of a wind turbine and its environment are known, the constant, k , is dependent on blade pitch angle and wind speed, via the turbine's tip speed ratio. Tip speed

ratio is the ratio of the rotor blade tip speed to the wind speed. The system's efficiency for this specific case is calculated using the following Equation (9) and several substitutions based on Equations (4), (5), and (8):

$$k = \lambda\beta = \left(\frac{v_{\text{rotor blades}}}{v_{\text{wind}}} \right) (\text{blade tilt angle [deg]})$$

$$0.25 < C_p < 0.45(\text{max}@0.59) \quad (9)$$

$$\eta = kC_p = \frac{(v_{\text{rotor blades}})(\beta)(C_p)}{v_{\text{wind}}}$$

The specific case efficiency is significant, because it takes into account all factors that could influence the wind turbine's functional capabilities. As the wind speed increases, the rotor speed increases, thus increasing the efficiency of the turbine. Greater wind speeds can be found at higher altitudes, which accounts for the increasing tower heights of many commercial wind turbines. The conditions of the specific case system yield a detailed power curve that is widely accepted for all commercial wind turbine applications.

The Matlab simulation performed for the wind turbine system utilized a pre-published, detailed example file [6]. The example file was scripted for a horizontal-axis wind turbine with generator back-up that was connected to a utility grid at 75 kW. The simulation was run on the example file without modifications; the desired power output, P_{out} , was equal to the nominal mechanical output power of 1.5 MW. Over a period, T , of 60 seconds, the wind speed followed an exact parabolic fit of

$$y = \frac{-1}{20}(x - 30)^2 + 25$$

meaning that the maximum wind speed of 25 m/s was reached at $\frac{1}{2}$ of the period (30s) and the power output was a step function relative to wind speed. Because the simulation assumed the general model of the wind turbine, it was presumed that its maximum possible efficiency would be 33%. The efficiency of the generator was calculated using Equation (10) [6, 7]:

$$\%Efficiency = \frac{P_{\text{Generator}}}{P_{\text{turbine}}} \times 100 \quad (10)$$

Solved for this specific case, Equation (10) yields an efficiency of 90%. Furthermore, Equation (10) can be combined with the turbine's efficiency via Equation (8) and the trace measurements on the wind speed curve in order to calculate both the maximum speed of the wind exiting the turbine and the maximum input power of the turbine at its peak efficiency. The calculation for this specific case is:

$$\eta : (0.33)(25\text{m/s}) = 8.25\text{m/s} = v_{\text{air out of turbine}}$$

$$\eta : \frac{150\text{kW}}{0.33} = 454.5\text{kW} = P_{\text{in}}$$

It is difficult to calculate any other characteristic properties of the wind turbine (tip speed ratio, area swept out by rotor blades, etc.) without having more data regarding the machine's physical features. If more data were known, it would be possible to employ the specific case equations for turbine efficiency, which could ultimately lead to a more thorough evaluation and verification of the simulated system.

Hybrid System Simulation with Grid Connection

The hybrid system operates on dual three-phase, 60 Hz input signals. The wind turbine inputs an AC signal, while the PV array inputs a DC signal that must be inverted prior to the load bus. Figure 7 shows how the two AC signals are joined at the load bus, which feeds into the main grid and/or an alternate AC load. The hybrid model outputs a significant amount of data to the Matlab workspace, which is vital for system analysis and proof of concept before construction. To prove the hybrid model's precision, separate simulations were run for the PV array and the wind turbine, with each being simulated under standalone and grid-connected conditions. In all four of these cases, the Simulink model compiled and the simulation ran smoothly without errors. The four individual results verify the precision of the hybrid system, if the hybrid system is considered as the sum of its parts.

Conclusions

The project accomplished its main objectives for background research and simulation of PV, wind, and hybrid renewable energy systems. Each system was designed as a microgrid that is capable of operating in parallel with or independent from the main electrical grid. A connection to the main grid was not required for this project, but it was simulated for each system (PV, wind, and PV-wind hybrid) in order to thoroughly verify the integrity of each system's design. The simulation results for the PV and wind systems were conclusive. The 40-kW PV system yielded $I_{\text{MPP}} = 2.198\text{A}$, $V_{\text{MPP}} = 16.4\text{ kV}$, and $P_{\text{MPP}} = 36.0\text{ kW}$, under the assumed conditions of 25% efficiency for all PV cells. The 150-kW (P_{out}) wind system yielded wind speeds of 17.68 m/s (in) and 8.25 m/s, and an input power of 454.5 kW. When the PV side of the system was separated from the wind side of the system, and when each was run as

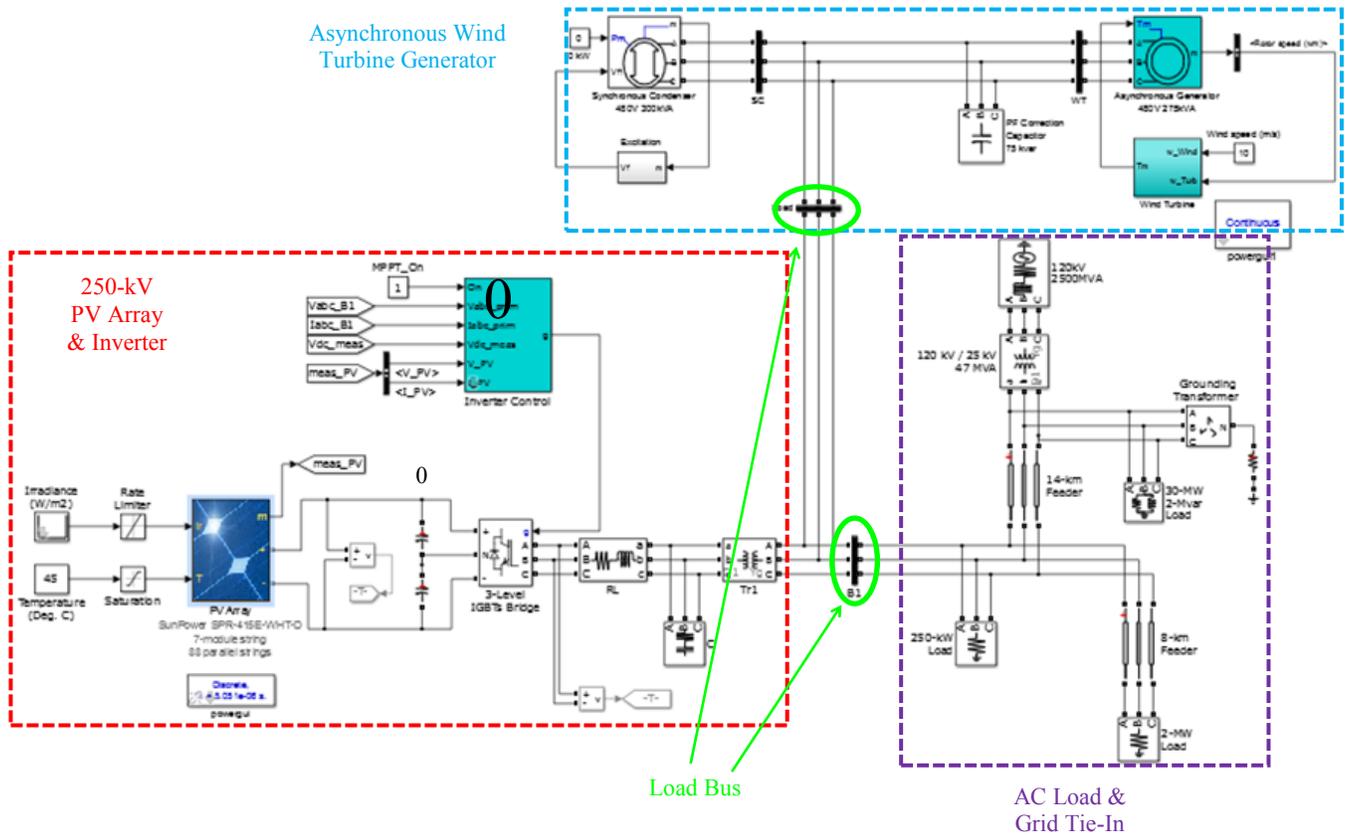


Figure 7. Complete Simulink Model of the Hybrid System

its own simulation, the two simulations yielded conclusive results similar to those of the standalone systems. This proves that the hybrid system model is complete and would yield a conclusive result, if it were to be run on a full license of Matlab.

The hybrid system simulation results dictate the specifications of components that must be chosen in order for the system to be constructed and implemented. The most important (and most restrictive) of these components are the PV array, wind turbine, and renewable energy converter. Based on these results, a set of assumptions could be made about the system's operating conditions, which would allow for components to be chosen for a hypothetical construction.

References

- [1] Sumathi, S., Kumar, L. A., & Surekha, P. (2015). *Solar PV and Wind Energy Conversion Systems: An Introduction to Theory, Modeling with MATLAB/SIMULINK, and the Role of Soft Computing Techniques*. (1sted.). Springer International Publishing.
- [2] Hart, D. W. (1996). *Introduction to Power Electronics*. (1sted.). Prentice Hall.
- [3] Rashid, M. H. (2014). *Power Electronics: Circuits, Devices, and Applications*. (4thed.). Pearson Educational Press.
- [4] Hybrid Wind and Solar Electric Systems. (n.d.). Retrieved from <http://energy.gov/energysaver/hybrid-wind-and-solar-electric-system>.
- [5] PV Battery: Photovoltaic System Component. (n.d.). Retrieved from <http://www.solardirect.com/pv/batteries/batteries.htm>
- [6] Three-Phase Asynchronous Wind Turbine Generator. (n.d.). Retrieved from <http://www.mathworks.com/help/physmod/sps/examples/three-phase-asynchronous-wind-turbine-generator.html>
- [7] Reid, R., Saulnier, B., & Gagnon, R. (n.d.). Wind Turbine Asynchronous Generator in Isolated Network. Retrieved from <http://www.mathworks.com/help/physmod/sps/examples/wind-turbine-asynchronous-generator-in-isolated-network.html>

Biographies

DOMINICK LAURIA is currently a graduate student and lab instructor in electrical engineering at the University of Hartford. He earned a BS degree in Audio Engineering Technology from the University of Hartford. He has one year of industry experience in audio and electronic equipment repair. He may be reached at dlauria@hartford.edu

SARAH LAMB is currently a graduate student in electrical engineering and business administration at the University of Hartford. Ms. Lamb earned a BS degree in physics from the University of Connecticut and taught AP/IB physics and robotics extensively at the secondary school level before returning to graduate school full-time. She is also a professional race mechanic with USA Cycling. She may be reached at slamb@hartford.edu

AKRAM ABU-AISHEH is an associate professor of Electrical and Computer Engineering at the University of Hartford. Professor Abu-aisheh is a senior IEEE member and he has ten years of industry experience in the area of fiber optic telecommunication systems and power electronics. Professor Abu-aisheh's research interests include optical communications and power electronics. Professor Abu-aisheh has MS and BS degrees in Electrical Engineering from the University of Florida and a PhD in Electrical Engineering from the Florida Institute of Technology. Professor Abu-aisheh may be reached at abuaisheh@hartford.edu

INVESTIGATING THE EFFECTS OF STEREO CAMERA BASELINE ON THE ACCURACY OF 3D PROJECTION FOR INDUSTRIAL ROBOTIC APPLICATIONS

Wutthigrai Boonsuk, Eastern Illinois University

Abstract

Stereo cameras have been used in robotics for decades. With recent advances in sensor technology, they have become one of the most widely used sensors for automation in robotic applications. Fulfilling the need for rapid detection and recognition of complex environments, stereo cameras allow robots to quickly sense an environment by reconstructing 3D scenes from disparity maps. However, typical stereo cameras have a fixed baseline, which lessens the quality of disparity maps when objects get too close to the cameras. The baseline of a stereo camera needs to be adjusted to accommodate the minimum range between the cameras and the object. As the accuracy of 3D scenes can be essential for certain robotic tasks, such as welding, riveting, measuring, and assembly, the focus of this current study was on investigating the effects of different stereo camera baselines on the accuracy of 3D projections generated from disparity maps. The results showed a correlation between stereo camera baseline and valid surface areas of the target object. This finding can be useful for researchers wanting to design and develop an effective stereo camera system and improve the quality of its 3D projection. Although this current study focused on a stereo vision system for industrial robotic applications, the results may also be applied to other robotic applications such as navigational robots and autonomous vehicles.

Introduction

Although stereo cameras have been utilized in the robotics field for decades, only recently have they become one of the most widely used sensors for automated robotics applications. Stereo cameras allow robots to sense their surroundings by reconstructing 3D scenes of the environment from disparity maps. Despite their wide use, however, typical stereo cameras have a fixed baseline—the distance between the two cameras—so disparity maps become distorted when objects get too close to the cameras. Figure 1 illustrates a typical configuration of a stereo camera system with a fixed baseline. In order to reduce this problem, the baseline of the stereo cameras needs to be shortened in order to accommodate the minimum range between the cameras and the object. In this study, the author investigated the effects

of different stereo camera baselines on the accuracy of 3D reconstruction of scenes from disparity maps. This information is crucial for finding the optimal baseline of a stereo camera system when a fixed baseline is used, and as a first step towards the development of a flexible baseline stereo camera system.

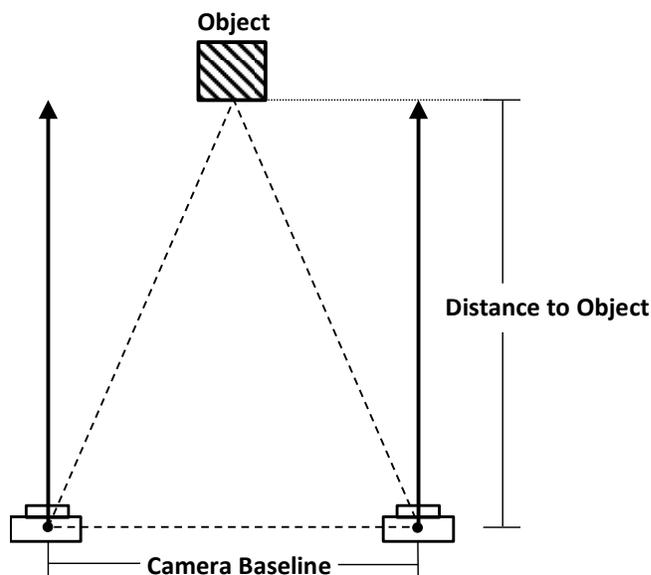


Figure 1. Stereo Camera System with a Fixed Baseline

The fundamental idea behind this study stems from the active stereo vision approach in robotics applications. This approach is a framework of stereo vision that manipulates camera parameters, lighting conditions, and motor controllers [1-3]. Conventional stereo cameras for robotic arms have a fixed baseline length [4], which can cause problems when the object gets too close to the cameras. Multi-baseline stereo cameras have been introduced in earlier research as superior to a set of stereo cameras with different baseline lengths [5]. By using a multi-baseline and multi-resolution approach [6], the depth error can be kept constant because the camera baseline and resolution can be varied in proportion to the depth. Rather than capturing a series of images from stereo cameras, another approach [7] is a stereo system that utilizes a high-speed slider to adjust the baseline length of the stereo camera. A similar method for constructing a stereo camera system was utilized in this

present study. In a previous study by Rovira-Mas et al. [8], the authors identified the best combination of baseline length and focal length lenses suitable for agricultural robotic vehicles with a working range of 6-12 meters; however, past studies have not examined the relationship between baseline adjustment and the accuracy of the 3D projection from the cameras, and have tended to pick the best combination from the pool. The purpose of this current study was to extend prior work by investigating the relationship between baseline adjustment in stereo cameras and the accuracy of the 3D projection from the camera. The findings are intended for use in industrial robotic arm applications.

In this study, the author analyzed the relationship between three parameters: 1) baseline of the stereo camera system, 2) distance to object, and 3) accuracy of the 3D scenes constructed from the disparity map. The accuracy of the disparity map is essential for building a high-quality 3D surface, particularly for automated robotic tasks that require high precision, such as welding, riveting, measuring, and assembly tasks. Thus, a set of experiments was designed in this study to identify how different stereo camera baselines may influence the disparity map when the distance to the object varies.

Methodology

In this study, a custom-built stereo camera rig consisting of two Raspberry Pi cameras (5 megapixels; focal length = 3.6 mm) was used. Figure 2 shows these cameras connected to two Raspberry Pi boards (model B+), mounted on linear motor screws that allow the cameras to be separated from 30-120 mm (1.1-4.7 inches). An Arduino Pro Mini (microcontroller board) was used to control the rotation of the motors in order to adjust the baseline. Stereo video signals were wirelessly transmitted via Wi-Fi. On a desktop computer, OpenCV and Point Cloud Library (PCL) were used to analyze real-time video signals and develop disparity maps using the Semi-Global Block Matching (SGBM) algorithm. A series of experiments was conducted in the laboratory. Figure 3 shows that the centers of the cameras were separated (i.e., baseline) at 15 mm increments (30, 45, 60, 75, 90, 105, and 120 mm).

This camera system was planned to be implemented on a FANUC LR Mate 220iC, which has a maximum arm reach of approximately 700 mm (27 inches). Thus, Figure 4 illustrates that the object was first placed at 700 mm from the cameras and moved closer in increments of 100 mm, up to 250 mm, which was the minimum arm reach of this robot. Figure 5 shows the target object, a 30x30 mm plastic cube with a 20-mm diameter hole. This target has typically been used for a robot gripping test in the lab and to test the accu-

racy of the disparity map. The disparity map was generated for each pair of baseline and object distances. The 3D scene was then reconstructed and analyzed for its accuracy. The higher the surface area that the camera picked up, the more effective the baseline would be at that specific distance to the object.

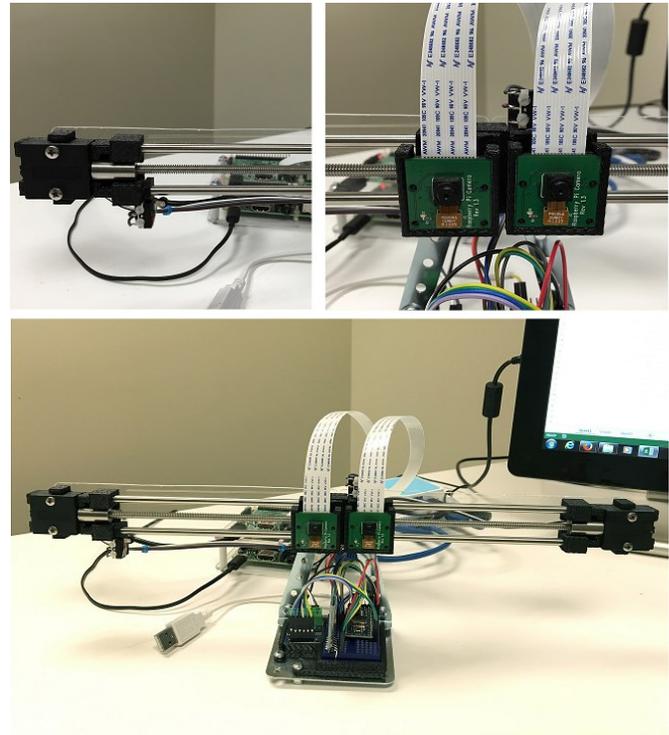


Figure 2. Custom-built Stereo Camera System

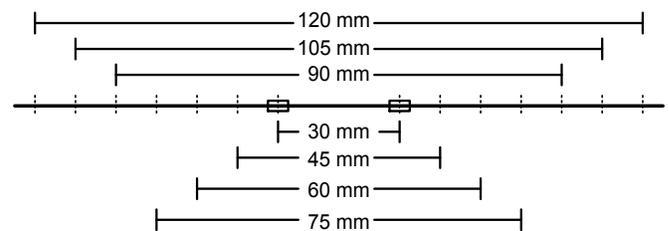


Figure 3. Baselines Used in the Experiments

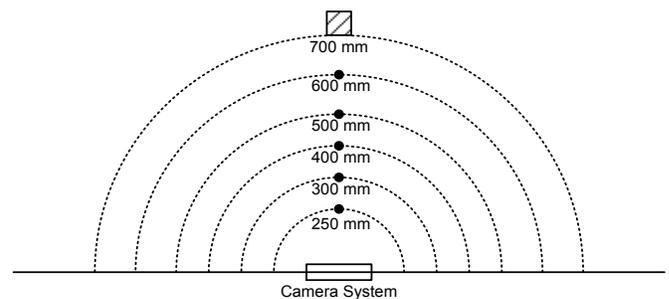


Figure 4. Distance to Target Object from the Camera System

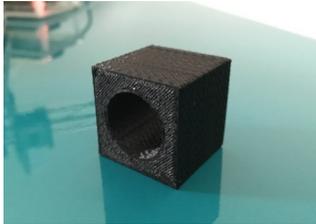
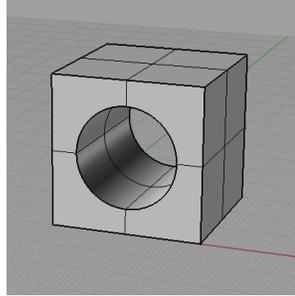
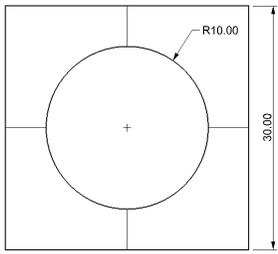


Figure 5. Target Object

Experiments and Results

Figure 6 shows the start of the experiments in which the stereo camera system was calibrated using a checkerboard. After calibration, camera matrices were generated and stored as text files that later were used to compute disparity maps and generate surface point clouds of the target object. Figure 7 shows how, for each baseline setting, the target object was moved from 700 to 250 mm. A recalibration was performed for each baseline adjustment in order to obtain proper disparity maps. A pair of images from the stereo camera system was then recorded for generating the disparity map and point cloud.



Figure 6. Camera Calibration Process

The position of the target and camera baseline significantly affected the quality of the disparity map. At baseline lengths greater than or equal to 90 mm, disparity maps from the stereo camera generated extreme surface errors at all target distances, implying that baseline lengths of 90 mm and wider were unsuitable for detecting objects in the range under 700 mm. Therefore, the results of this study were discussed based on baselines of 30, 45, 60, and 75 mm in the settings.

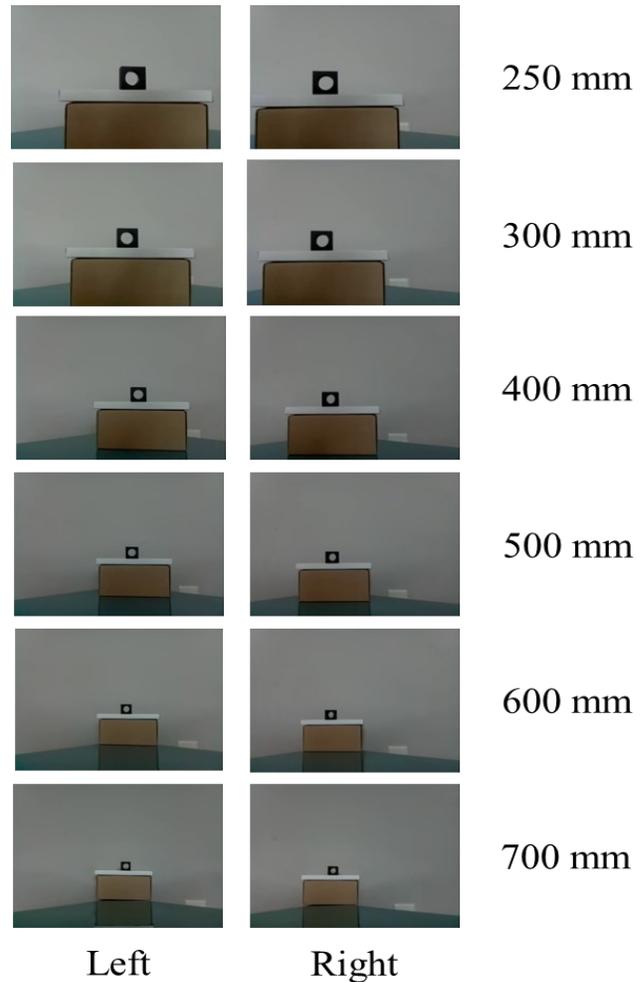


Figure 7. Stereo Images for a Baseline of 45 mm

The three-dimensional surfaces of the target were subsequently generated from the point clouds. Post-processing procedures were applied to the point clouds in order to remove noise and irrelevant points from the scene. Figure 8 illustrates an example of a point cloud after post-processing procedures. Surface areas (in square millimeters) were calculated to assess the ability of the stereo camera to pick up the surface of a target object when different baselines were used.

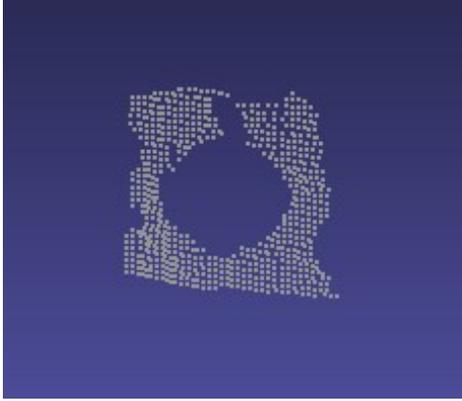


Figure 8. Example of the Point Cloud of a Target Object

Figure 9 and Table 1 present the results. Figure 9 displays a stereo camera with a 30-mm baseline that was able to pick up the most surface areas of the target object at distances of 250 and 300 mm. At 400 mm distance, the 45-mm baseline captured more surfaces than other lengths. At 500 and 600 mm distances, cameras with baselines of 45, 60, and 75 mm had very similar results, but were all better than the 30-mm baseline. At 700 mm, all baselines yielded similar results.

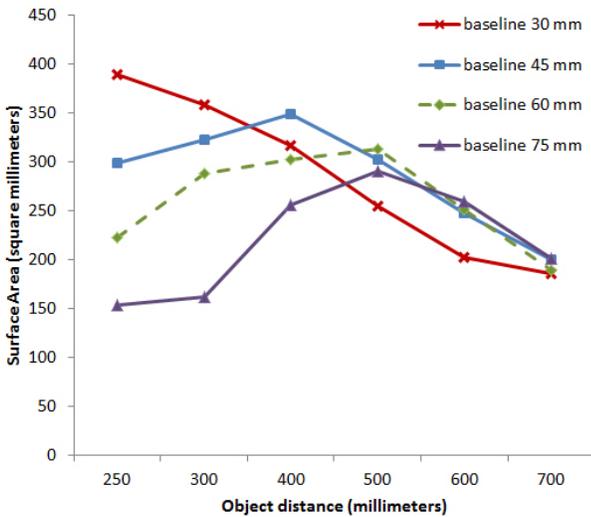


Figure 9. Chart of Surface Areas of the Target

Table 1. Surface Areas of the Target at Different Baselines

Object Distance	Surface Areas (Square Millimeters)			
	30 mm	45 mm	60 mm	75 mm
250 mm	389.32	298.61	222.57	153.36
300 mm	357.21	322.54	287.54	161.24
400 mm	316.47	347.93	301.26	254.87
500 mm	254.32	301.59	312.25	289.65
600 mm	201.97	247.63	251.27	259.31
700 mm	185.68	199.87	188.63	201.23

Discussion

Results from the experiments showed that stereo camera baselines could affect the 3D projection of a target object at different distances. In general, shorter baselines performed better when the object was at shorter distances, whereas longer baselines tended to perform better when the object was at greater distances. In addition, each baseline seemed to have an optimal point where it could perform best at a particular distance. Several factors might need to be considered before implementing this system, including calibration parameters, object size, and lighting conditions. Obtaining optimal calibration parameters for a camera system is challenging. These parameters appeared to be unique for each baseline. Future experiments can extend this study by identifying the relationship between baseline and those calibration parameters, as well as creating an automatic system that can adjust without the need for recalibration each time. In this study, the size of the target object remained constant. Using various sizes of objects may help identify the effectiveness of the 3D projection for larger baselines. Finally, lighting conditions can influence the performance of the camera system, especially when the target object is placed further from the camera. In this study, the lighting conditions were held constant across all experiments. Future research can manipulate the lighting conditions to confirm the results of the study.

Conclusions

In summary, the author of this current study implemented a preliminary stereo vision system for industrial robotics. The results suggest that adjusting the stereo camera baseline can influence the accuracy of 3D projection from disparity maps. An optimal stereo system should have a dynamic baseline that varies depending on the distance between the cameras and the target object.

References

- [1] Hogue, A. & Jenkin, M. R. (2014). Active Stereo Vision. *Computer Vision: A Reference Guide*, Springer US, 8-12.
- [2] Klarquist, W. N., & Bovik, A. C. (1998). Fovea: A foveated vergent active stereo vision system for dynamic three-dimensional scene recovery. *Robotics and Automation. IEEE Transactions*, 14(5), 755-770.
- [3] Murray, D. W., McLauchlan, P. F., Reid, I. D., & Sharkey, P. M. (1993). Reactions to peripheral image motion using a head/eye platform. *Proceedings of International Conference on Computer Vision*, (pp. 403-411). Los Alamitos, CA.

-
- [4] Saxena, A., Driemeyer, J., & Ng, A. (2008). Robotic grasping of novel objects using vision. *The International Journal of Robotics Research*, 27(2), 157-173.
 - [5] Okutomi, M., & Kanade, T. (1993). A Multiple-Baseline Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4), 353-363.
 - [6] Gallup, D., Frahm, J. M., Mordohai, P., & Pollefeys, M. (2008). Variable baseline/resolution stereo. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (pp. 1-8). Anchorage, Alaska.
 - [7] Nakabo, Y., Mukai, T., Hattori, Y., Takeuchi, Y., & Ohnishi, N. (2005). Variable baseline stereo tracking vision system using high-speed linear slider. *Proceedings of IEEE International Conference on Robotics and Automation*, (pp. 1567-1572). Barcelona, Spain.
 - [8] Rovira-Mas, F., Wang, Q., & Zhang, Q. (2010). Design parameters for adjusting the visual field of binocular stereo cameras. *Biosystems Engineering*, 105(1), 59-70.

Biography

WUTTHIGRAI BOONSUK is an assistant professor of Applied Engineering and Technology at Eastern Illinois University. He earned his MS and PhD degrees in Industrial and Manufacturing System Engineering from Iowa State University. Dr. Boonsuk also received his second MS degree in Human Computer Interaction from Iowa State University. His research interests include 3D stereoscopic applications, manufacturing systems, rapid prototyping, robotic and controller systems, virtual reality, and geographic information systems (GIS). Dr. Boonsuk may be reached at wboonsuk@eiu.edu

TOWARDS THE IMPROVEMENT OF HIGHWAY-RAIL INTERSECTIONS WITH AN INTELLIGENT GRADE CROSSING SYSTEM

Osileke Osipitan, New York State Department of Transportation

Abstract

Safety at highway-rail intersections (HRIs) is currently dependent on installed active warning systems such as flashing lights and gates. The volume of trains, highway vehicles, and pedestrians traversing the HRIs has increased over the years. Despite the improvements and activation of warning devices, crashes at HRIs still occur. In this study, the author examined crashes that involved motorists, who violated activated warning devices and other causes combined at HRIs from 2000-2015 in New York State. The methodology involved gathering accident data for the 16-year period from the New York State Department of Transportation. Descriptive and inferential statistics were employed to analyze the data. Chi-square goodness-of-fit was used to test the hypothesis to determine the statistical significance difference between the total number of accidents that occurred as a result of violation of activated warning devices by motorists as well as all other accidents. While the minimum expected frequency for the respective classified accidents was approximately 234, the test indicated that the number of accidents involving motorists that violated activated warning devices and other causes was statistically significant and are unequal ($X^2(1) = 4.336, P=0.037$). Findings indicated that accidents, which occurred as a result of motorists' violation of activated warning devices, were more than other causes of accidents. The intelligent grade crossing system (IGCS), which consists of the integration of intelligent transportation systems used for roadways with intelligent railroad systems technology towards enhancing safety at HRIs, was described and recommended towards improving safety at HRIs. In addition, the New York State law should be reviewed so that cameras installed as a result of this solution could aid law enforcement and reduce motorists' violation of activated warning devices at HRIs.

Introduction

A highway-rail intersection (HRI) is a point at grade where rail and road systems meet. Measures have been taken to reduce crashes between vehicles and trains through provision of active warning devices at public HRIs, which are used interchangeably with public at-grade railroad crossings in this study. Despite improvements made at these

HRIs, accidents still occur and continue to be a major problem at public railroad crossings, specifically in urban areas. Since trains have the right-of-way, motorists needing to cross the railroad track must stop and give way to approaching trains.

For adequate warning, public crossings have been equipped with gates and flashing lights, which are regarded as active warning devices, to reduce collisions between highway vehicles and trains. When flashing lights only, or flashing lights with gates, are absent at a given HRI, such a crossing is considered passive. The upgrading of a passive crossing with active warning devices improves safety at the crossing. There is higher potential for conflicts between the two systems of transportation due to of the volume of trains and vehicles passing over active crossings. About half of all HRI accidents between 1994 and 2003 occurred at railroad crossings that were equipped with active warning devices [1]. The only method for eliminating exposure between highway users and trains is to separate both systems from at-grade.

Since the cost of a grade separation or installation of a bridge for eliminating a high-risk candidate HRI is so expensive, it is necessary to find solutions to safety improvements at railroad grade crossings. The volume of motor vehicles traversing a crossing is increasing, while pedestrians, bicyclists, and motorists do foul the tracks. Similarly in urban centers, commuter train movements have increased over the years, based on population growth and demand for public transit. Therefore, in order to improve mobility and combat crashes at highway-rail intersections, the author investigated crashes involving motorists that violated active warning devices, and all accidents based on other causes at HRIs in New York State, so as to provide intelligent grade crossing systems to help reduce crashes.

Background

The highway-rail intersections in New York State (NYS) include both public and private crossings. Public crossings are highway-rail intersections that are open and used by the general public. The installations of warning devices at these crossings were funded by federal and state governments, while the maintenance of such devices remains in the hands

of railroad organizations. In addition, the highway approaches are being maintained by the municipalities having jurisdiction over the public roads. Private crossings are HRIs that consists of roadway, which leads to private properties and are maintained by private owners. The installation and maintenance of warning devices are based on agreements made between the railroads and owners of the private crossings. As of 2016, in New York State, there were 2911 public railroad crossings and 2944 private railroad crossings. Figure 1 indicates the percentages of public and private HRIs in New York State.

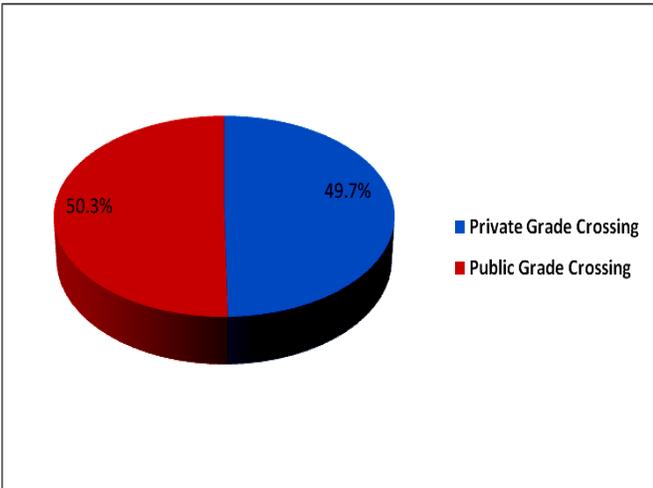


Figure 1. Public and Private HRIs in New York State

The improvements of public crossings have gulped millions of dollars annually through share-cost funding between the federal and state governments, and sometimes with local authorities, in order to mitigate risks encountered by the general public using the intersections. The public crossings attract a high volume of road traffic, while urban cities, specifically in downstate New York, attract a high volume of train and highway traffic. Despite the improvements made to warning devices at public HRIs, addressing the causes of accidents with advance awareness to users of the system as well as the ability of train operators/engineers to stop quickly during emergencies is quite necessary. The New York State Department of Transportation classified causes of accidents based on the following:

- Deliberate violation of activated gates by motorists
- Deliberate violation by pedestrians and/or bicycles
- Vehicle striking train
- Outside factors and conditions
- Crossing characteristics
- RR error/malfunction

In this study, the author investigated causes of accidents involving motorists, who deliberately violated activated warning devices, as well as other causes at public HRIs in

NYS. These at-grade railroad crossings require government intervention and funding to reduce fatalities, thereby saving the lives of the traveling public that uses them.

Problem Statement

Highway-rail intersections have claimed many lives in the U.S., including NYS, and caused countless injuries. The federal government has appropriated funds for states to improve public crossings and install active devices, which includes flashing lights and gates to warn motorists. Despite this investment, lives are still being lost. However, while it is difficult to eliminate these crossings, it is also expensive to grade-separate all of the crossings. Motorists and other road users need to safely cross the tracks in the course of travel. Therefore, this study provided a solution by integrating intelligent transportation systems with a positive train control system at HRIs in order to alert highway users.

Methodology

In this study, the author examined the crashes at HRIs for a 16-year period in New York State in order to complement the improvements made to current warning systems at highway-rail intersection (HRI) locations, where crashes have occurred. The methodology involved gathering accident data at HRIs for the years 2000-2015 from the New York State Department of Transportation. The data was the population of all accidents occurring during the period of study. The data were categorized into accidents that occurred as a result of driver violation of activated gates, and all other causes during the 16-year period. Data retrieved were copied into Microsoft Excel. These data were sorted and checked for errors. The data were then imported into SPSS 20 for statistical inferences, using the entire data population. The descriptive and inferential statistics were employed to analyze the data. In other to test statistical significance, a non-parametric chi-square goodness-of-fit test was conducted to determine the significant difference between the total number of accidents by motorists who violated activated warning devices and the total of all other accidents caused by other factors. The alpha level was set at 0.05. Therefore, the following hypothesis was tested:

Hypotheses

H₀: There is no statistically significant difference in the number of accidents that occurred as a result of violation of activated gates by motorists and all other accident causes at HRIs from 2000-2015 in New York State.

H₁: There is a statistically significant difference in the number of accidents that occurred as a result of violation of activated gates by motorists and all other accident at HRIs from 2000-2015 in New York State.

Findings

The accident data gathered during the study period were analyzed using Microsoft Excel as well as SPSS 20 for descriptive and inferential analyses. The total number of HRI accidents that occurred during the study period was 467. Figure 2 indicates the trend of the total accidents per year from years 2000 to 2015.

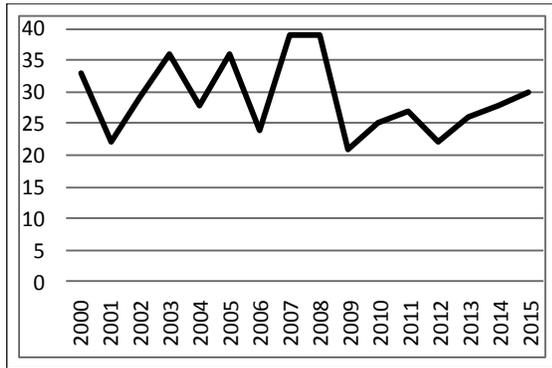


Figure 2. Trend of the Total Accidents per Year from 2000-2015

These accidents fluctuated based on the period. They were collections of all types of accidents. Despite the improvements made to HRIs annually, there was no indication of consistency or steady reduction in total annual accidents. The peak of the total annual accidents at HRIs occurred in years 2007 and 2008, while the fewest accidents occurred in 2009. Relative to the cumulative percentage of the number of accidents, Figure 3 indicates that over 50% of the total accidents occurred between 2007 and 2015.

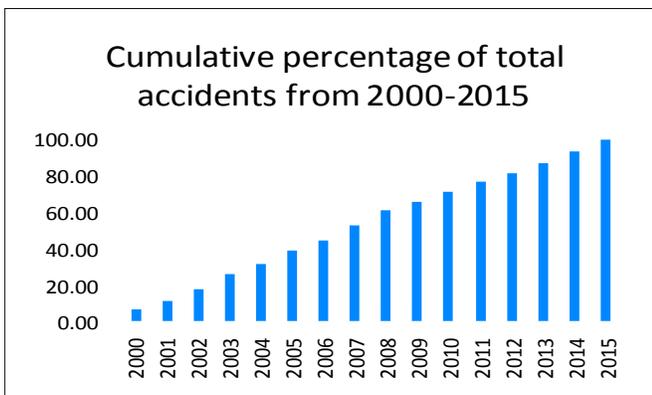


Figure 3. Cumulative Percentage of Accidents from 2000-2015

During this period, there were improvements at railroad crossings but not all candidate crossings were improved simultaneously, because funds appropriated by the federal government for annual improvement for these HRIs have remained the same—about six million dollars—while project costs have been increasing. In addition, traffic volume at these crossings is also increasing. The accident occurrence at railroad crossings in this study was classified as motorists violating the activated gates as well as accident occurrences as a result of other causes. Figure 4 indicates that 55% of the accidents occurred as a result of deliberate violations by motorists, while other causes were 45% throughout the study period.

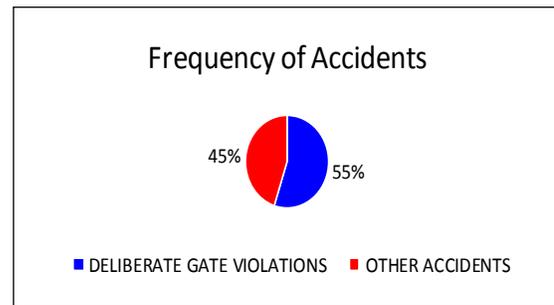


Figure 4. Accidents as a Result of Deliberate Gate Violations to Other Accidents at Railroad Crossings from 2000-2015 in New York State

Table 1 shows the number of accidents based on the classification. A total of 467 accidents occurred during the 16-year period.

Table 1. Accident Classification

	Observed N	Expected N	Residual
Motorists Gate Violation	256	233.5	22.5
Other Causes	211	233.5	-22.5
Total	467		

Out of the 467 accidents at HRIs, 256 of them involved motorist gate violations, while 211 involved all other causes including, deliberate violation by pedestrians and/or bicyclists, weather conditions, vehicle striking trains, malfunction of warning devices, and crossing characteristics. The minimum expected frequency indicated in Table 1 was approximately 234, which satisfies the assumption of a minimum of five in each category for the chi-square test. The chi-square goodness-of-fit test conducted determined whether the number of accidents caused by motorists that violated warning devices were equal to all other accidents caused by other factors at HRIs in New York State. Table 2 shows the test of significance.

Table 2. Chi-Square Test Statistics

	Accident Classification
Chi-Square	4,336 ^a
Df	1
Asymp. Sig	0.037

a. 0 cells (0.0%) had expected frequencies less than five. The minimum expected cell frequency was 233.5

The chi-square test of significance, indicated that accidents that occurred as a result of violation of activated warning devices and other accident causes was statistically significantly different ($X^2(1) = 4.336, p = 0.037$). Over 45 more accidents occurred as a result of violation of active warning devices over other accident causes altogether. The expected cell frequency of 234 was not met.

Discussion

Based on the findings, most of the accidents occurred in years 2007 and 2008. Over half of the number of accidents occurred from 2007 to 2015. The activated devices meant to warn motorists were violated by drivers either driving around the gates or not heeding to the signals. This was termed as deliberate actions by the drivers, who could not wait for the approaching trains to pass the crossings. This action surpassed other accident causes altogether, which included violation by pedestrians and/or bicyclists, weather conditions, vehicle striking trains, malfunction of warning devices, and crossing characteristics. This justified the relative assertion that most accidents involving commuter rail lines occur when people ignore railroad crossing gates. When the safety signals at a railroad crossing are ignored, it is dangerous and illegal in New York State, in accordance with the vehicle and traffic law (section 1170), which imposes criminal penalties and fines against those disobeying active warning devices [2]. Since it is very expensive to grade-separate each of these high-risk crossings, an intelligent grade crossing system (IGCS) could complement the active warning devices that are currently installed at HRI and help mitigate the accidents.

Intelligent Grade Crossing

An intelligent grade crossing system (IGS) utilizes technology that would help mitigate crashes and collisions at HRI. It is a location where an intelligent transportation system (ITS) used for the roadway is integrated with an intelligent railroad system, most especially the positive train control (PTC) system [3]. It is the ultimate solution for railroad grade crossing safety. It utilizes many ITS applications

used for roadways such as vehicle detection, Global positioning system (GPS) tracking, advance traveler information systems, etc., towards enhancing safety [4] as well as railroad transponders and wayside detection systems. The ITS technology was applied to roadway systems to improve safety by integrating communication and information technology [5]. The PTC system provides information on train speed as well as the location of the train in real-time. It uses communication-based technology to control trains for effective prevention of trains from accidents, especially when a train engineer or operator cannot take appropriate action to stop the train when required [6]. The technology involves computers on-board in trains, wayside interface devices with servers from the control center using wireless communication networks, positioning system such as global positioning system (GPS), and transponders to continuously control rail operation in real-time [7].

Based on the aforementioned application of ITS technology for roadways, it could also be applied to vast transportation infrastructure and vehicles. The field devices, which include cameras, detectors, and dynamic message signs, could be integrated with the railroad system and the existing warning devices at a grade crossing location. A dynamic message sign (DMS) is an electronic sign positioned along highways. It is used to display information on traffic conditions, travel time, incidents, and roadway construction [8]. It was formerly called a changeable message sign and sometimes called a variable message sign, which is needed as a link between transportation agencies and the public they serve for traffic conditions [9].

The DMS term was developed within the NTCIP (National Transportation Communication for ITS Protocol) in order to create a standard that would support changeable message sign and variable message sign with a common set of data [10]. In this study, the system's primary information was to help provide the position and speed of trains at HRI to motorists approaching highway-rail intersections. The DMS should be placed in a visible position to attract roadway users. Messages indicated must be shown in a manner that can be understood by motorists.

The closed circuit television (CCTV) camera is a field device within the ITS. It is installed on highway corridors to capture images and sends information to centers at remote locations. Such centers consists of equipment including large-screen monitors, servers, computers, and cable networks that link and communicate with the devices from such centers to the field [11]. The sensors used to detect and collect traffic data along the highway were applicable to this study. The non-intrusive sensors were required for detecting vehicle and train presence as well as speed. These sensors and CCTV can detect stalled vehicles as well as any block-

age at HRIs and send such information to the cabin of train engineers/operators and the control centers for immediate response [4].

Architecture

The architecture encompassed layers that include institution, transportation, and communication. The institution entails the organization that will provide the funding mechanism and policies as well as the effective implementation, operation, and maintenance of the system. The architecture is needed for planning and project development. From the transportation layer point of view, the user service applicable to this study is the HRI. Based on the user service, railroad crossing safety requires integration of both the highway and rail systems relative to logical and physical architecture. According to USDOT-RITA, the national architecture consists of logical and physical architecture, which also provides a framework for designing transportation systems that define ITS user services [12].

The key components of the architecture are vehicles and field, which are regarded as the environment, while roadway users and centers are remote areas. The HRI user service in the national architecture uses ITS technologies to improve train control and detect/alert highway traffic so as to avoid or mitigate the severity of crashes between trains and highway vehicles at HRIs. It was indicated that nine constituents make up the HRI national architecture and include track circuit, wayside signals, flashing lights and gates, traffic signals, dynamic message signs (DMS), and surveillance cameras [13]. Presently, the operational speed of most commuter traffic in New York State is 79 mph. The standard speed rail subservice used by light rail transit, commuter, and heavy trains with operational speeds up to 80 mph in the national architecture was adopted for this study. Their integration with each other required digital data link communication networks as well as dedicated short-range communication (DSRC) so that they could talk to each other on the same protocol in real-time. The architecture is a standard that will reduce time when tailored accordingly. The process specification applicable to the roadway system as well as market and equipment packages for standard rail crossings in the national architecture was adopted for implementation. The packages consist of the system elements, which entail the advanced traveler information system (ATIS), the advanced traveler management system (ATMS), and the advanced detection system (ADS).

The System Engineering Process

The system engineering process (SEP) follows various steps that guide accomplishment of the design for the IGS at

a given HRI in order to avoid excessive cost of implementation. The concept of operation would be tuned in line with the overall goal of improving safety. The major feature is precise warning time, highway motorists' information, and the tendency for trains to stop during emergencies at HRIs. Field devices, such as CCTV and other detective devices, are integrated with a grade crossing controller, railroad wayside technology, and train cab-based technology towards information sharing between the railroad and the municipalities in charge of the roadway. The DMS is located on roadway approaches. For proper functioning, the installed devices at crossings send information to the railroad center, while the controller at the grade crossing communicate with the train operator when a stalled vehicle is detected prior to arrival at the HRI. The municipality or regional transportation center, railroad center, and emergency management center should be integrated with applicable packages so that they are interoperable. The design should accommodate all future protocols. The system functionality would conform to specifications, test, and verification. The system must continue to fulfill its objective and be failsafe, based on continuing operation and maintenance, as well as upgrade. Figure 5 shows an IGC design.

System Deployment

The system would use the existing grade crossing controller and track circuitry at the grade crossing approach. Devices such as CCTV and loop detectors are installed at the crossing. Real-time conditions, such as obstacles detected at a crossing, are processed through the grade crossing controller and conveyed directly to the approaching train cabin and the railroad center through wireless communication or transponders. In order to determine the position of the train, the GPS mounted on antennas on the train help determine the location of the train in real-time. Similarly, transponder systems that were mounted on track beds or rail ties could help determine train locations. This could be achieved through the reader antennas mounted on the trains to capture power from installed transponders located at intervals along the tracks. The transponders have the advantage of indicating track locations and loss of signals because they use radio frequency (RF) signals. The train can locate itself on the map through the identity of the transponder. The read data that are transmitted to the on-board computer for speed and position of the train are then relayed to the field devices such as DMS to alert motorists through its integration through the grade crossing controller to the railroad center.

Where there are multiple tracks, particularly at crossings near train stations, it also helps alert pedestrians of the approach of secondary trains. The main attribute of the train-based technology is that an on-board computer determines

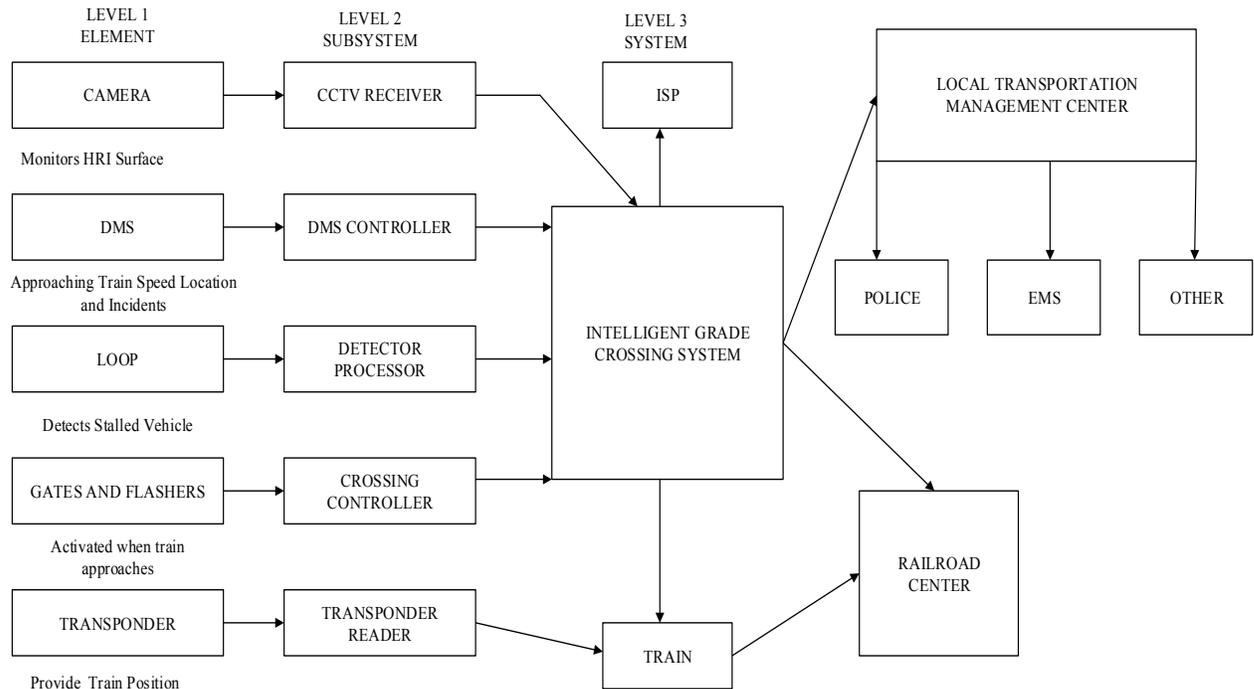


Figure 5. Intelligent Grade Crossing System

its location and transmits data to the system for processing and dissemination [2]. Figure 6 indicates the HRI layout with a deployed IGC.

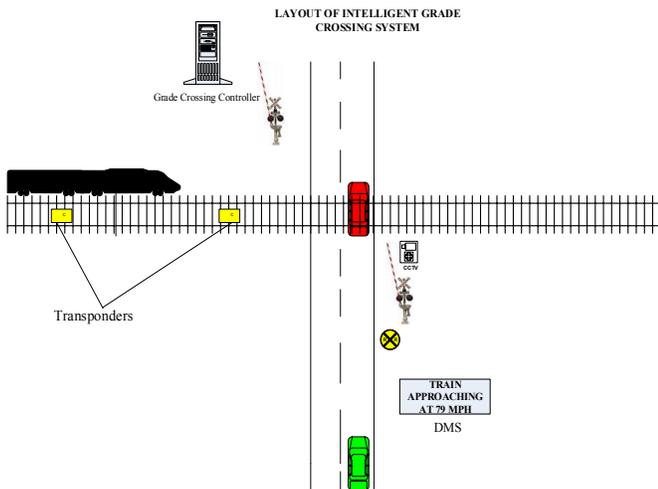


Figure 6. Intelligent System Deployment

The integration of the rail system with the ITS requires appropriate standards to ensure proper exchange of information between the traffic management center (TMC) and

any other centers and field devices. The standards would ensure interoperability, reduce costs, and avoid proprietary rights, since products of multiple vendors can be chosen. This would allow the centers to talk to each other for effectiveness. Notable standards include the National Transportation Communication for ITS Protocols (NTCIP), which provides the communication rules and defined applicable data for controlling field devices from a given center. Traffic management data dictionary (TMDD) adds additional vocabulary not in NTCIP for center-to-center communication. Incident management (IM) consists of various standards developed by IEEE to address interfaces between the emergency management center and traffic management center as well as other centers, and provide data elements and messages. The integration of local TMCs with the railroad center allows them to have access to information relative to incidents at the local TMC to manage the corridor for traffic routing.

Conclusions

In order to eliminate crashes between trains and highway vehicles at crossings, a grade separation is needed. Because the cost of replacing an HRI with a bridge is extremely expensive, the high risk of crashes at existing crossings within

a railroad corridor can be mitigated with intelligent grade crossing (IGC) technology, which integrates a detection system, DMS, and railroad wayside system. Based on the causes of accidents in New York State, deliberate/illegal driver actions dominate these factors. Drivers fail to heed to devices when activated, with the aim of beating the train before it arrives at the crossing. However, the trains are killing the drivers and their occupants. Other identified causes include vehicles getting stalled on the tracks, drivers abandoning their vehicles, suicide, etc. These accidents could be minimized by providing an IGC system. The IGC will provide drivers with the speed and location of trains through a dynamic message sign so that they can stop in a timely manner, even when visibility at the HRI is poor. Similarly, train engineers or operators would be informed ahead of any stalled vehicles at the crossings, so that the train could be stopped before arriving at the crossing. In the event of an emergency, rail transponders could help force the train to stop.

Presently, there are no cameras at HRIs; similarly, state law does not allow enforcement with CCTV. In order to prevent drivers from deliberately going around the gates at HRIs in New York State, statewide enforcement would be necessary through CCTVs connected to the IGC system, rather than relying only on a police presence before imposing penalties and fines. This would help reduce the number of accidents at HRIs. Therefore, as a starting point, the author recommends that this system be applied to a select group of high-risk, high-volume crossings to curb impatient motorists.

References

- [1] Yeh, M., & Multer, J. (2007). Traffic Control Devices and Barrier Systems at Grade Crossings: Literature review. *Transportation Research Record*, 2030, 69-75.
- [2] Metropolitan Transportation Authority (n.d.). Railroad Crossing Safety. Retrieved May 10, 2016, from <http://web.mta.info/mta/police/safety/rcs.html>
- [3] USDOT-Federal Railroad Administration (n.d.). Intelligent grade crossings. Retrieved April 26, 2016 from <http://www.fra.dot.gov/Page/P030>
- [4] Zhang, L., & Schurr, K. (2005). How technologies will bring us safer and smarter railroad crossings. Mid America Transportation Center, University of Nebraska, Lincoln
- [5] Patel, R. (2005, January). *Intelligent transportation systems (ITS) operational support contracts implementation plan*. Final Report, UTRC Region 2, New York.
- [6] USDOT-Federal Railroad Administration. (n.d.).

- Positive train control information: Railroad safety. Retrieved April 26, 2016, from <http://www.fra.dot.gov/Page/P035>
- [7] Henry, L. (2015, April 21). Railroad crossings: A potential lifesaver. *All Analytics*. Retrieved April 26, 2016, from http://www.allanalytics.com/author.asp?section_id=2310&doc_id=27726
- [8] US Ohio Department of Transportation (n.d.). Retrieved April 20, 2016, from <https://www.dot.state.oh.us/Divisions/Operations/Traffic/FAQs/Pages/DMS.asp>.
- [9] Dude C. & Ullman, G. (2006). Dynamic message sign message design and display manual. Retrieved May 8, 2016, from <http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/0-4023-P3.pdf>
- [10] NYSDOT – Office of Traffic Safety and Mobility (2011) Variable message sign guidelines. Retrieved April 28, 2016 from <https://www.dot.ny.gov/divisions/operating/oom/transportation-systems/repository/VMS%20%20Guidelines%20Aug%202011.pdf>
- [11] Osipitan, O.O. (2016). Using intelligent transport system technology to manage work zones in Lagos, Nigeria. *Technology Interface International Journal*, 16(2), 13-20
- [12] USDOT-FHWA (n.d.) Intelligent Transportation System Architecture and Standards. Retrieved April 28, 2016 from http://ops.fhwa.dot.gov/its_arch_imp/policy_1.htm
- [13] Schulz, L.A. (1998). Application of ITS at railroad grade crossings. Upper Great Plains Transportation Institute, North Dakota State University. Department Publication 123. Retrieved April 23, 2016, from <http://www.ugpti.org/pubs/pdf/DP123.pdf>

Biography

OSILEKE OSIPITAN is an intermodal transportation specialist with the New York State Department of Transportation. He received his PhD in Technology Management, specializing in Construction Management, from Indiana State University. He is a Chartered Quantity Surveyor and professional member of the Royal Institution of Chartered Surveyors. He is also a chartered member of the Chartered Institute of Logistics and Transport. Dr. Osipitan may be reached at osipitan@hotmail.com

EPISTEMIC UNCERTAINTY EFFECTS ON RESISTANCE FACTORS BASED ON STATIC TOP-DOWN LOAD TESTS OF DRILLED SHAFTS IN SANDS

Jiliang Li, Purdue University Northwest; Jinyuan Zhai, University of Akron; Li Tan, Purdue University Northwest

Abstract

Drilled shafts are often used as deep foundations for many structures and buildings, due to their high axial and lateral resistance bearing capacity. For example, many highway bridges are built on drilled shaft foundations. Based on the statistical data of 18 conventional top-down compression load tests of drilled shafts embedded in sands from the National Cooperative Highway Research Program (NCHRP) Project 24-17, the focus of this study was on calibration of the total resistance factors for drilled shafts. Actual measured total resistance was interpreted using the drilled shaft field test data. The in-situ soil properties from standard penetration testing (SPT), correlated with knowledge-based uncertainty (correction factors in American Association of State Highway Transportation Officials [1] and Liao and Whitman [2]) and the dimensions of drilled shafts, were used to calculate the predicted total resistance using the SHAFT program. The SHAFT program is consistent with the 1999 Federal Highway Administration (FHWA) method [3]. To further consider the knowledge and data-based uncertainty because available data are “incomplete” and inevitably contain variability, a Monte Carlo simulation method was used to perform calibration. Based on the normally distributed loads and log-normally distributed resistance bias from the test data, the resistance factors for the Strength I limit state were determined at a target reliability index of 3.0, and the results were compared with those reported in the literature.

Introduction

Drilled shafts are often used as deep foundations for many structures and buildings, due to their high axial and lateral resistance bearing capacity. For example, many highway bridges are built on drilled shaft foundations. As a more rational approach considering uncertainty in design, the load and resistance factor design (LRFD) has increasingly been used and become a mandatory design for all state Department of Transportation (DOT) and FHWA-funded bridge projects since 2007. The drilled shaft foundation for bridge design with the LRFD methodology helps ensure that the entire structure has the same compatible

reliability index, β , from the substructure of underground foundations to the above ground superstructures. For this purpose, updated and calibrated resistance factors from an adequate number of good-quality test data for each design method are needed [4-8]. However, the resistance factors that are calibrated from high-quality test data are often not available in geotechnical engineering. Obtaining the calibrated resistance factors by fitting LRFD to an allowable stress design (ASD) has been performed to maintain a consistent level of reliability with past practice; as an example, see Barker et al. [9]. The bridge design specifications based on the American Association of State Highway Transportation Officials (AASHTO) [1] provide resistance factors for design of the drilled shafts under axial loads. In fact, in AASHTO, a significant number of resistance factors in the foundation design are selected based on the calibration with ASD. To further improve LRFD implementation and to achieve compatible reliability with the superstructure, calibrated resistance factors based on actual field load test data rather than calibration to traditional factor of safety from ASD are sorely needed.

Every engineering system has uncertainties, from analysis to design, construction, and operations. Traditionally, engineers have to deal with a significant number of uncertainties through conservative assumptions and apply safety factors to cover the effects of underlying uncertainties [10, 11]. The level of conservativeness is difficult to quantify, as the assumptions and safety factors are often based on engineering judgments. There are two major sources of uncertainty: aleatory uncertainty associated with natural randomness and epistemic, or knowledge-based, uncertainty associated with inaccuracies in the prediction and estimation of reality [10-12]. In the drilled shaft resistance factor calibration, available data are often incomplete or insufficient, and typically display variability. First, this fact can result in the aleatory uncertainty associated with randomness of the drilled shaft calculated and measured capacity from field test data. Second, there is the epistemic or knowledge-based uncertainty, due to limited understanding and imperfect knowledge of the real world, for example, and the effects of different soil parameters correlated from the AASHTO and Liao and Whitman correction factors [1, 2] for standard penetration test (SPT) numbers. Besides devel-

oping the total resistance factors for the axially loaded drilled shafts in sands collected in the National Cooperative Highway Research Program (NCHRP) Project 24-17, the objective of this study was to investigate the knowledge-based uncertainty effects—in particular, the statistical effects of the two most widely used SPT number correction factors in the AASHTO and Liao and Whitman methods [1, 2], respectively, on the final drilled shaft calibrated total resistance factors.

The total resistance of an axially loaded drilled shaft consists of skin resistance and tip resistance. The recommended design methods for these two components by AASHTO and FHWA are based on O’Neill and Reese’s 1999 study [3]. From traditional top-down load tests, it is often difficult to separate tip and skin resistance from the measured total resistance. In this study, the FHWA method, based on the O’Neill and Reese study and load test data, were used for the calibration of the total resistance factors for drilled shafts in sand soils [3]. For top-down loaded drilled shafts, there are no available skin and tip resistance measurements of drilled shafts in the database; therefore, the total resistance factors calibrated for the drilled shafts in sand were based on 18 valid drilled shaft datasets.

Conventional Top-Down Load Tests

The conventional top-down load test simulates more realistically the actual loading conditions from superstructures for drilled shaft foundations, in which the axial load versus settlement is measured. In this study, data from 18 loaded drilled shafts in sands were collected from the NCHRP drilled shaft database. The load, corresponding to a settlement at 5% of the shaft diameter (“0.05B”) or plunging load, was defined as the nominal capacity, as recommended by O’Neill and Reese [3]. The selection of this criterion was based on a previous study by Paikowsky [16] for LRFD calibration consistency. Zhang et al. [13] showed that the FHWA’s “0.05B” method produced the closest and consistent capacities with the mean value of the capacities. And this failure criterion has been further con-

firmed by Liang and Li [8]. The FHWA drilled shaft database contains 261 conventional top-down statically loaded tests. Among them, the data for 18 drilled shafts in sands of the conventional static top-down tests’ failure load with strictly exhibiting FHWA failure load criterion [3] have been collected for the resistance factors calibration in this study.

FHWA Design Method

The FHWA design method [14] for cohesive, cohesionless soils and weak rock geomaterials was based on the work by O’Neill and Reese [3], assuming smooth rock socket surfaces with closed joints [15]. Paikowsky’s study [16] for LRFD is also partly based on the FHWA method, where rough sockets were assumed. The assumption of smooth surface is more commonly used in practice and expected to yield a lower predicted load capacity of drilled shafts. Some drilled shaft data collected in this study were for sandy granular soil with less than 50 blow count/foot and weak rock having a wider range of unconfined compressive strengths than those for IGMs (0.5 MPa to 5 MPa) [13]. The methods for calculating skin, tip, and total resistance in cohesionless soils are documented in studies by O’Neill and Reese [3] and Reese and O’Neill [14]. In this current study, the following conditions were met or assumed: a) the geomaterials were sands, and b) the concrete had slump of 152.4 mm (127 ~ 229 mm or 5~9 inches) recommended from SHAFT and a sand unit weight of 18.86 kN/m³ (120) lb/ft³ (if no information was available).

According to AASHTO and Das and Sobhan [17], the knowledge-based uncertainty effects, in particular, the two most widely used SPT number correction factors, in the AASHTO [1] and Liao and Whitman methods [2] are listed in Table 1. Das and Sobhan [17] compared the different correction factors proposed by Liao and Whitman [2], Skempton [18], Seed et al. [19], AASHTO [1], Peck et al. [20], and Bazaraa [21] concluded that the Liao and Whitman method was better. In this current study, the focus of the investigation was limited to the epistemic or knowledge

Table 1. Liao and Whitman [2] and AASHTO [1] Correction Factors, C_N , for SPT Number

C_N , Correction factor, methods	C_N (σ' in ksf)	C_N (σ' in US ton/ft ²)	C_N (σ' in kN/m ² , or kPa)
Liao and Whitman (1986)	$\sqrt{\frac{2000}{\sigma'}}$	$\sqrt{\frac{1}{\sigma'}}$	$9.78\sqrt{\frac{1}{\sigma'}}$
AASHTO (2007) [1], Peck et al. (1974)	$0.77 \log_{10}\left(\frac{40}{\sigma'}\right)$	$0.77 \log_{10}\left(\frac{20}{\sigma'}\right), \sigma' \geq 0.25tsf$	$0.77 \log_{10}\left(\frac{1912}{\sigma'}\right), \sigma' \geq 25kPa$

based uncertainty effects using Liao and Whitman and AASHTO correction factors, C_N , for SPT number and soil internal friction angle in study of the resistance factors calibration. As shown in Table 1, σ' is the current overburden effective stress, while the correction factor, C_N , comes from Liao and Whitman's 1986 method [2] and AASHTO [1], based on the work by Peck et al. [20]. Epistemic (or knowledge-based) uncertainty of the real world was considered, for example, through the correction factors, C_N . The soil internal friction angle, ϕ (or PHT, as shown in Table 3) for computing the predicted capacity, has correlation with the SPT number, according to Peck et al. [20], as shown in Equation (1):

$$\phi \approx 54 - 27.6034 \cdot e^{(-0.014N')} \quad (1)$$

The calibrated soil internal friction angle, ϕ (or PHT), was used in the SHAFT program [15] for evaluating the predicted bearing capacity, which was agreeable and consistent with the 1999 FHWA design method [3]. Note that the SHAFT program was developed based on the drilled shaft FHWA design methodology.

Summary of Resistance and Bias Statistics

The FHWA database contains drilled shaft dimensions and different construction methods. In addition, in-situ soil profiles, elevation, soil descriptions, and parameters are also available for calculating the nominal total resistance via predicted skin and tip resistance. The shaft-soil profiles and parameters required for obtaining the predicted capacity using the FHWA method via the SHAFT program were employed for prediction of skin, tip, and total resistance. The actual measured total resistance versus settlement can be obtained and graphically interpreted. Table 2 summarizes the statistical information for the drilled shafts measured (interpreted) resistance and calculated (predicted) resistance.

Table 2. Measured and Predicted Total and Unit Resistance of Drilled Shafts in Sand

Unit resistance KPa (tsf)	Measured	Predicted with two different correction factors, C_N	
		AASHTO (Peck et al.)	Liao and Whitman
μ_{UR}	265(2.77)	144 (1.50)	142 (1.48)
σ_{UR}	135(1.41)	67 (0.704)	66 (0.692)
COV_{UR}	0.51(0.51)	0.47 (0.469)	0.47 (0.467)

In Table 2, COV_{UR} is the coefficient of variation of unit resistance of drilled shafts installed in sand soils; σ_{UR} is the standard deviation of unit shaft resistance; and, μ_{UR} is the mean of the 18 drilled shafts unit resistance. Table 3 is a summary of 18 load test cases, including shafts diameter, embedded length, dominant type of sandy soil, method of installation, measured load at displacement of 0.05 diameter B, predicted failure load from SHAFT, and bias factor defined as ratio of actually measured capacity over predicted capacity. Table 4 summarizes the bias parameters used for resistance factor calibration via a Monte Carlo simulation method.

Calibration of Drilled Shaft Total Resistance Factor

Among the reliability analysis methods for calibration, two are most widely accepted: the first order of reliability method (Yang [7], Hasofer and Lind [22], Ellingwood et al. [23], and Phoon et al. [24]) and the Monte Carlo method (Allen et al. [5] and Roberts [25]). In this paper, the Monte Carlo method was adopted. The calibration in this study followed the recommended procedures by Allen et al. [5]. Due to limited scope, only the Strength I limit state was considered and the limit state function was adopted [5, 8, 26], as given by Equation (2):

$$g = \phi R - \gamma_{LL}LL - \gamma_{DL}DL \quad (2)$$

where, g is the safety margin; ϕ is the resistance factor; R is the nominal resistance; γ_{LL} and γ_{DL} are the live load and dead factors, respectively; and, LL and DL are the nominal live and dead loads, respectively.

Paikowsky's [16] parameters were adopted, as shown in Table 4. Both live and dead loads were assumed to be normally distributed. This assumption was consistent with the work by Nowak [27]. Note that bias λ was calculated by dividing the measured nominal resistance from the top-down load test data by the corresponding predicted value using SHAFT 5.0. Statistical analysis on the bias λ values was then performed, as shown in Figure 1 and listed in Table 4. The bias from the Liao and Whitman [2] method is generally larger than that obtained from the AASHTO method, but not significantly. Figure 1 shows a histogram of the bias values of the total resistance, ranging from 0.72 to 3.73. Figure 2 shows the cumulative distribution function (CDF) curves of standard normal variable Z with respect to different bias values. Based on current and previous studies [5, 26, 28], the bias was assumed to follow the lognormal distribution. Note that the mean values and the standard deviation values in Figures 1 and 2 can be used in the calibration for engineering conservativeness. Furthermore,

Table 3. Summary of 18 Load Test Cases

DSI D#	Install construction method	Drilled Shaft diameter B (in)	Length (ft)	Failure criterion plunging or settlement at B 5%	Qult (ton) interpreted from load test	UNIT Qult (tsf) interpreted from load test	PHT based on Liao & Whitman's Qult (ton) predicted with SHAFT 5.0	PHT based on AASHTO's Qult (ton) predicted with SHAFT 5.0	PHT based on Liao & Whitman's UNIT Qult (tsf) predicted with SHAF T 5.0	PHT based on AASHTO's UNIT Qult (tsf) predicted with SHAF T 5.0	Liao & Whitman Bias = measured/calculated	AASHTO Bias = measured/calculated
225	Casing	23.62	32.10	1.181	214	2.191	192	196	1.9657	2.0067	1.1146	1.0918
245	Dry	24.00	25.00	1.200	450	5.730	215	215	2.7375	2.7375	2.0930	2.0930
249	Casing	36.00	70.15	1.800	500	1.008	452	452	0.9115	0.9115	1.1062	1.1062
259	Dry Casing	30.00	29.50	1.500	770	5.317	393	393	2.7139	2.7139	1.9593	1.9593
276	Wet	36.00	100.00	1.800	965	1.365	807	832	1.1417	1.1770	1.1958	1.1599
281	Wet	30.00	45.00	1.500	676	3.060	490	512	2.2183	2.3179	1.3796	1.3203
282	Wet casing	45.96	60.00	2.298	450	0.651	442	442	0.6394	0.6394	1.0181	1.0181
283	Wet	36.00	59.00	1.800	450	1.079	598	623	1.4339	1.4938	0.7525	0.7223
284	Wet	30.00	76.50	1.500	670	1.784	217	225	0.5779	0.5992	3.0876	2.9778
301	Dry	33.60	30.00	1.680	650	3.519	378	375	2.0464	2.0302	1.7196	1.7333
313	Wet	36.00	65.50	1.800	950	2.052	523	534	1.1296	1.1534	1.8164	1.7790
332	Dry	30.00	22.00	1.500	320	2.963	107	107	0.9908	0.9908	2.9907	2.9907
334	Dry	24.00	20.00	1.200	162	2.578	73	73	1.1618	1.1618	2.2192	2.2192
335	Dry	18.00	20.00	0.900	137	3.876	55	55	1.5562	1.5562	2.4909	2.4909
336	Dry	18.00	17.00	0.900	125	4.161	38	38	1.2649	1.2649	3.2895	3.2895
337	Dry	24.00	20.00	1.200	200	3.183	67	67	1.0663	1.0663	2.9851	2.9851
435	Wet	30.00	17.00	1.500	235	2.816	63	63	0.7550	0.7550	3.7302	3.7302
817	Wet	27.80	24.60	1.390	268	2.585	244	255	2.3533	2.4593	1.0984	1.0510

Table 4. Bias Lognormal Distribution Parameters Used for the Monte Carlo Simulation for Resistance Factors Calibration

Bias λ statistics	AASHTO (18 Cases)		Liao and Whitman (18 Cases)	
	Total and unit resistance lognormal distribution	"Fit to tail" 10 total and unit resistance bias of sand	Total and unit resistance lognormal distribution	"Fit to tail" 10 total and unit resistance bias of sand
μ_{λ_R}	1.984	1.294	2.003	1.312
σ_{λ_R}	0.916	0.398	0.912	0.392
COV_{λ_R}	0.462	0.308	0.456	0.298
ϕ	0.45	0.50(0.51)	0.45	0.50 (0.52)

the bias mean value, μ_b , and the bias standard deviation, σ_b , in Table 4 can be determined following a “fit to tail” strategy [5]. Detailed procedures to develop the standard normal variables, z , and the CDF plots can be found in the Transportation Research Circular E-C079 [5]. Table 4 summarizes the “best fit” lognormal distribution parameters used in the calibration.

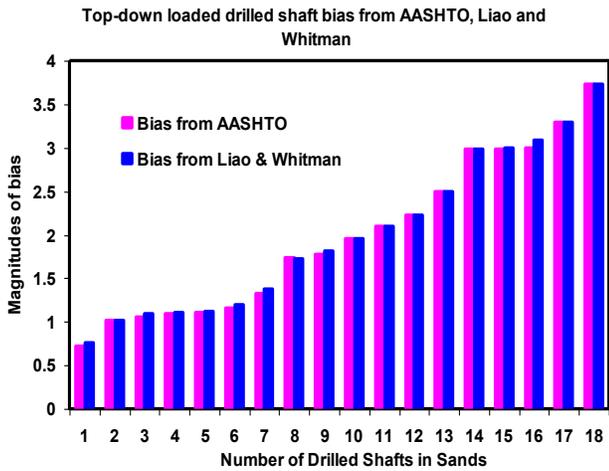


Figure 1. Histogram of Total Resistance Bias

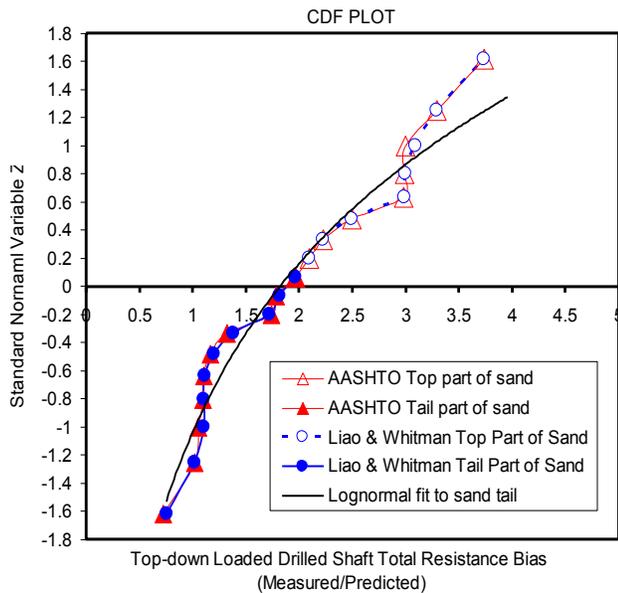


Figure 2. CDF Plot of the Bias of Total Resistance

Prior to the Monte Carlo simulation, a target reliability index, β_T , of 3.0 (approximately corresponding to the probability of failure, $P_f = 0.001$) and a ratio of dead load over live load (say DL / LL = 2.0) were selected. The variation of DL/LL does not appear to influence the calibrated resistance factors [5]. It is easy to appreciate that, depending

on the different methods of predicting the resistance, there will be considerable uncertainty related to the analysis method, especially when accurate soil profile properties are not available. To consider and implement uncertainty into the reliability-based LRFD design, the Monte Carlo simulation method, therefore, can be employed to investigate the uncertainty effects on resistance factors for the total unit resistance and, ultimately, the drilled shaft load and resistance factors calibration. Additional information on the Monte Carlo simulation details and their applications are available in studies by Ang and Tang [10, 11]. For the Monte Carlo simulation and calibration of resistance factors, a MATLAB program was written for the reliability analysis based on actual measured drilled shaft static load test results and predicted nominal resistance. Table 5 lists the statistical information of load components for calibration.

Table 5. Statistics and Load Factors [16]

Load type	Bias	Coefficient of Variation	Load Factor Used
Dead load	$\lambda_{DL} = 1.05$	$COV_{DL} = 0.1$	$\gamma_{DL} = 1.25$
Live load	$\lambda_{LL} = 1.15$	$COV_{LL} = 0.2$	$\gamma_{LL} = 1.75$

According to Harr [29], an approximately 4500 Monte Carlo trials are required for a confidence level of 90% achievement. The required number of Monte Carlo trials is based upon the need for achieving a particular level of reliability [29, 30]. In this current study, a Monte Carlo simulation with 10,000 trials was conducted. The mathematical expressions of the functions are complicated and omitted here. Details of the functions and calibration procedures can be found in the Circular E-C079 [5]. If the calculated β value is different from the target reliability index, β_T , the trial resistance factor must be changed and iterations must continue until $\beta = \beta_T$. The corresponding resistance factors calibrated from this procedure were achieved.

Results of Total Resistance Factors Calibrated for Drilled Shafts—Load Tests

Table 6 summarizes the total resistance factors calibrated from the load test data. The knowledge-based uncertainty of SPT correction factor effects from AASHTO and Liao and Whitman appear to be negligible, because each calibrated total resistance factor is rounded up to 0.50. In comparison with the current total resistance factor of $\phi = 0.55$ in cohesionless soils in AASHTO [1], the total resistance factors obtained in this current investigation generally agreed with the recommended resistance factor of 0.55 reported by Paikowsky [16] and AASHTO [1], which

fall on the more conservative side. It should be mentioned that the resistance factor calibrated in AASHTO, according to Paikowsky's work, was partly based on 1988 FHWA methodology [14], while the current study was based on the 1999 FHWA [3] method via the SHAFT program [15]. The slight differences in the calibrated total resistance factors may partly be due to the different versions of FHWA, that is, the calibration via the 1999 FHWA method instead of FHWA 1988 and/or different soil conditions.

Table 6. Drilled Shafts Total Resistance Factors Calibration Comparison

Current study $\beta_T = 3.0$	ϕ calibrated by fit to tail
Based on AASHTO, SPT C_N	0.50 in sand
Based on Liao and Whitman, SPT C_N	0.50 in sand
Paikowsky and AASHTO	0.55 in cohesionless soils
Abu-Farsakh et al.	0.60 for mixed soil
Liang and Li	0.35 for mixed soil

Conclusions

The total resistance factors considering epistemic or knowledge-based uncertainty effects, herein the two most widely SPT correction factors [1, 2], were calibrated and rounded up to 0.50 for recommendation for drilled shafts design through the 18 case studies. This current study indicated that the epistemic uncertainty of the two SPTB correction factors on the resistance factors was negligible. This finding generally agrees with the research results from Abu-Farsakh et al. [31], while the calibrated resistance factors considering the knowledge-based uncertainty via the SHAFT program were close to the calibrated resistance factor proposed in AASHTO and fall on the conservative side. The slight difference in the calibrated resistance factors summarized in Table 6 may partly be due to the fact that the current calibration was based on the FHWA 1999 method instead of the FHWA 1988 method, and also due to the different soils from different areas. Construction procedures and methods may also have affect the resistance factor calibration. Future work is needed to study the methods recommended by Brown et al. [32] and FHWA 2010 methods in different states; for example, the work reported by Abu-Farsakh et al. [31]. Further work will also verify any further updates, since the design method changes and different soil conditions may affect the recommended resistance factors endorsed by AASHTO [33] and those reported in the literature.

References

- [1] AASHTO (2007). *LRFD bridge design specifications* (3rd ed.). Washington, D.C.: AASHTO.
- [2] Liao, S., & Whitman, R. V. (1986). Overburden correction factor for SPT in sand. *Journal of Geotechnical Engineering*, 112(3), 373-377.
- [3] O'Neill, M. W., & Reese, L. C. (1999). *Drilled shafts: Construction procedures and design methods*. (Report no. FHWA-IF-99-025). Washington, D.C.
- [4] Liang, R., & Nowari, N. O. (2000). Resistance factors for driven piles. *ASCE Geotechnical Special Publication* (Report No. 100), 178-191.
- [5] Allen, T. M., Nowak, A. S., & Bathurst, R. J. (2005). *Calibration to determine load and resistance factors for geotechnical and structural design* (#E-C079). Washington, D.C.: Transportation Research Board.
- [6] Yang, K., & Liang, R. Y. (2007). Resistance factors for lateral design of drilled shafts. In *Proceedings of the DFI Annual Conference*. Denver, CO.
- [7] Yang, L. (2006). *Reliability-based design and quality control of driven piles*. (Unpublished doctoral dissertation). University of Akron, Akron, OH.
- [8] Liang, R., & Li, J. (2009). Resistance factors calibrated from FHWA drilled shafts static top-down tests data base. In *Proceedings of Selected Sessions of the International Foundation Congress and Equipment Expo*. Orlando, FL.
- [9] Barker, R. M., Duncan, J. M., Rojjiani, K. B., Ooi, P. S., Tan, C. K., & Kim, S. G. (1991). *Manuals for the design of bridge foundations* Report No. 343 NCHRP). Washington, D.C.: Transportation Research Board.
- [10] Ang, A. H-S., & Tang, W. H. (1990). *Probability concepts in engineering planning and design, vol. 2: Decision, risk and reliability*. New York: John Wiley & Sons.
- [11] Ang, A. H-S., & Tang, W. H. (2007). *Probability concepts in engineering, vol. 1, Emphasis on applications to civil and environmental engineering* (2nd ed.). New York: John Wiley & Sons.
- [12] Course hero. (2016). Retrieved from <https://www.coursehero.com/file/8207565/ANG2ed-1-r/>
- [13] Zhang, L. M., Li, D. Q. & Tang, W. H. (2005). Reliability of bored pile foundations considering bias in failure criteria. *Canadian Geotechnical Journal*, 42, 1086-1093.
- [14] Reese, L. C., & O'Neill, M. W. (1988). *Drilled shafts: Construction procedures and design methods* (Report No. FHWA-HI-88-042). Washington, D.C.
- [15] Ensoft, Inc. (2001). *SHAFT version 5.0 for Windows*. Austin, TX: Ensoft, Inc.

- [16] Paikowsky, S. G. (2004). *Load and resistance factor design (LRFD) for deep foundations* (Report No. 507, NCHRP). Washington, D.C.: Transportation Research Board.
- [17] Das, B. M., & Sobhan, K. (2014). *Principles of geotechnical engineering* (8th ed.). Boston: Cengage Learning.
- [18] Skempton, A. (2005). *Series 14—Skempton's published books and articles*. Retrieved from <http://www.cv.ic.ac.uk/SkemArchive/HDMS/Civils/Skempton/Web/SKEMS014.htm>
- [19] Seed, H. B., Lee, K. L., Idriss, I. M., & Makdisi, F. I. (1975, July). The slides in the San Fernando dams during the earthquake of February 9, 1971. *Journal of the Geotechnical Engineering Division, 101* (GT7), 651-688.
- [20] Peck, R. B., Hanson, W. E., & Thornburn, T. H. (1974). *Foundation engineering*. New York: John Wiley & Sons.
- [21] Bazarara, A. R. (1967). *Use of stander penetration test of estimating settlements of shallow foundation on sand*. (Unpublished doctoral dissertation). University of Illinois, Champaign-Urbana, IL.
- [22] Hasofer, A. M., & Lind, N. C. (1974). Exact and invariant second moment code format. *Journal of Eng. Mechanics Division, 100*(EM1), 111-121.
- [23] Ellingwood, B., Galambos, T. V., MacGregor, J. G., & Cornell, C. A. (1980). *Development of a probability based load criterion for American national standard A58* (Report No. 577). Washington, D.C.: National Bureau of Standards.
- [24] Phoon, K. K., Kulhawy, F. H., & Grigoriu, M.D. (1995). *Reliability-based design of foundations for transmission line structure* (Report No. TR-105000). Palo Alto, CA: Electric Power Research Institute.
- [25] Roberts L. A. (2006). *Reliability-based design and analysis of deep foundations at the service limit state*. (Unpublished doctoral dissertation). University of Missouri, Kansas City, MO.
- [26] Yang, X., Han, J., Parsons, R. L., & Henthorne, R. W. (2008). Resistance factors for drilled shafts in weak rock based on o-cell test data. *Soil Mechanics: Transportation Research Record No. 2045. Journal of the Transportation Research Board, 62-67*.
- [27] Nowak, A. (1995). *Calibration of LRFD bridge design code* (Report No. 368, NCHRP). Washington, D.C.: Transportation Research Board.
- [28] Li, J., & Liang, R. (2008). Skin resistance factors calibrated from o-cell tested drilled shafts of FHWA Database. In *Proceedings of the International Geotechnical Conference, Vol. 3* (pp. 771-778). Moscow, Russia.
- [29] Harr, M. E. (1996). *Reliability-based design in civil engineering*. Mineola, NY: Dover.
- [30] Baecher G. B., & Christian J. T. (2003). *Reliability and statistics in geotechnical engineering*. West Sussex, UK: Wiley.
- [31] Abu-Farsakh, M. Y., Xinbao, Y., & Zhongjie, Z. (2012). Calibration of side, tip, and total resistance factors for LRFD of drilled shafts. In *Proceedings of the TRB Annual Meeting*. Washington, D.C.
- [32] Brown, D. A., Turner, J. P., & Castelli, R. J. (2010). *Drilled shafts: Construction procedures and design methods* (Report No. FHWA NHI-10-016). Washington, D.C.: FHWA.
- [33] AASHTO (2012). *LRFD bridge design specifications* (6th ed.). Washington, D.C.: AASHTO.

Biographies

JILIANG LI is an assistant professor of civil engineering at Purdue University Northwest. He earned a DEng (PhD of Engineering Science) in Mining Engineering from the University of Science and Technology Beijing, China, and holds a PhD degree in Civil Engineering from the University of Alabama. Dr. Li has consulted on residential, commercial, and subdivision projects, as well as large-scale state DOT and federal projects. His interests include mining rock mechanics, engineering, civil/geotechnical engineering including statics, mechanics of materials, soil and rock mechanics, engineering geology, surface water hydrology, design of earth structure, and foundation engineering. Dr. Li may be reached at Jiliang.Li@pnw.edu

JINYUAN ZHAI is with the Department of Mechanical Engineering at the University of Akron. She earned her PhD in mechanical engineering from the University of Akron in 2016. Her dissertation was in modeling ductile damage of metallic materials. She has also worked as an assistant professor of mechanical engineering at the University of Science and Technology, Beijing. Dr. Zhai received her BEng and MEng degrees in engineering machinery from Taiyuan University of Science and Technology. She may be reached at jy21stc@gmail.com

LI TAN is a professor of electrical engineering at Purdue University Northwest. He received his MS degree in structural mechanics and MS and PhD degrees in electrical engineering from the University of New Mexico. Dr. Tan authored/co-authored two textbooks: *Digital Signal Processing: Fundamentals and Applications* (2nd ed.), Elsevier/Academic Press, 2013; *Analog Signal Processing and Filter Design*, (2nd ed.), Linus Publications, 2016. Dr. Tan has also served as an associate editor for the *International Journal of Modern Engineering* and the *International Journal of Applied Research and Innovation*. Dr. Tan may be reached at lizhetan@pnw.edu

A COMPARISON OF THE MINIMUM RESOLUTIONS OF TWO DIGITAL IMAGE CORRELATION-BASED TOOLS IN MAKING STRAIN MEASUREMENTS

Niranjan Desai, Purdue University Northwest; Christos Georgakis, Technical University of Denmark;
Gregor Fischer, Technical University of Denmark

Abstract

Structural health monitoring (SHM) is an important technique that helps structural engineers enhance the safety of critical structures and facilitates efficient maintenance of existing structures, while also assisting in the economic operation of the structure. SHM involves implementing a strategy that identifies and characterizes damage or undesirable performance in engineering structures. This investigation compared the minimum resolutions of two state-of-the-art digital image correlation-based software packages, ARAMIS and iMETRUM. The smallest strain accurately measurable using these tools was determined. This industry-affiliated study was performed in order to investigate initially undetected damage in the connections of the Storstrom Bridge in Denmark, whose detection would have prevented its propagation, resulting in lower repair costs.

Laboratory tests were performed on a specimen that modeled a steel beam-to-column connection of the bridge. In these tests, a shear force and bending moment were developed at the connection as it was loaded, and the corresponding strains that developed were measured using both conventional strain gauges as well as the two tools based on this digital image correlation (DIC). The minimum resolutions of both state-of-the-art systems used in this investigation were determined. Due to the challenges faced in making these small-strain measurements even under controlled laboratory conditions using ARAMIS, it was concluded that it would be unrealistic to use this tool in a real-world situation to measure strains as small as those that would need to be measured to detect the onset of damage in bridge connections.

The strains determined using iMETRUM were less sensitive to the presence of specimen vibrations produced by external perturbations in the vicinity of the specimen. Hence, to ascertain whether it is practical to use iMETRUM to measure small-strains in a real-world situation, more work needs to be done in making small-strain measurements using this tool in the presence of vibrations induced in the test specimen due to external disturbances.

Introduction

Structural health monitoring (SHM) is emerging as a vital tool to help civil engineers improve the safety and maintenance of critical structures and assists infrastructure owners with timely information for the continued safe and economic operation of their structures. SHM involves implementing a strategy that identifies and characterizes damage or undesirable performance in engineering structures. The conventional SHM process involves using global techniques that determine the health condition of structures by studying changes in their dynamic properties or responses. The system is observed over time by periodically sampling the dynamic response or deterioration obtained using an array of sensors. This is followed by extracting damage-sensitive features from these measurements and analyzing these features using statistical methods, thereby assessing the present condition (or current state of health) of the structure.

There also exist techniques to monitor the structural health at a local scale. Some of the techniques more commonly used on a local scale include the acoustic emission (AE) technique, the X-ray radiographic technique, and the interferometry technique [1-3]. A recent investigation of the condition of the Storstrom Bridge—which is a road and railway arch bridge, connecting Falster and Zealand in Denmark, and which is also on the rail line between Copenhagen, Denmark, and Hamburg, Germany—showed damage in local regions of the bridge (for example, the connections). On further investigation, it was inferred that this damage originated in regions of the connections hidden from the naked eye, such as behind gusset plates, and had propagated from there over time. The undetected damage in this bridge makes for an interesting and relevant situation that suggests that **local** structural damage may remain undetected using conventional SHM techniques that analyze changes in the global structural response parameters. Hence, this current study focused on applying a state-of-the-art DIC technique to monitor structural health, that is specifically applied to bridges, at a local level.

The goal of this investigation was to determine the minimum resolution of two state-of-the-art DIC systems by ascertaining the lowest possible strains that they are capable

of accurately measuring. The accuracy of the two systems in measuring small strains (strains in the range of 0.015% - 0.075%) was compared. It is important to know the minimum strain accurately measurable by a standard DIC-based SHM technique, so as to detect damage at its onset, thereby preventing it from spreading, as was observed in the case of the regions of the Storstrom connections that were hidden from external view. Since this investigation was motivated by connection damage in the Storstrom Bridge, the smallest strain accurately measurable using the DIC technique was determined by performing tests on a laboratory specimen that replicated a typical beam-column connection in the real bridge and computing the strains that developed in this connection using the DIC-based tools ARAMIS and iMETRUM.

This investigation expands upon an initial study that determined the accuracy of the DIC-based ARAMIS tool in measuring small-strains [4]. This investigation attempted to determine the accuracy of the DIC-based tool iMETRUM in making small-strain measurements of the same order of magnitude as those measured using ARAMIS, and also compared the accuracy of these two state-of-the-art tools. It was decided that iMETRUM would be used in the current investigation because, according to iMETRUM suppliers, iMETRUM is capable of overcoming many of the challenges faced in making small-strain measurements that were made using ARAMIS [4].

Background Work

As mentioned, the AE, the X-ray radiography, and the interferometry technique constitute some of the contemporary techniques used to monitor structural health on a local scale. Generally, it can be seen that the main challenges in applying these techniques are that they are expensive, labor-intensive, and involve having to make contact with the structure by placing sensors on it [1-3]. An alternative technique that can be used to monitor structural health involves the use of a photogrammetry system (using principles of image analysis and DIC), which applies a post-processing algorithm to analyze images of the structure captured during its loading phase, and computes the resulting strains and deformations in the structure [5, 6].

Photogrammetry is a non-contact measurement technology and has the following advantages [7]: it can be used to measure difficult-to-access structures; it has a fast measuring speed, and records large amounts of information in a short time by acquiring images; it is highly precise; it involves a simple process, and is less labor-intensive; and, it permits additional analysis of visual records at a later time.

Image analysis and DIC techniques involve first capturing a reference image of the object whose strains and deformations are to be measured. This represents the object in its undeformed state. As the object deforms, additional images are collected. The software package that is used to process these images recognizes the surface structure of the object and allocates coordinates to the image pixels. It compares the digital images and computes the object displacements and deformations. If the object surface has relatively few contrasting features or characteristics (homogenous surfaces), a stochastic color spray pattern is generally applied on the surface so that the different images can be compared and the strains and deformations computed [8]. Any random pattern can be used (for example, spray paint). However, the pattern must be such that there is an adequate contrast in the grayscale and surface pattern on the specimen surface [5, 6]. This can be achieved, for example, by using black and white spray paint to apply a random pattern on the surface of the object [5, 6].

The photogrammetry system captures the images during the loading process. Following this, it computes the deformation and the strain of the selected surface using a post-processing algorithm [5, 6]. There have been several successful applications of photogrammetry in the fields of aerospace engineering, chemistry, biology, architecture, and biomechanics, to name a few [7]. A non-contact photogrammetry system was used by Lin et al. [9] to measure the deformation of a pressure airship model. These structures are generally comprised of flexible materials, making them large and soft. Consequently, it is important to investigate the static and dynamic deformations of these structures in way that involves non-contact measurements.

Photogrammetry was used by Lee and Al-Mahaidi [10] to investigate the load-deformation characteristics of reinforced concrete T-beams strengthened with carbon fiber-reinforced polymer (CFRP) plates. Digital photogrammetry techniques were used to measure the vertical deflections of bridges by Jauregui et al. [11]. In all of the aforementioned cases, it was observed that the measurements obtained using the photogrammetry system showed a good correlation with experimental results and conventional measuring techniques. Also, Jauregui et al. [11] believe that photogrammetry in bridge engineering can provide an excellent alternative to conventional high-cost measurement systems. Close-range photogrammetry has not been as popular in bridge engineering applications as it has been in other areas [7]. Research activity on the application of this technique in bridge-related projects has been minimal and widely dispersed over the last 25 years [11]. However, there have been a few pioneering applications of the methodology applied in this field. There is a significant potential for growth in the

application of close-range photogrammetry in the field of bridge engineering. Additionally, the fast growth of digital imaging and computer technologies, the availability of inexpensive, off-the-shelf digital cameras, and soft-copy photogrammetry software systems has made more bridge engineering applications of photogrammetry involving deformation and geometry measurement a possibility [7].

On performing an overview of the cases involving the application of close-range photogrammetry to bridge engineering, it was observed that the field work for bridge geometry measurement has been reduced by more than 50%, while maintaining the same level of accuracy compared to conventional surveying methods [7]. It has also been found to be useful in historic bridge rehabilitation projects [7]. It can be concluded that close-range photogrammetry is a powerful, non-contact measurement technique that can provide unique solutions for a wide variety of bridge engineering applications [7]. There exist more recent applications of this technique in the field of bridge engineering. Waterfall et al. [12] looked at three case studies involving the application of DIC techniques towards obtaining measurements of bridges; namely, a post-tensioned concrete road bridge, a steel girder railway bridge, and a steel highway bridge. On completing these case studies, it was concluded that the measurements obtained using the advanced DIC system showed a good match with those obtained using conventional methods. Additionally, this technique makes it possible to measure large areas and several points, including inaccessible regions, efficiently, quickly, and in a cost-effective manner, making its implementation in industry more likely.

Recently, Busca et al. [13] compared several state-of-the-art vision-based measurement techniques (specifically, cameras) against traditional ones, by conducting tests on the response of an existing bridge during train transit. The goal of the research was to attempt to move from a 1D measurement at a single point, towards an approach where a single camera can obtain measurements in two or more points up to continuous measurement. It was observed that, for the single-point and multi-point measurements, the approach was promising, in that the limits in moving from single-point to two- and higher-point measurements were not that harsh. Also, a good representation of the dynamics of the system was obtained from the measurements, with the same limits of the displacement sensors. For the case of continuous measurement, an attempt was made to get the image of the complete structure using no targets at all, just working on the image contrast between the main beams and the background sky. It was concluded that the accuracy of the results obtained using this targetless approach was not as good as that obtained using targets. Yarnold et al. [14] implemented a comprehensive SHM system on the Tacony-Palmyra Bridge in Philadelphia, PA. The objective of the

project was to preserve the structure using developments in SHM technology. The instrumentation used to monitor the structure included electrical resistance and vibrating wire strain gages, tilt sensors, a weather station, and cameras that were placed at selected locations along the structure. A live web portal and a customized playback program were used for the integration of the data and video. The live web portal permitted real-time viewing of the data and video over the Internet. The playback program allowed specified events, such as bascule openings and overloaded vehicular passage, recorded by the structural monitoring software, to be viewed. With the long-term SHM system in place, it has become possible to make important structural measurements, like maximum strain release during bascule openings and maximum strain variation due to vehicular overload. The structure's deformation response to loading and environmental effects was also developed.

Methodology

A typical beam-to-column moment connection of the Storstrom Bridge was replicated in the laboratory. A point load directed vertically downward was applied on the beam at an eccentricity from the connection at the column. This developed a shear force and bending moment at the connection, in a manner that occurred in the real-life joint. As the load was gradually increased over the course of the test, strains were measured at six different locations at the connection using the DIC-based ARAMIS and iMETRUM tools, as well as conventional strain gauges (described subsequently). These measurements were taken at the same instant of time, so as to enable a comparison between the readings obtained by the DIC-based approach and the conventional strain gauges. The accuracy of the strains obtained using the DIC approach were determined by comparing these values against those obtained using the conventional strain gauges. Finally, the lowest strains accurately measurable using both the DIC-based ARAMIS and iMETRUM tools were subsequently obtained.

Figure 1 shows the laboratory specimen that was designed by Furlan [15] and which was comprised of three steel S355/Fe510 plates (10 mm thick), a main plate (850 mm long by 150 mm wide), and two other plates, referred to as secondary plates (both of which were 300 mm long by 150 mm wide). One edge of each of the two secondary plates was bolted to an edge of the main plate via nine high-resistance bolts. This resulted in the edge of the main plate being sandwiched between the two secondary plates. The other end of each of the two secondary plates was welded to a fixed plate, which was bolted to a support that held the entire system in place. The junction at which the edge of the main plate was sandwiched between the edges of the two

secondary plates represented a typical beam-to-column bolted connection, with the main plate representing a beam. From here on, the region where the main plate is sandwiched between the two secondary plates using bolts will be referred to as the connection. Additionally, the other end of the main plate (or beam) was left free, and the external point load was applied vertically downward at this location during the test.



Figure 1. Laboratory Specimen (Adapted From Furlan [15])
Note. From “Use of Digital Image Correlation for Steel Strain Section Evaluation,” by M. Furlan, 2014, unpublished master’s thesis, Technical University of Denmark. All images reprinted with permission.

The connection was designed by Furlan [15] in accordance with the Eurocode 3 (EC3) provisions. It was comprised of nine high-resistance bolts of class 8.8. This configuration ensured that the stress in the bolts remained below their yield stress. Consequently, the bolts were in the same initial condition every time the test was repeated. The load was applied at the free end of the beam using a mechanical setup. A hydraulic jack was also available as another option. While being easier to use, it was harder to control than the mechanical setup over the low range of loading being applied in this experiment (a maximum load of 13.10 kN was applied). Furthermore, the vibrations created by the mechanical setup were less than those caused by the hydraulic jack. Vibrations led to noise in the experimental output (strain-versus-load diagram), which was undesirable. Using the mechanical setup, the load was manually increased by incrementing its value in a stepwise manner. The load cell was calibrated and the load values, which were read in mV, were converted to the corresponding values in kN. The process used to calibrate the load cell is described in the next subsection. The specimen was loaded to 70% of its yield stress. This ensured that it remained undamaged over the

entire course of the test, and was in the same condition every time the test was repeated. The value of the loading that caused the specimen to be at 70% of its yield stress was determined by Furlan [15]. He developed a complex finite element model that incorporated the essential characteristics of the real specimen, using the ABAQUS software package. He concluded that the specimen yielded at 18.71 kN. Hence, in the current test, the specimen was loaded in increments to a maximum load of $0.7 \cdot 18.71 = 13.1$ kN

The load cell needed to be calibrated, since its output was in the form of voltage (in units of mV), which needed to be converted to equivalent units of force (kN). The load cell was calibrated by progressively loading it in five increments of 2.5 kN and one increment of 0.75 kN, noting the corresponding output voltage readings in mV. Thus, the load cell was loaded to a final value of 13.25 kN, which was slightly greater than the maximum value of load applied to the specimen during the tests (13.1 kN). These readings were plotted to develop a graph of load in kN (y-axis) versus voltage in mV (x-axis). Figure 2 shows this plot, which is linear, since the specimen was in its elastic region of response. Based on this calibration, the following linear relationship was obtained that was subsequently used to convert the load-cell voltage output in mV into an equivalent value of force in kN:

$$Force(kN) = [5780 * voltage(mV)] + 39.877$$

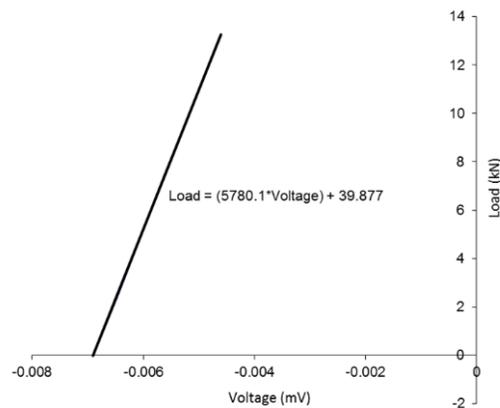


Figure 2. Calibration of the Load Cell

As the specimen was loaded, strains were measured at six locations in the connection at which high-stress variations were observed. These locations were determined by Furlan [15] and based on the results he obtained during his ABAQUS-based finite element model analysis of the specimen. These strains were measured using both conventional strain gauges as well as DIC-based techniques. The strain gauge readings obtained were used as a reliable reference to which the strains measured by the DIC-based approach

were compared, and their accuracy determined. In order to accomplish this task, six strain gauges were attached on the same face of the specimen at the connection (see Figure 3): three were attached to the outer surface of one of the secondary plates at the connection, and the remaining three were attached to the main plate. Digital images were captured of the opposite face of the connection on which the strain gauges were placed, thereby enabling the computation of strains in hidden regions. On both the main plate and the secondary plate in the connection region, the locations at which the strain gauges were attached were selected so as to include the following regions:

- A region in which tensile strain in the main plate (beam) develops. Hence, one strain gauge was placed near the top (upper) edge of the beam, where the tensile strain was maximum.
- A region in which a compressive strain in the beam develops. Hence, one strain gauge was placed near the bottom (lower) edge of the beam, where the compressive strain was maximum.
- A region in the beam (within the connection region) very close to the neutral axis, where the value of strain was minimum.
- A region in which tensile strain in the secondary plate develops. Hence, one strain gauge was placed near the top (upper) edge of the secondary plate in the connection, where the tensile strain was maximum.
- A region in which a compressive strain in the secondary plate develops. Hence, one strain gauge was placed near the bottom (lower) edge of the secondary plate at the connection, where the compressive strain was maximum.
- A region in the secondary plate (within the connection region) very close to the neutral axis, where the value of strain was minimum.

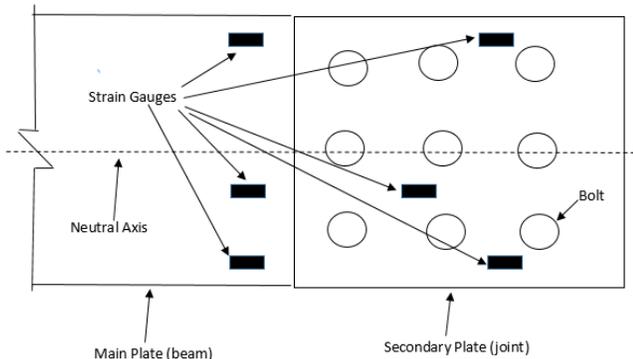


Figure 3. Location of the Strain Gauges

At the connection, the strains developed in the secondary plates were much smaller than those developed in the main

plate (beam), because, in the former case, there were three plates bolted together, while, in the latter case, there was only one plate. Hence, this positioning of strain gauges provided a wide range of strain measurements in different regions of the connection: a) maximum connection tensile strains in both the beam (main plate) and the connection (secondary plates), the strains in the secondary plate being of a significantly lower value; b) maximum connection compressive strains in both the beam and the secondary plates at the connection, the strains in the secondary plate being of a significantly lower value; c) minimum connection strain values near the neutral axis, in both the beam as well as the secondary plate, the strain in the secondary plate being significantly lower than that in the beam.

Like the load cell output, the strain gauge output readings were also in mV. Consequently, the strain gauges were also calibrated in order to convert the output mV readings into their equivalent readings in units of strain (length/length). Traditional electrical resistance strain gauges provided by the laboratory that had a gauge factor, k , of 2.13 were used. For a single active gauge and three dummy resistors in a Wheatstone bridge configuration, strain was related to the output voltage (mV), gauge factor, and input voltage (5 mV) as follows: $\text{strain} = [(4 * \text{output voltage}) / (k * \text{input voltage})]$

Strain Measurement Using the DIC-Based Approach

As described earlier, the strains that developed in the specimen were measured using both traditional strain gauges (to provide a reference for comparison) and a DIC-based approach, using both the ARAMIS and iMETRUM DIC-based tools. The readings using the latter approach were compared against those obtained using the former. In order to make strain measurements using ARAMIS, digital images of the specimen were captured as it deformed under a progressively increasing load using a Nikon D800 digital camera (see Figure 4), supported by a tripod stand (see Figure 5). These images served as the raw data that was processed by the ARAMIS DIC-based algorithms to determine the specimen strain.

In order to make strain measurements using iMETRUM, a video of the specimen was captured as it deformed under a progressively increasing load using a GigE PoE camera (specifications: frame rate = 15 fps, resolution = 2452 x 2056 pixels) supported by a tripod stand. The video served as the raw data that was processed by the iMETRUM DIC-based algorithms to determine the specimen strain. This camera was recommended by the iMETRUM supplier for the strain magnitudes being measured in this test.

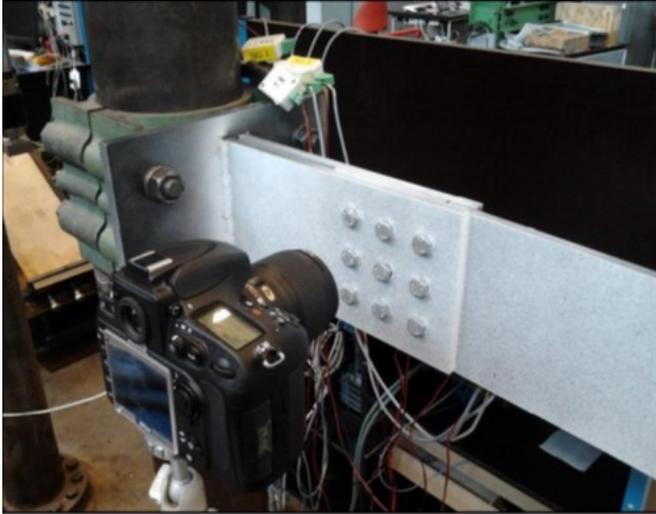


Figure 4. Placement of the Camera Relative to the Specimen (Adapted From Furlan [15])



Figure 5. Tripod Stand Used to Support the Camera (Adapted From Furlan [15])

The following parameters influenced the accuracy of the strains measured (relative to the strain gauge readings) on post-processing the digital images of the specimen:

- The focal length of the camera lens: While making the strain measurements using ARAMIS, three lenses with focal lengths of 24 mm, 60 mm, and 105 mm were used; the best results were obtained using the lens with the 60 mm focal length. A 60 mm lens was also recommended by Furlan [15] for this applica-

tion. The iMETRUM supplier recommended and provided a high-resolution, low-distortion lens, having a focal length 35 mm, for the strain magnitudes being measured in this test. Hence, this lens was utilized.

- The distance of the camera from the face of the specimen: Another important parameter that influenced the quality of the output readings was the distance of the camera from the face of the connection. Each location of the specimen to which a strain gauge was attached was captured by the camera in isolation. Initially, it was attempted to capture all six locations at once. However, in order to accomplish this, the camera needed to be placed at a distance further away from the specimen than the optimum distance (see next paragraph) that gave the best readings. By doing so, the strains obtained on post-processing the images were inaccurate, showed a significant amount of noise, and did not compare well with the strain gauge readings. This was because the values of strain being measured in this experiment were very low and the camera had to be placed as close to the specimen as possible in order to accurately capture these values without losing focus of the specimen. Another issue with focusing on all six strain gauges in one picture was that the two plates of the connection that the camera was digitally capturing did not lie in the same plane, thus it was not possible to bring them both into focus in the same image.

Due to the very small values of strain being measured in this experiment, it was necessary to place the camera close to the face of the connection. In order to find the optimum value for this distance, several tests were performed by placing the camera at distances ranging from 7.5 to 20 cm from the face of the connection. Based on the test results, it was found that a distance of 10 cm from the face of the connection gave optimum results. At distances greater than this value, the strain readings obtained after post-processing the images were inaccurate in comparison to the traditional strain gauge readings that were used as a benchmark. At distances less than 10 cm from the connection face, it was not possible to bring the specimen into focus. A value of 10 cm was also recommended by Furlan [15] in order to obtain good results. An automatic focus camera was used in the ARAMIS tests.

While making the measurements using iMETRUM, the surface of the lens was placed at approximately 28 cm from the face of the specimen. This is a standard distance recommended by the supplier for the magnitude of strain being measured. Unlike the case of the ARAMIS camera, the iMETRUM camera did not focus on the specimen automati-

cally. In order to bring the specimen into focus, the technique recommended by the iMETRUM supplier was used. Placing the lens 28 cm away from the specimen surface, any external light source illuminating the specimen was turned off and the specimen was exposed only to the ambient light in the laboratory. Under these conditions, the camera shutter was completely opened. Now, the specimen surface was viewed through the camera. The surface appeared blurry and unfocused. The lens was gradually moved back and forth by a short distance (falling within about 5 cm) until the specimen surface came into focus.

- The spray pattern applied across the specimen surface: ARAMIS and iMETRUM process the digital images captured during the test by tracking the movement of single pixels and comparing the successive photographs obtained at each stage of loading. In order for the software to perform this task, it is essential that there be an adequate contrast in the grayscale and surface pattern on the specimen surface. Figures 6 and 7 show the final spray patterns used in two locations. These helped the researchers to visualize the spray pattern used. This was accomplished by using two different colors—a uniform white background on the specimen surface to conceal any imperfections on its surface, and a black spray that was discontinuously applied over the white background. This created black points whose displacements could be tracked by the software algorithm. Furlan [15] investigated the effect of the spray pattern on the accuracy of the results and, after trying several different patterns, observed that the use of a finer pattern led to more accurate results. Consequently, Furlan's pattern was used in this investigation.

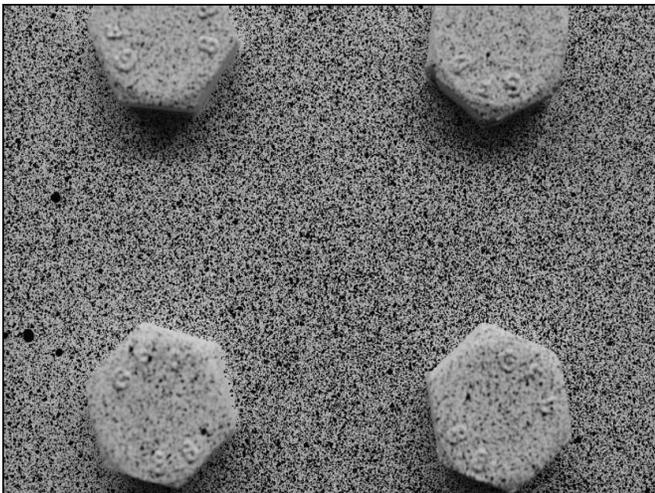


Figure 6. Final Spray Pattern—Joint Middle, Adapted From Furlan [15]

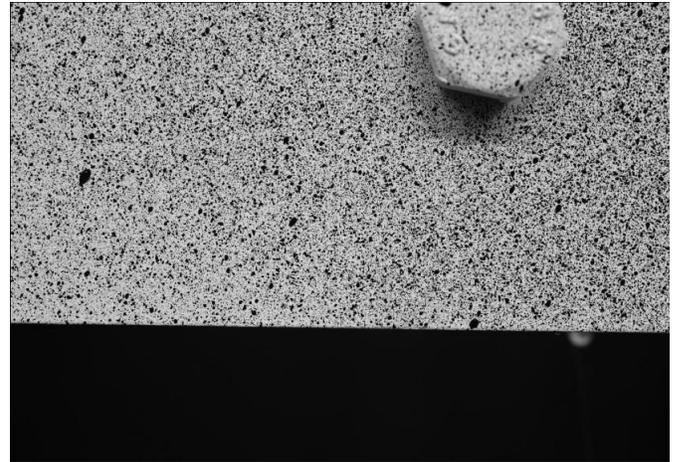


Figure 7. Final Spray Pattern—Joint Bottom, Adapted From Furlan [15]

- The existence of external vibrations while conducting the test: The camera was programmed to automatically capture an image every 45 seconds. Durations of 30 seconds, 60 seconds, and 75 seconds were also tried. It was observed that 45 seconds was an optimum time interval, since it was long enough to give the specimen enough time to settle after each manually applied load increment, yet of an appropriate duration to complete the experiment in a reasonable amount of time. Furthermore, programming the camera to automatically capture the images circumvented the need to make contact with the camera manually to capture each image. The experiment was extremely sensitive to vibrations, and even small perturbations caused by the most delicate of physical contact with the sensor created noise in the output readings. iMETRUM was less sensitive than ARAMIS to vibrations in the test specimen. iMETRUM made a continuous video recording of the deformation of the specimen as it was being loaded. This video recording served as the raw data that was processed by the DIC algorithms incorporated in the iMETRUM software. Vibrations existing in the specimen were captured in the video recording. While processing the video recording, the iMETRUM algorithms filtered out the noise created by the specimen vibrations.
- The lighting: While performing the strain measurements using ARAMIS, the specimen surface being digitally captured was brightly illuminated using two external light sources, in addition to any ambient light that already existed in the laboratory. In order to obtain accurate readings, it was important to ensure that the entire surface of the connection was uniformly illuminated and there were no shadows from nearby objects. An LED panel light and stand were pro-

vided by the suppliers of iMETRUM. Moderate illumination uniformly applied across the specimen surface, about as intense as the ambient light, was adequate for the iMETRUM measurements.

- The relative orientation of the specimen and lens faces: In order to obtain accurate readings, it was essential that the face of the camera lens be oriented parallel to the face of the connection being photographed. This was accomplished using a built-in feature of the digital camera, such that when viewed through the camera, the region of the specimen in focus was covered with red squares. A high concentration of red squares across the surface meant that the specimen surface in focus was parallel to the camera. A sparse concentration of red squares across the specimen surface implied that the camera position required adjustment in order to ensure that it was parallel to the specimen surface. In the case of iMETRUM, the lens surface was manually placed approximately parallel to the specimen face. A slight relative inclination between the two surfaces did not affect the accuracy of the strains measured using iMETRUM, since its algorithms were designed to account for this.

Several tests were conducted at each strain gauge location, and the strains obtained using ARAMIS and iMETRUM were compared with those obtained using traditional strain gauges at each location of the specimen under scrutiny. In order to be concise, only representative results at each location are presented. To save space, more than one set of resulting plots are presented at only the middle of the joint, where results are shown from two different tests. This was the stiffest part of the specimen and the strains here were consequently the lowest, thereby reaching the limits of the DIC-based techniques in measuring strains.

Analysis and Results

Figures 8-14 show comparisons between the strains measured using the DIC-based tools, ARAMIS and iMETRUM, and those measured using the traditional strain gauges. Table 1 shows the peak strain measured using ARAMIS, iMETRUM, and the strain gauges at the six locations investigated. Table 2 shows the percentage change between these peak strain values and those measured using the strain gauge value as the reference value.

Table 1. Peak Strains Measured using the DIC-Based Approach

Location	Strain (%): ARAMIS		Strain (%): iMETRUM		Strain (%): Strain Gauge		Type
	Test 1:	Test 2:	Test 1:	Test 2:	Test 1:	Test 2:	
Joint Middle	0.027	0.0116	0.022	0.0212	0.0149	0.0145	Compressive
Beam Middle	0.0245		0.0285		0.0264		Compressive
Joint Bottom	0.0298		0.03		0.0309		Compressive
Joint Top	0.0255		0.0299		0.029		Tensile
Beam Bottom	0.0723		0.0725		0.072		Compressive
Beam Top	0.0725		0.0745		0.073		Tensile

Table 2. Percentage Change between Peak DIC-Strains, Relative to the Strain Gauges

Location	% Change ARAMIS		% Change iMETRUM	
	Test 1:	Test 2:	Test 1:	Test 2:
Joint Middle	44.82 (+*)	25 (-*)	47.65 (+)	46.21 (+)
Beam Middle	7.2 (-)		7.95 (+)	
Joint Bottom	3.56 (-)		2.91 (-)	
Joint Top	12.07 (-)		3.1 (+)	
Beam Bottom	0.41 (+)		0.69 (+)	
Beam Top	0.685 (-)		2.055 (+)	

* (+/-) = increase/decrease relative to strain gauge value

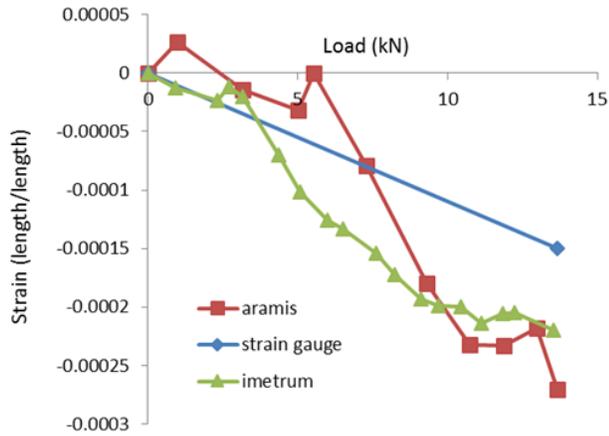


Figure 8. Comparison of Strains in the Middle of the Joint (Test 1)

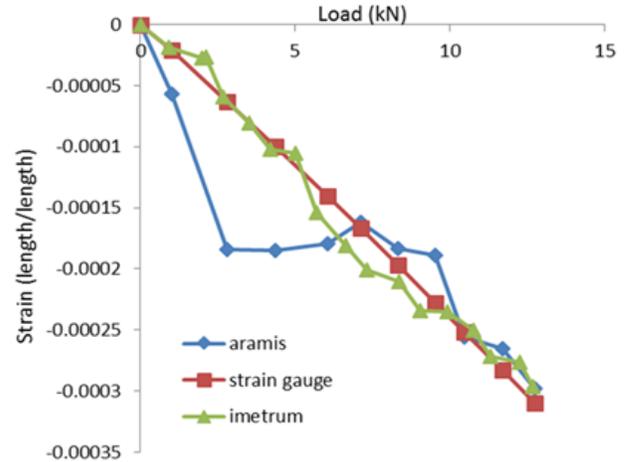


Figure 11. Comparison of Strains at the Bottom of the Joint

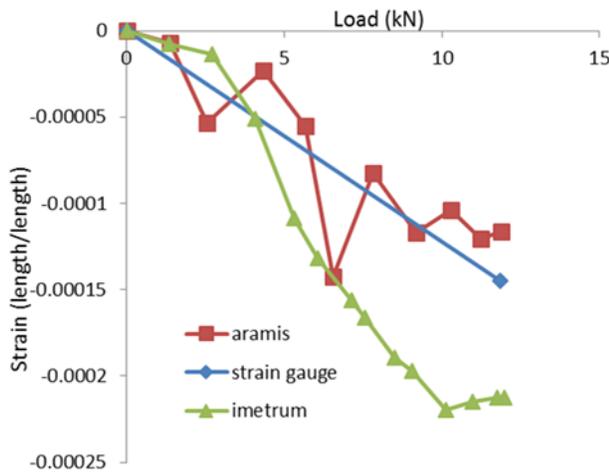


Figure 9. Comparison of Strains in the Middle of the Joint (Test 2)

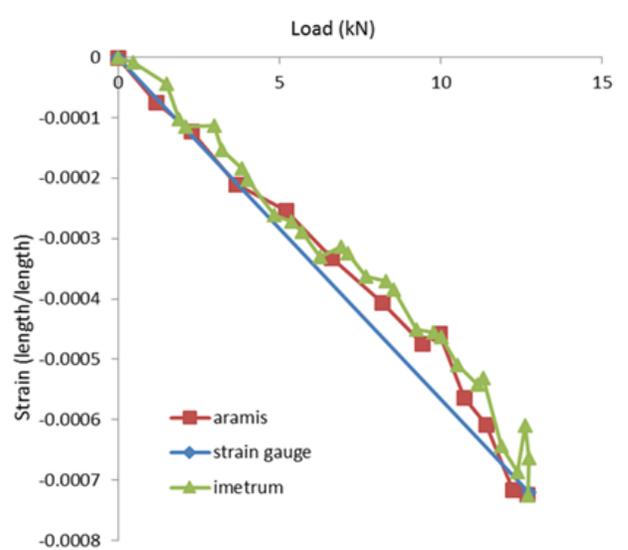


Figure 12. Comparison of Strains at the Bottom of the Beam

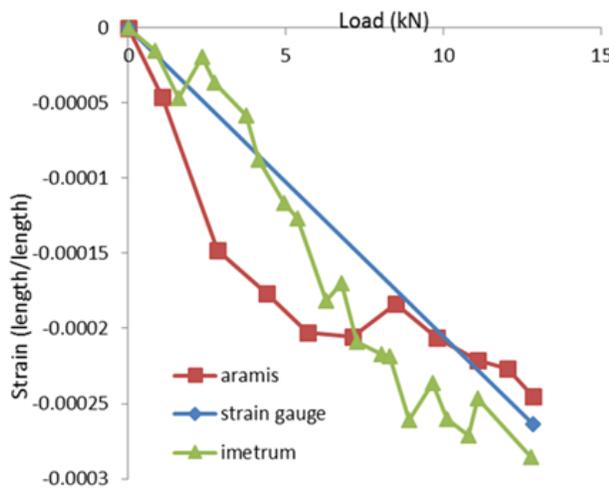


Figure 10. Comparison of Strains in the Middle of the Beam

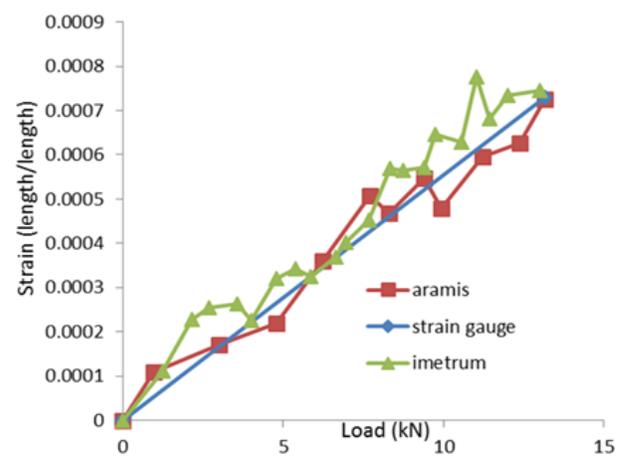


Figure 13. Comparison of Strains at the Top of the Beam

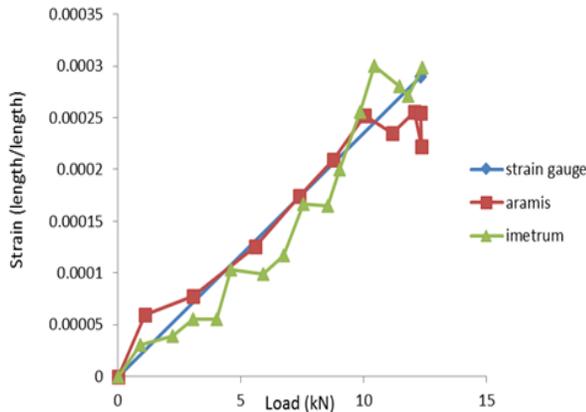


Figure 14. Comparison of Strains at the Top of the Joint

Based on the measured strain values, the accuracy of the DIC-based approach was determined for measuring strains lying within three ranges: 0-0.015% (by using the results in Figures 8 and 9); 0.015%-0.03% (by using the results in Figures 10, 11, and 14); and, 0.03%-0.075% (by using the results in Figures 12 and 13). R-squared (coefficient of determination) values were computed for each of the plots. At each strain gauge location, these values served as a measure of the closeness of the DIC-based strains to the straight-line model obtained via the strain gauge measurements. Since Figures 8 and 9 display strains having values in the range of 0-0.015%, of all the figures presented, they provided the most meaningful means for comparing strains having a value less than or equal to 0.015% (range 1 mentioned previously). While computing the R-squared values for the strains in ranges 2 and 3 discussed above, only those data points corresponding to strain values that lay within the range under scrutiny were considered in the R-squared computation. Consequently, the results in Figures 10, 11, and 14 were used to compare strains in the range of 0.015-0.03% by considering only those values of strain greater than 0.015%, and not considering strain values less than 0.015% in these figures.

Similarly, since the results in Figures 10, 11 and 14 were used to compare the strains in the range of 0.015%-0.03%, the values in Figures 12 and 13 were used to determine the accuracy of strains measured by the DIC-based approach for strains lying within the range of 0.03%-0.075% by considering only those strain values in Figures 15 and 16 greater than 0.03%, and not considering those values less than 0.03% in the R-squared value computation. Table 3 shows the R-squared values.

Discussion

1. The strain gauge responses showed a linear strain versus load behavior. This was due to the fact that the material was loaded within its elastic range of response.
2. In all of the figures presented (Figures 8-14), the strain versus load response obtained from the DIC-based ARAMIS and iMETRUM readings followed a general pattern, similar to the strain versus load response obtained using the strain gauges. Additionally, the maximum values of strain measured by the DIC-based approach were nearly identical to those measured by the strain gauges for strains in the range of 0.03%-0.075%. The percentage change between these maximum values (see Table 2, figures obtained by using the strain gauge value as the reference value) was most pronounced for strain measurements taken at the middle of the joint, with a difference of 44.82% in test 1 and 25% in test 2 using ARAMIS, and 47.65% and 46.21% using iMETRUM. This is reasonable, since the middle of the joint was the stiffest of all the locations investigated and developed the lowest strains in the specimen, on the order of approximately 0.015%. At locations that developed strains greater than 0.015%, and falling within the range of 0.015%-0.03%, the percentage changes between the maximum values measured by the strain

Table 3. R-Squared Values

Figure	Strain Range (%)	Type	R-squared ARAMIS	R-squared iMETRUM
8	0 – 0.015	compressive	0.5728	0.3772
9	0 – 0.015	compressive	0.7015	0.388
10	0.015 – 0.03	compressive	0.7382	0.8461
11	0.015 – 0.03	compressive	0.6827	0.98
12	0.03 – 0.075	compressive	0.9725	0.91
13	0.03 – 0.075	tensile	0.9716	0.8952
14	0.015 – 0.03	tensile	0.9252	0.9345

gauges versus those measured using the DIC-based approach predictably decreased and ranged from 3.56%-12.07% using ARAMIS and 2.91%-7.95% using iMETRUM. Finally, for even higher strains in range 3 (falling within the range of 0.03%-0.075%), the percentage change between peak strains measured using both approaches reduced to less than 0.7% in the case of ARAMIS and about 2% in the case of iMETRUM. Thus, purely from the perspective of peak strain computed by the DIC-based approach, it can be seen that this approach can accurately determine strains to a value of approximately 0.03% using both ARAMIS and iMETRUM, with the error in the peak strain measurement being less than 0.7% in the former case and no more than 2% in the latter.

3. Of all the comparisons shown in Figures 8-14, the most pronounced variation between the values of the strains obtained using the DIC-based approach versus the strain gauge values occurred in the middle of the joint (Figures 8 and 9). This is reasonable, since the middle of the joint was the stiffest of all the regions investigated and developed the lowest strains in the specimen, on the order of approximately 0.015%.
4. At the middle of the beam and the top and bottom of the joint (Figures 10, 11, and 14), where the peak strains fell within the range of 0.025%-0.03%, it can be seen that the strains measured by the DIC-based approach using ARAMIS resemble the strain gauge measurements more closely for strain values greater than approximately 0.015% (i.e., within the range of 0.015%-0.03%). This is reasonable since, based on the discussion in points 2 and 3, of all the locations studied in this investigation, the DIC-based approach was the least accurate for measuring strains less than 0.015%. However, in the case of iMETRUM, the strain measurements showed a greater resemblance to the strain gauge values for strains less than 0.015% in these plots than those shown by the ARAMIS values. Overall, in these plots, the pattern displayed by the iMETRUM strain resembled the pattern displayed by the strain gauge measurements and were better than the ARAMIS strains.
5. Figures 12 and 13 show the comparisons between the DIC-based strains and the strain gauge values for strains having a maximum value of approximately 0.073%. While interpreting these figures (see also point 6), it is meaningful to focus on the comparison of strains falling within the range of approximately 0.03%-0.075%, since the other figures focus on strains less than 0.03%. The peak strains in all figures, except for Figures 12 and 13, were within the range of 0.015%-0.03% and permitted a closer analysis and magnified view of strains of this order of magnitude.
6. The R-squared values presented in Table 3 were used as being representative of the accuracy of the DIC-based approach. These values lie between 0 and 1. A higher R-squared value implies a more accurate DIC-based strain measurement. The R-squared values obtained from Figures 12 and 13 were 0.9725 (at the bottom of the beam) and 0.9716 (at the top of the beam) for the ARAMIS measurements, and 0.91 and 0.8952 for the iMETRUM measurements at the same locations. As discussed in point 5, comparisons at these locations provide meaningful results for strains falling within the range of 0.03%-0.075%. Hence, it can be concluded that the DIC-based approach most accurately measured strains in the range of 0.03%-0.075%, to an accuracy of approximately 97% using ARAMIS, and about 90% using iMETRUM.
7. Comparisons at all of the other locations (other than the top and bottom of the beam) involved comparing strains that fell within the range of 0-0.03%. By subdividing this range into two smaller intervals, (0-0.015% and 0.015%-0.03%), it can be seen that the average R-squared value for the interval 0-0.015% was 0.6375 for the ARAMIS measurements and 0.3826 for the iMETRUM measurements. Additionally, the average R-squared value for the interval 0.015%-0.03% was 0.782 for the ARAMIS measurements and 0.9202 for the iMETRUM measurements. Hence, the DIC-based strain values were more accurate in the range of 0.015%-0.03% than they were in the range of 0-0.015%, which is reasonable, since the strain values in the former range are larger than those in the latter. Hence, it can be concluded that the DIC-based approach using ARAMIS does not produce accurate strain measurements for strains in both the 0-0.015% range (average R-squared = 0.6375) and 0.015-0.03% range (average R-squared = 0.782). Additionally, it can be concluded that the DIC-based approach using iMETRUM does not accurately measure strains less than 0.015% (average R-squared = 0.3826). However, for strains in the 0.015-0.03% range, the average R-squared values of 0.9202 obtained for the iMETRUM measurements leads to the conclusion that iMETRUM can be used to measure strains in this range to an accuracy of over 90%.
8. It can, therefore, be concluded that the minimum strain (in a region of a structure hidden from the na-

ked eye) that can be measured accurately by the DIC-based approach using the ARAMIS tool is 0.03%, and 0.015% for the iMETRUM tool is. This value of strain represents the minimum resolution of these systems in measuring strains.

Possible Causes of Error

1. Vibrations in the test specimen caused during manual load application: The strain readings computed using the DIC-based ARAMIS were very sensitive to any disturbances in the experimental specimen. Since the loads were applied manually, load application created vibrations in the specimen. Digital images were captured every 45 seconds in order to allow the specimen to settle and the vibrations to cease. However, it is possible that there were some vibrations, albeit negligible, in the specimen at the time of capturing the digital images. Measurements made using iMETRUM were less sensitive to the presence of specimen vibrations.
2. Vibrations in the test specimen due to external movement and disturbances in the lab (examples: other tests being conducted in the laboratory, large lab door being closed, people walking around, floor vibrations, etc.): While conducting the tests, it was observed that external vibrations created a significant fluctuation in the output strain readings obtained using the DIC-based ARAMIS. Despite attempting to create a highly controlled environment in order to minimize any external perturbations, it was not possible to completely isolate the specimen. Hence, small effects due to external vibrations could have marginally influenced the DIC-based strain measurements made using ARAMIS. The strain measurements made using iMETRUM were less sensitive to the presence of external vibrations in the vicinity of the test specimen.
3. During the calibration of the strain gauges, the mV output reading from the strain gauges was converted into a strain value by using a strain gauge factor of 2.13, provided by the supplier. It is possible that there were minor differences between the theoretical value and the real value. Hence, the strain gauge readings were not perfect.
4. Strain gauges were located on the rear surface of the specimen, and pictures were taken on the front surface in order to measure strains in the hidden region. Additionally, while using the software to compute the specimen strains (at the locations investigated in this

study) after it processed the digital images, it was required to specify the location on the specimen's image on the software's graphical user interface at which the strain was desired. Since the strain gauges were located on the rear surface of the specimen, their exact location on the image could not be pinpointed with perfect accuracy. The strains were computed at locations that were very close to their exact locations. Small differences between the exact locations and the ones at which the strain values were obtained using the DIC-based tools could have caused minor variations between the strain gauge values and those obtained using the DIC approach.

5. While capturing the images, it was essential that the surface of the specimen being photographed was perfectly parallel to the camera surface. This was ensured at the start of the experiment. However, there was an extremely small movement of the specimen in the z direction, as the load was applied. This minute movement could have possibly introduced some minor error in the DIC-based calculations, since it resulted in the loss of the initial perfectly parallel orientation of the specimen surface with respect to the sensor surface.
6. The camera has to be placed perfectly parallel to the ground. Even a slight tilt can cause errors in the strain measurements.

Comparison between ARAMIS and iMETRUM

Vibrations of the specimen affected the ARAMIS measurements more than they did the iMETRUM measurements. Hence, it was easier to make these measurements using iMETRUM than it was using ARAMIS. iMETRUM made a continuous video recording of the deformation of the specimen as it was being loaded. This video recording served as the raw data that was processed by the DIC algorithms incorporated in the iMETRUM software. Vibrations existed in the specimen and were captured in the video recording. While processing the video recording, the iMETRUM algorithms filtered out the noise created by the specimen vibrations. While using ARAMIS, however, discrete digital images of the specimen captured at different stages of the loading process served as the raw data for the algorithms. There was a waiting period between the capture of consecutive images, because even the slightest disturbance in the specimen adversely affected the strain measurements produced by ARAMIS. Hence, considerable caution needed to be exercised during the performance of the test in the case of ARAMIS. This was not so in the case of iMETRUM. As

there was no waiting period between the application of load increments in the iMETRUM test, the duration of the test was also shorter than in the case of ARAMIS. The test took about 3-4 minutes in the case of iMETRUM and about 12-15 minutes in the case of ARAMIS. Since several tests were performed at each strain gauge location in order to obtain a consistent set of good results, this time difference was an important parameter. Finally, the measurements made using iMETRUM were not as sensitive to the intensity of the external light source as those made using ARAMIS. The only requirement was that the light needed to be applied uniformly across the specimen. In the case of the tests performed using ARAMIS, two light sources were used to brightly illuminate the specimen in order to obtain more accurate readings. iMETRUM can be used to make strain measurements in about one-third the amount of time as ARAMIS; it has a better minimum resolution in measuring strains than ARAMIS does (0.015% using iMETRUM versus 0.03% using ARAMIS); it is easier to use and does not require the user to be as sensitive to perturbing the specimen as ARAMIS does; and, it can be used to make measurements under the effect of a light source of lower intensity than is needed by ARAMIS.

Summary and Conclusions

Damage observed in the connections of the Storstrom Bridge in Denmark was hidden from the naked eye and, hence, was not appropriately addressed at its onset, leading to its propagation. Identification of such damage at its onset can prevent its spread, thereby saving money and time on repairs, and also increasing the safety of the structure. Thus far, DIC-based SHM techniques have been used to monitor the structural condition of bridges on a global scale (for example, by measuring global deflections of bridges). In doing so, local damage such as that observed in the Storstrom Bridge can go undetected. Consequently, in this investigation, the DIC technique was applied in analyzing the strains of a small, localized region (a beam-to-column connection) of the Storstrom Bridge hidden from the naked eye (behind the gusset plates), and the lowest possible strains that could be accurately measured using this methodology were determined.

A beam-to-column connection in the Storstrom Bridge was replicated in the laboratory and subjected to a progressively increasing load that produced a shear force and bending moment at the connection. The strains that developed in the connection were measured using state-of-the-art DIC-based ARAMIS and iMETRUM tools, and were compared with the strains measured using strain gauges as a reference. It was determined that the DIC-based ARAMIS and iMETRUM tools can accurately measure strains as low as

0.03% and 0.015%, respectively. Due to the challenges faced in making these small-strain measurements with ARAMIS, even under highly controlled laboratory conditions, it was concluded that it is currently unrealistic to use ARAMIS in a real-world situation to measure strains as small as those that would need to be measured in order to detect the onset of damage in bridge connections. The strain measurements made using iMETRUM were less sensitive to the presence of vibrations created in the specimen, due to external moving loads. Consequently, in order to determine whether it is realistic to use iMETRUM in a real-world scenario involving the presence of vibrations in the structure being monitored that are produced by external moving loads, additional work needs to be done where the minimum strain that can be measured using iMETRUM is determined when the specimen is subjected to external vibrations.

References

- [1] Kaphle, M., Tan, A., & Thambiratnam, D. (2009). Structural health monitoring of bridges using acoustic emission technology and signal processing techniques. *Proceedings of the 13th Asia Pacific Vibration Conference*, New Zealand.
- [2] Ahlborn, T. M., Shuchman, R., Sutter, L., Brooks, C., Harris, D., Burns, J., et al. (2010). *The State-of-the-Practice of Modern Structural Health Monitoring for Bridges*. Report prepared at the Department of Civil and Environmental Engineering, Michigan Tech Transportation Institute, Michigan Technological University, Houghton, Michigan.
- [3] Gentile, C., & Bernardini, G. (2010). An interferometric radar for non-contact measurement of deflections on civil engineering structures: laboratory and full-scale tests. *Structure and Infrastructure Engineering*, 6(5), 521-534.
- [4] Desai, N. (2016). Small-Strain measurement in bridges using the digital image correlation (DIC) technique. *Proceedings of SPIE 9805, Health Monitoring of Structural and Biological Systems (980530)*.
- [5] Winkler, J., Fischer, G., & Georgakis, C. T. (2012). Localized bending fatigue behavior of high-strength steel monostrands. *Proceedings of the 6th International Conference on Bridge maintenance, Safety and Management (IABMAS)*, Italy.
- [6] Winkler, J., Fischer, G., & Georgakis, C. T. (2012). Monitoring of localized deformations in high-strength steel cables. *Proceedings of the NDE/NDT for Highways and Bridges: Structural Materials Technology (SMT) 2012 Conference*, New York, USA.
- [7] Jiang, R., Jauregui, D. V., & White, K. R. (2008). Close-range photogrammetry applications in bridge measurement: Literature review. *Measurement: Jour-*

nal of the International Measurement Confederation (IMEKO), 41, 823-834.

- [8] Perez, L. S., Fischer, G., & Georgakis, C. T. (2012). *Determination of Localized Strains and Global Deformations Using Video System*. Master's Thesis, Technical University of Denmark, Kgs. Lyngby, Denmark.
- [9] Lin, G., Tan, H., & Wang, C. (2008). Pressure Airship Model Structural Analysis and Deformation Measurement Using Photogrammetry. *Proceedings of the International Conference on Experimental Mechanics*, Nanjing, China.
- [10] Lee, T. K., & Al-Mahaidi, R. (2008). An experimental investigation on shear behaviour of RC T-beams strengthened with CFRP using photogrammetry. *Composite Structures*, 82, 185-193.
- [11] Jauregui, D. V., White, K. R., Woodward, P. E., & Leitch, K. R. (2003). Noncontact Photogrammetric Measurement of Vertical Bridge Deflection. *Journal of Bridge Engineering*, 8(4), 212-222.
- [12] Waterfall, P. M., Macdonald, J. H. G., & McCormick, N. J. (2012). Targetless precision monitoring of road and rail bridges using video cameras. *Proceedings of the 6th International Conference on Bridge maintenance, Safety and Management (IABMAS)*, Italy.
- [13] Busca, G., Cigada, A., Mazzoleni, P., Zappa, E., & Franzi, M. (2012). Cameras as displacement sensors to get the dynamic motion of a bridge: performance evaluation against traditional approaches. *Proceedings of the 6th International Conference on Bridge maintenance, Safety and Management (IABMAS)*, Italy.
- [14] Yarnold, M. T., Moon, F. L., Aktan, A. E., & Glisic, B. (2012). Structural monitoring of the Tacony-Palmyra Bridge using video and sensor integration for enhanced data interpretation. *Proceedings of the 6th International Conference on Bridge maintenance, Safety and Management (IABMAS)*, Italy.
- [15] Furlan, M. (2014). *Use of Digital Image Correlation for Steel section Strain Evaluation*. Master's Thesis. Supervisors: Dr. Christos Georgakis and Dr. Claudio Modena.

Biographies

NIRANJAN DESAI is an assistant professor of civil engineering at Purdue University Northwest. His areas of research interest include structural health monitoring and nonlinear dynamic analysis. In 2014, he was awarded the "Alan Yorkdale Memorial Award" for his journal paper titled "A study of the out-of-plane performance of brick veneer wall systems in medium rise buildings under seismic loads." Dr. Desai may be reached at desai39@pnw.edu

CHRISTOS GEORGAKIS is a full professor in the Department of Civil Engineering at the Technical University of Denmark and Head of the Civil Engineering Structural Dynamics Group, CESDyn. CESDyn is a research-oriented group working in the field of civil engineering structural dynamics. Research activities are mainly based on experimental work investigating the effect of environmental loading on structural response. Core areas are wind loading in general, additional effects due to rain, snow and ice accretion, seismic loading, and the impact of human-induced vibration. Christos Georgakis has published 16 journal articles in structural dynamics and vibration control and is co-author of the book *Cable Supported Bridges: Concept and Design*, in which he authored the chapter on monitoring of bridge structures. Dr. Georgakis may be reached at cg@byg.dtu.dk

GREGOR FISCHER is an associate professor in the Department of Civil Engineering at the Technical University of Denmark. Dr. Fischer may be reached at gf@byg.dtu.dk

INSTRUCTIONS FOR AUTHORS: MANUSCRIPT REQUIREMENTS

The INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH AND INNOVATION is an online/print publication designed for Engineering, Engineering Technology, and Industrial Technology professionals. All submissions to this journal, submission of manuscripts, peer-reviews of submitted documents, requested editing changes, notification of acceptance or rejection, and final publication of accepted manuscripts will be handled electronically. The only exception is the submission of separate high-quality image files that are too large to send electronically.

All manuscript submissions must be prepared in Microsoft Word (.doc or .docx) and contain all figures, images and/or pictures embedded where you want them and appropriately captioned. Also, for all accepted manuscripts, each figure, image, or picture that was imported into your Word document must be saved individually as a **300dpi or higher JPEG (.jpg)** file and submitted separately; your manuscript and figure numbers must be used in the title of the file (e.g., **J13-F-18 Figure 4**); that means one additional file for each image imported into your manuscript. These 300dpi images do NOT need to be embedded in your manuscript. For tables or graphs created directly in Word, you do not need to submit them as separate files.

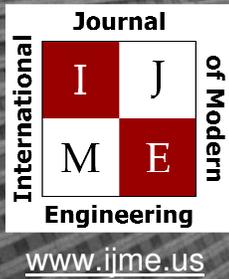
Included below is a summary of the formatting instructions. You should, however, review the [sample Word document](http://www.ijeri.org/formatting_guidelines) on our website (www.ijeri.org/formatting_guidelines) for details on how to correctly format your manuscript. The editorial staff reserves the right to edit and reformat any submitted document in order to meet publication standards of the journal.

The references included in the References section of your manuscript must follow APA-formatting guidelines. In order to help you, the sample Word document also includes numerous examples of how to format a variety of scenarios. Keep in mind that an incorrectly formatted manuscript will be returned to you, a delay that may cause it (if accepted) to be moved to a subsequent issue of the journal.

1. **Word Document Page Setup:** Two columns with 1/4" spacing between columns; Top of page = 3/4"; Bottom of page = 1" (from the top of the footer to bottom of page); Left margin = 3/4"; Right margin = 3/4".
2. **Paper Title:** Centered at the top of the first page with a 22-point Times New Roman (Bold), Small-Caps font.
3. **Page Breaks:** Do not use page breaks.
4. **Body Fonts:** Use 10-point Times New Roman (TNR) for body text throughout (1/8" paragraph indentation); 9-point TNR for author names/affiliations under the paper title; 16-point TNR for major section titles; 14-point TNR for minor section titles; 9-point TNR BOLD for caption titles for tables and figures; other font sizes as noted in the sample document.
5. **In-text Referencing:** List and number each reference when referring to them in the body of your document (e.g., [1]). The first entry must be [1] followed by [2], [3], etc., continuing in numerical order to the final entry in your References section. Again, see the sample Word document for specifics. Do not use the End-Page Reference utility in Microsoft Word. You must manually place references in the body of the text.
6. **Tables and Figures:** Center all tables and figures. Captions for tables must be above the table, while captions for figures are below; all captions are left-justified.
7. **Page Limit:** Manuscripts should not be more than 15 pages (single-spaced, 2-column format).
8. **Page Numbering:** Do not use page numbers.
9. **Publication Charges:** Manuscripts accepted for publication are subject to mandatory publication charges.
10. **Copyright Agreement:** A Copyright Transfer Form must be signed and submitted by all authors on a given paper before that paper will be published.
11. **Submissions:** All manuscripts and associated files must be submitted electronically.

MANUSCRIPTS should be submitted to Dr. Philip D. Weinsier, manuscript editor, at philipw@bgsu.edu along with a copy to editor@ijeri.org.

FILES containing your high-quality images should ONLY be submitted to philipw@bgsu.edu.



Print ISSN: 2157-8052
Online ISSN: 1930-6628



INTERNATIONAL JOURNAL OF MODERN ENGINEERING

ABOUT IJME:

- IJME was established in 2000 and is the first and official flagship journal of the International Association of Journal and Conferences (IAJC).
- IJME is a high-quality, independent journal steered by a distinguished board of directors and supported by an international review board representing many well-known universities, colleges and corporations in the U.S. and abroad.
- IJME has an impact factor of **3.00**, placing it among the top 100 engineering journals worldwide, and is the #1 visited engineering journal website (according to the National Science Digital Library).

OTHER IAJC JOURNALS:

- The International Journal of Engineering Research and Innovation (IJERI)
For more information visit www.ijeri.org
- The Technology Interface International Journal (TIIJ).
For more information visit www.tiij.org

IJME SUBMISSIONS:

- Manuscripts should be sent electronically to the manuscript editor, Dr. Philip Weinsier, at philipw@bgsu.edu.

For submission guidelines visit
www.ijme.us/submissions

TO JOIN THE REVIEW BOARD:

- Contact the chair of the International Review Board, Dr. Philip Weinsier, at philipw@bgsu.edu.

For more information visit
www.ijme.us/ijme_editorial.htm

INDEXING ORGANIZATIONS:

- IJME is indexed by numerous agencies.
For a complete listing, please visit us at www.ijme.us.

Contact us:

Mark Rajai, Ph.D.

Editor-in-Chief
California State University-Northridge
College of Engineering and Computer Science
Room: JD 4510
Northridge, CA 91330
Office: (818) 677-5003
Email: mrajai@csun.edu



www.tiij.org



www.ijeri.org

The International Journal of Engineering Research & Innovation (IJERI) is the second official journal of the International Association of Journals and Conferences (IAJC). IJERI is a highly-selective, peer-reviewed print journal which publishes top-level work from all areas of engineering research, innovation and entrepreneurship.



IJERI Contact Information

General questions or inquiries about sponsorship of the journal should be directed to:

Mark Rajai, Ph.D.

Founder and Editor-In-Chief

Office: (818) 677-5003

Email: editor@ijeri.org

Department of Manufacturing Systems Engineering & Management

California State University-Northridge

18111 Nordhoff St.

Room: JD3317

Northridge, CA 91330