# Graphics Interface 2014

Montreal, Quebec, Canada

May 7-9, 2014

# Poster Session Proceedings

Edited by Christopher Batty

*Sponsored by the Canadian Human-Computer Communications Society*

The abstracts in this volume comprise the proceedings of the meeting mentioned on the title page. They reflect the authors' opinions and appear here without change. Their inclusion in this document does not necessarily constitute an endorsement by the editors or the Canadian Human-Computer Communications Society.

# Contents

## Human Computer Interaction



## Computer Graphics

# Facebook use in Bhutan: A Comparative Study

**Foad Hamidi**
Department of Computer Science and Engineering
Lassonde School of Engineering, York University,
Toronto, Canada M3J 1P3
fhamidi@cse.yorku.ca

**Melanie Baljko**
Department of Computer Science and Engineering
Lassonde School of Engineering, York University,
Toronto, Canada M3J 1P3
mb@cse.yorku.ca

## ABSTRACT

Technology is adopted and used in novel ways in different contexts. We present results from a survey concerning the use of the Facebook social network in the Kingdom of Bhutan and compare the results with similar surveys conducted in the United States. The results uncover differences and similarities between the ways the technology is used in the two very different contexts. The comparison shows that Facebook users in Bhutan are more likely than their American counterparts, to browse the profiles of and add social network members that they have not met previously.

**Keywords:** Design, Human Factors.

**Index Terms:** H.5.m. [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

## 1 INTRODUCTION

Many theories exist on why certain technologies become popular so fast and "go viral". These dynamics become even more complex, and interesting, when differences between cultures in which the same technology is used affect the human-technology interaction as well. An examination of the novel ways technology is appropriated when designed within one culture and used in another can provide valuable insights into the dynamics of human-technology interaction. In this work, we examine some of the dynamics of the use of social networks, and specifically Facebook, in the novel context of Bhutanese society.

Facebook is currently the most popular website on the Internet [1] and has more than one billion users. Since its inception in 2004, it has been the subject of many studies examining the reasons behind its success [7, 8]. Among the factors at work, a need for a tool to facilitate the development and maintenance of social capital, especially in the face of globalization and increased mobility has been recognized [5]. Additional recognized factors have been the creation of personas, the easy sharing of information including recommendation of content, and creative expression [2, 6, 8].

Bhutan is an Asian Himalayan kingdom geographically located between India and China. It is unique in that as a society it has been careful to adopt new technologies: both television and the Internet arrived in Bhutan only in 1999 and Thimphu is the only capital in the world without a traffic light.

Bhutan is unique in its approach to technology: on the one hand, the Bhutanese society is wary of mindless adaption of new trends and, on the other, quick to adopt technology that proves to be useful. The survey presented was conducted on a college campus and despite the restrictions, more than 98% of the students who participated in the survey have Facebook membership and are active on the site.

Since opening its borders to the world in the 1950s, Bhutan has gone through several major transformations. There are many people who travel abroad to work and study and many young people move from small towns and villages to the growing capital of Thimphu. Previous research has shown that social networks help maintain contact with family and friends in the place of origin, as well as, facilitating the creation of new relationships in new environments such as a college campus [4]. Our hypothesis is that the differences between the use of Facebook in Bhutan and the United States reflect some of the characteristics of the context in which they are used.

## 2 A TALE OF TWO SURVAYS

Comparing technologies within different contexts, and at different times, is a daunting task. Changes especially in human-technology interaction, are very rapid and many technologies, adapt and change radically over time. Despite these concerns, we believed a good starting point was to conduct a survey similar to previous surveys conducted with a similar population (i.e., college students) in the United States. Thus, we based our survey on two studies conducted previously by Ellison, Steinfield and Lampe [4, 5]. Note that this work does not aim to emulate those studies in Bhutan; rather, we compare part of the reported results in those studies with information gathered in Bhutan to open a discussion about the differences and similarities between Facebook use in the two very different contexts.

In addition to general demographic questions such as age, gender and years in the program of study, we used the *Facebook Intensity* (FBI) scale [5], the "actual friends" variable and the connection strategies scale [4].

The students who participated in the survey, a total of 58 participants, were all pursuing a Bachelor of Computer Applications, which is a three-year program at the *Royal Thimphu College* (RTC), a popular private college located close to Bhutan's capital, Thimphu. All the students in this program participated in the survey. 20% of the students were in their first year, 43% in their second year and 27% in their third year. 75% were male and 25% female. The average age was 22.4 years. Despite the campus being far from the capital Thimphu (about 15 kilometers), or perhaps because of it, only 55.1% of the students live on campus.

Facebook use rates among the students are high. 98.2% of the students have a Facebook account and 94.8% use a laptop to access the Internet. The average time of Internet use per day was

reported as 4.4 hours. The reported Facebook use per week was 3.2 hours. There was a large variance (*SD* = 3.7 and *SD* = 3.9, respectively) for both of these values.

## 3 RESULTS

The results are presented in relation to previous data reported in two studies [4, 5]. The *Facebook Intensity Scale* (FBI) was 3.29 (*SD* = 0.23) which is significantly ($t_{506}$= 7.16, ns) higher than the value reported by Ellison et al., which was 3.06 (*SD* = 0.23) [5].

The number of reported Facebook friends was very high. 44.8% reported having more than 400 friends and the average number of total Facebook friends was 445.8 (*SD* = 257.6). Of these on average 35% were identified as *actual* friends. This number is higher than the 25% reported by Ellison et al. [4]. Further, of the total friends 15.6% were RTC students, 44.6% were from another town and 22.6% were from their hometown.

The *online to offline* measure, at 3.42 (*SD* = 0.96), was not significantly higher ($t_{509}$= 1.9, p>0.5) than previous results, which were 3.64 (*SD* = 0.79) [5]. However, the *offline to online* measure, at 3.19 (*SD* = 0.06), was significantly higher than previous results ($t_{506}$= 9.01, ns), which were 1.97 (*SD* = 1.03) [5].

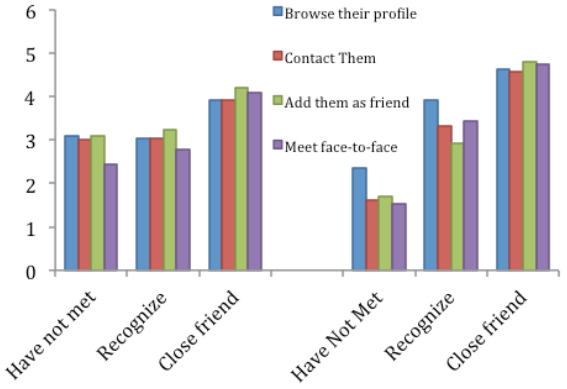



**Figure 1. Comparing the results of connecting strategy from [4] (shown on the right) and this study (shown on the left).**

The results of the connection strategy scale are shown in Figure 1 alongside their counterparts from the work of Ellison et al. [4].

There is a clear difference between the offline to online measure and other elements of the connection strategy. In particular, the results show that in Bhutan, Facebook users add other users that they do not necessarily know offline more frequently and are more likely to browse the profiles of and add people they have not met previously. Incidentally, this result confirms the first author's subjective experience while travelling in Bhutan, whereby he received many friend requests from people he had not met offline before. This was in contrast to his experience in North America, where he had socially met most of the people who had sent requests, before.

Another related result is the higher number of friends reported. While this could be due to the social network's maturation, it might also be an outcome of the difference in connection strategy.

## 4 DISCUSSION

Many difficulties arise when comparing the use of technology in such different contexts. First, there are differences in the population: while the American studies involved college students, they were chosen randomly and across disciplines. In our case, we only had access to students within a specific program. Second, social networks are dynamic themselves and comparing their use over time is very difficult [3]. For example, Facebook has shifted from a profile-centered system to a news-centered one and this affects user behavior on it. Finally, we faced challenges in the use of language and concepts when applying the survey in a new culture and had to discard some results, as some answers to the questions clearly did not correspond to our intended information.

These preliminary results show that there are clear differences between Facebook use in Bhutan and the US. Future surveys and interviews can explore this area further.

## 5 ACKNOWLEDGMENTS

We would like to thank Dr. Tshering Cigay Dorji, Sonam Choden and Susuhangma Rai at the Bhutan Innovation and Technology Centre, as well as, Bhagawan Nath, Somnath Chaudhuri and Sourav Basu at the Royal Thimphu College. We also would like to thank Dr. Nicole Ellison and Dr. Charles Steinfield.

## REFERENCES

[1] Alexa. 2013. http://www.alexa.com/topsites

[2] Brandtzæg, P. B., Lüders, M., & Skjetne, J. H. 2010. Too many Facebook "friends"? Content sharing and sociability versus the need for privacy in social network sites. Intl. Journal of Human–Computer Interaction, 26(11-12), 1006-1030.

[3] Ellison, N. and Boyd, D. M. 2013. Sociality through Social Network Sites. In The Oxford Handbook of Internet Studies (Ed. William H. Dutton), Oxford University Press, 151-172.

[4] Ellison, N. B., Steinfield, C., and Lampe, C. 2011. Connection strategies: Social capital implications of Facebook-enabled communication practices. *New Media & Society 13*, 6, 873-892.

[5] Ellison, N. B., Steinfield, C., and Lampe, C. 2007. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. Journal of Computer-Mediated Communication 12, 4, 1143-1168.

[6] Hamidi, F., and Baljko, M. 2012. Using social networks for multicultural creative collaboration. In *Proc. ICIC 2012*, ACM Press, 39-46.

[7] Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* 3 (Mar. 2003), 1289-1305.

[8] Lacy, S. 2009. The Stories of Facebook, YouTube and MySpace: the people, the hype and the deals behind the giants of Web 2.0. Crimson Publishing.

[9] Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., and Crowcroft, J. 2012. The personality of popular facebook users. In Proc. of CSCW'12, 955-964.

# Illimitable Space System Demo

Miao Song*
Concordia University

Serguei A. Mokhov†
Concordia University

Peter Grogono‡
Concordia University

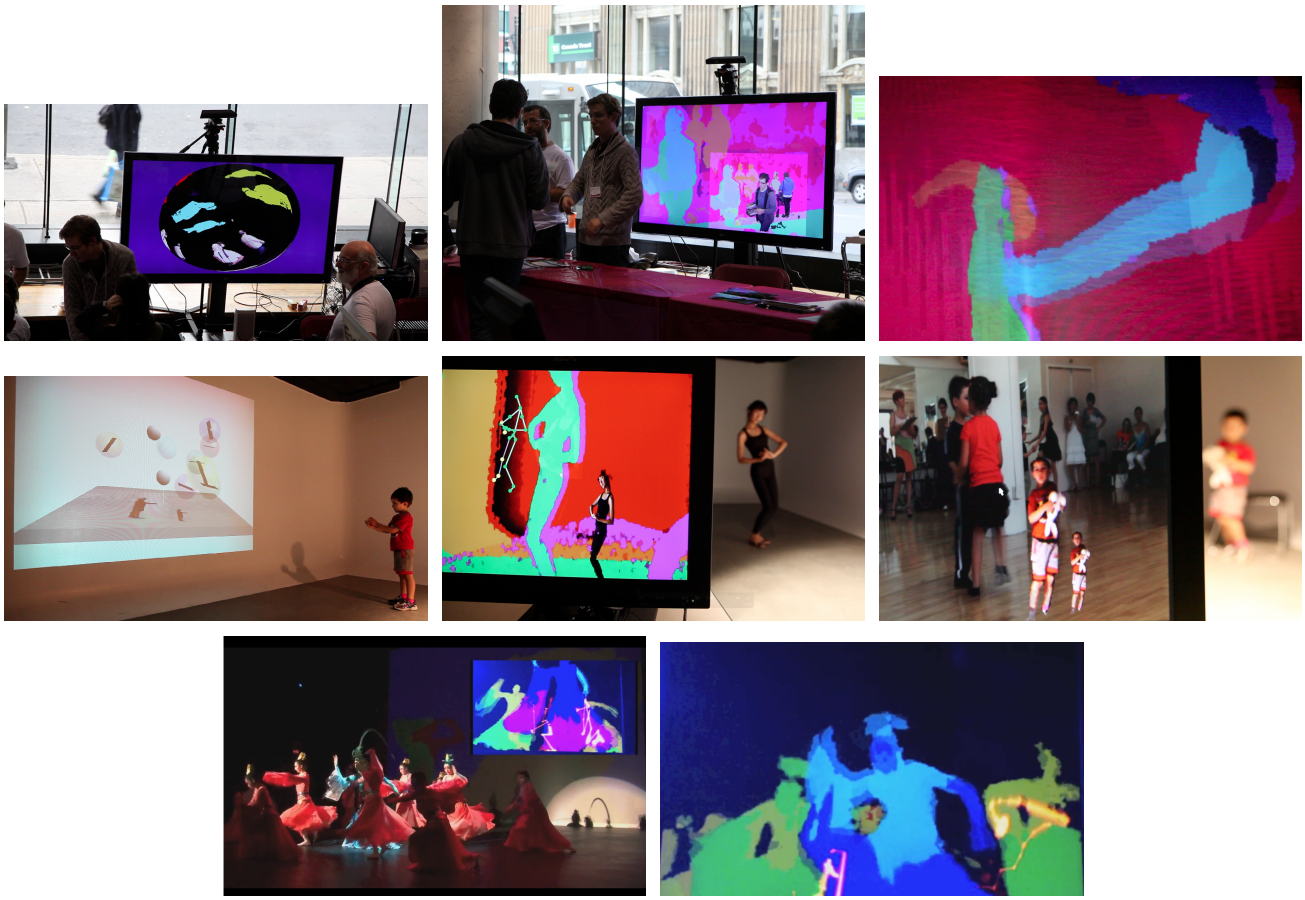


Figure 1: Various ISS Case Studies for Interactive Performance and Documentary Arts

## Abstract

In this work we showcase various multimodal interaction aspects of the research-creation *Illimitable Space System* (ISS) in terms of artistic performance (e.g., dance), music visualization, and interactive documentary controlled with gestures and voice depending on the mode chosen in realtime.

**Index Terms:** H.5.1 [Multimedia Information Systems]; H.5.2 [User Interfaces]: Input devices and strategies, Screen Design, Voice I/O; H.5.5 [Sound and Music Computing]: Signal analysis, synthesis, and processing;

*e-mail: m_song@cse.concordia.ca
†e-mail: mokhov@cse.concordia.ca
‡e-mail: grogono@cse.concordia.ca

## 1 ISS Overview

We demonstrate various multimodal interaction aspects of the *Illimitable Space System* (ISS) as a configurable artist's toolbox in terms of artistic performance (such as dance or theatre production), music visualization, and interactive documentary controlled with gestures and voice. Some of the system's applications were publicly exhibited in various locations.

The system's conceptual design is depicted in Figure 2. The participants, in some cases are artists and/or the audience, are central to interact with the installation providing motion, voice, and other captured data as an input to the system. The ISS processes the supplied input and produces a real-time graphical and audio feedback to the interacting audience [**?**] and the process repeats.

The logical pipeline of the performance installation presented in Figure 2, combines computer graphics and simulation, documentary montage, interactive media, and theatre performance into one research-creation piece, which transitions through the four primary states illustrated. The states are adapted to the need of each specific performance or installation.

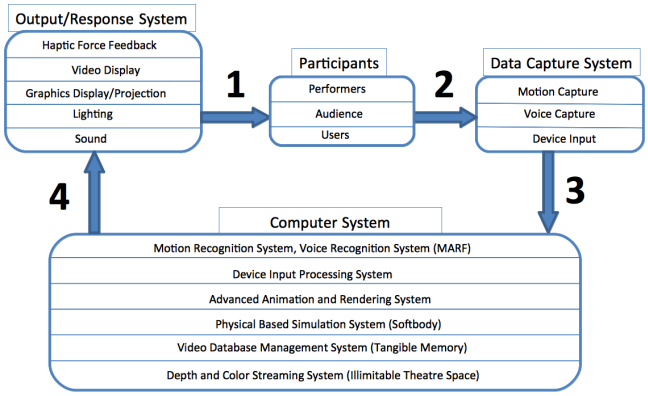



Figure 2: Conceptual Design of an ISS-Based Installation

- State 1: In either a makeshift initial theatrical space, such as a blackbox, or a classical theatre stage, where lighting, projection, sound props, and related facilities are properly set up for actors' performance, audience's participation, and other user group's interaction devices.

- State 2: Participants, including actors, audience, and other users may freely move in the designated physical space. They could move their body in different postures with different gestures, talk or sing, and could play musical instruments, or touch haptic-enabled devices. The state of human's performance, such as motion of actors' body movements, audience's voices and the forces of users applied through the haptic devices, are captured through the integrated Data Capture System to generate a response.

- State 3: The captured data from the theatrical space are transferred to the ISS's PACS (Performance Assistant Computer System) to be analyzed. The ISS mainly contains three big components: recognition, computing, and media output. The motion recognition system, voice recognition system, and device processing system are filtering, parsing, and sorting motion, audio, and device/sensor input streams captured into the system by various capture devices (such as Kinect). The ISS then computes and simulates physically based computer graphics and dynamic audio based on the mathematical and physically based algorithms, and fetch proper video footage from a video database management system following some querying algorithm.

- State 4: The outputting system then re-renders the newly computed computer generated imagery onto projectors, plays generated audio signals through speakers, and even controls dynamic lighting (interacted with originally as a result of participants' actions). Optionally, the outputting system then reconfigures the theatrical technologies afterward based on the computation results and state updates.

While the audience can be anyone, depending on the mode, the number of people who can *effectively* interact at the same time varies as follows: documentary: one audience member at a time (while multiple observers are possibles, the gestures and voice are accepted only from one participant); dance and theatre performance: depth visualization has no such limitations, but green-screening and avateering in virtual audience or with depth are limited to recognition of up-to six people (Kinect's limitation). What follows are sample interaction scenarios present in the linked videos.

### 1.1 Ascension

The last two images in Figure 1 are from the most recent (January 18–19, 2014) Chinese New Year Gala performance where the ISS was deployed during the actual dance called *Ascension* as a dynamic background based on motion tracking and particle systems capturing the dancers in real-time. We have learned a very large number of lessons and experience from this non-lab non-controlled large scale public deployment [**?**]. Prior to *Ascension* the system was deployed in various configurations at the Open House and Stewart Hall Exposcience events in 2012 and 2013 in Montreal, Concordia University. See the related videos at: `http://vimeo.com/channels/153466`.

### 1.2 Augmented Reality Audience

Revisiting own competition footage as a augmented reality audience and re-recording it again seems like a lot of fun some. This PoC recorded in a Concordia Hexagram production room replays pre-recorded dance competition footage with the real-time green-screening and augmenting video frames of a child from his own competition (`http://vimeo.com/50069419`).

### 1.3 Interactive Documentary: Tangible Memory Bubbles

Audience interacting with Projected *Tangible Memories* (based on the linear *I Still Remember* short documentary film). The Tangible Memories Bubbles project was made into installation in the the same video production room. The audience are trying to play with the projected memory bubbles from the interactive documentary's Tangible Memories by using gestures and voice to call out different color bubbles that contain various video footage in them.

### 1.4 Multimodal Musical Performance: The Depth of a Sleeve

Water-sleeve Chinese dance with music visualization blended together is an earlier example of the real-time visualization of the rendered depth in rainbow-like colors. When the dancer and her long water sleeves are at different distance from the Kinect device, the depth stream of the dancer is rendered in gradient shades. Additional, the rendered graphics is affected by various music beats in real-time.

## 2 Conclusion

We briefly illustrated various interaction modes of the ISS. The currently undergoing or future work covers the use of the ISS in some scenes of the larger-scale theatre production *Like Shadows* stages by the Central Academy of Drama, Beijing; digital forensic evidence management by extending the interactive documentary component, and others.

## Acknowledgements

This work was supported in part by the Department of Computer Science and Software Engineering, Faculty of Engineering and Computer Science, Hexagram Concordia, District 3, Concordia University and FQRSC.

## References

[1] M. Song. *Computer-Assisted Interactive Documentary and Performance Arts in Illimitable Space*. PhD thesis, Concordia University, 2012.

[2] M. Song and S. A. Mokhov. Dynamic motion-based background visualization for the Ascension dance with the ISS. [dance show, video], Jan. 2014. `http://vimeo.com/85049604`.

# QualiWand: Towards Optimising Feedback for Motion Capture System Calibration

Zlatko Franjcic*
Qualisys AB / Chalmers University of Technology

Paweł Woźniak
t2i Interaction Lab, Chalmers Univeristy of Technology.

**ABSTRACT**

This work in progress report presents preliminary results on the design of a feedback device that supports the task of calibrating a motion capture system. Calibrating motion capture systems is needed for their proper operation. It requires unique, dynamic actions and demands spatial awareness from the user. The goal of this inquiry is to find ways of providing feedback that will guide the user to perform the task with maximum efficiency so that the calibrated volume is optimal. This paper contains a description of the task and our initial design of a feedback system. We introduce QualiWand — and augmented calibration device that will enable us to study which feedback form is most optimal for the task. We then propose a plan for studying sound, visual and tactile feedback.

**Index Terms:** H.5.m [Information Interfaces and Presentation (e.g., HCI)]: Miscellaneous

## 1 INTRODUCTION AND RELATED WORK

Calibration methods for multi-view camera systems used for 3D computer vision often require the user to perform certain tasks in the physical world. Those tasks aim at setting up scenes with known geometry that can ideally be viewed in their entirety by all the cameras simultaneously with as little obstruction as possible. The images so taken are used to determine the parameters of a given projection model describing how a 3D point in the scene is mapped to 2D point on a camera's image plane [6].

The usual approach to "creating" scenes with known geometry is to introduce a physical reference object, sometimes called *calibration object*, into the measurement space. As Pribanić et al. [5] note, many different types of objects have been proposed over the course of years in an attempt to reduce manufacturing and other costs associated with the calibration object itself and lower the effort and expertise required from the user to perform a calibration. But even if the particular choice of calibration object enables users with little or no training to perform the steps outlined by the calibration procedure, is not always guaranteed that the user's actions will result in acquisition of image data that is suitable for the intended future measurement and 3D reconstruction. It is thus important that users understand which data is useful and which not with respect to the planned setup. Optimal solutions to that issue remain to be found.

Providing real-time feedback from the system during calibration task execution can guide the user by supplying information on past actions and directing the user to perform certain actions to obtain suitable calibration data for a given measurement. With this approach, non-expert users can perform a calibration with a reduced risk of having to repeat the procedure due to unsuitable data resulting from the user's actions.

In this work we examine the interactive aspects of a custom-designed calibration procedure based primarily on a one-dimensional calibration object, and used for a commercial multi-view motion capture system [3]. In particular, we study what type of feedback is appropriate for the custom-designed calibration procedure. Our aim is to investigate the feedback mechanism's effectiveness in reducing mistakes, improving the fraction of "good data" provided by the user and decreasing task execution time. Important aspects to consider in the feedback design are 3D perception augmentation and reduction of cognitive load. In a broader sense, we seek to investigate methods to design and evaluate feedback forms for complex dynamic tasks where the user constantly changes position and orientation.

Several past research efforts focused on augmenting the user experience of calibration tasks in different contexts. Flatla et al. [2] attempted making calibration tasks more pleasurable by introducing gamification. They designed games to determine color perceptibility, set optimal C:D ratios for input devices and measure the input range for a physiological sensor. Pfeuffer et al. [4] introduced a new gaze calibration procedure where they introduced the concept of blending the calibration into target applications. Our research can relate to these works as we are also aiming to make the calibration a more pleasurable experience. The work presented in this paper goes beyond the aforementioned inquires as it aims to provide a "sixth sense" not only augmenting the user experience of the calibration process, but also rendering the process more accurate.

## 2 CALIBRATION — AN INTERACTIVE TASK

We study the task of calibrating an optical, marker-based motion capture system designed by Qualisys AB[1]. A typical setup consists of four to up to a few dozens of cameras. Each camera can take up to 500 images per second of near-infra-red light reflected by retro-reflective near-circular markers. The calibration process of the Qualisys AB motion capture system makes use of a one-dimensional calibration object: two markers are mounted on a rigid bar [1], or wand, whereby the distance between the markers is known to a very high accuracy. Since images of two points alone are not sufficient for calibration [7], a second reference object is used. In particular, an L-shaped frame with four markers mounted to it is placed as a static reference object in the measurement space, whereby the distance between any two neighboring markers is again known with very high precision. The combination of measured markers from both the static and the moving object serves as the input to an algorithm that determines the best parameters, in terms of root-mean-square error, for the projection model by solving the so-called bundle adjustment problem [3].

The heuristic guideline for the user moving the wand through space is simple: the wand should be moved throughout the entire volume where 3D measurements will be made, and additionally the wand should be rotated uniformly in the entire volume. These simple instructions aim at reducing overfitting and bias towards a particular position and orientation of the line with respect to any of the cameras' optical axis. Figure 1 presents the calibration wand, the L-frame and a visualisation of the calibrated motion tracking volume.

While the calibration procedure and the heuristic guideline are easily explained to novices, the fact that there is currently no feed-

*e-mail: zlatko.franjcic@qualisys.com

[1] http://www.qualisys.com

(a) The wand (b) The L-frame (c) The calibrated volume

Figure 1: The calibration wand (a) features two reflexive markers placed on a T-shaped rod, while four static reference points are arranged on an L-shaped frame (b). As the distance between the markers is predefined, the system can be calibrated by analysing the view of the wand and the L-frame from multiple cameras. This results in a finite volume (c) where the system can reliably provide positional information.

back at all during actual task execution, leaves the user with no other choice then trial-and-error to gain sufficient experience to get "a feel" for what kind of wand movements lead to a satisfactory calibration. This situation can clearly be improved, as there are objective measures that can be used to give feedback in different possible ways on the quality of the calibration at any point during the calibration process itself. The quality can be measured using objective measures such as calibration residuals. Furthermore, the calibration can be assessed with respect to a volume defined by the user, that is, the volume within which the user intends to conduct motion measurements. Quality measures can be obtained on-the-fly thus creating a possibility to inform the user on the state of the calibration during the task.

## 3 Design

Designing for performing a complex and unique task is a challenge. Our initial design inquiry into the nature of the task and the context of the task resulted in identifying several feedback forms as possible solutions.

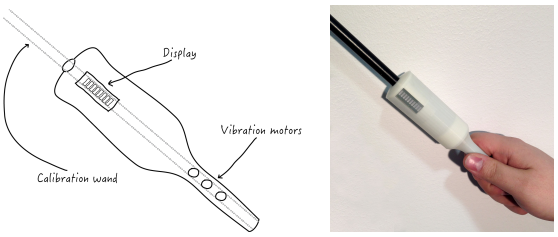
**Sound**. Auditory feedback may be useful as it does not require any additional infrastructure and it can easily accommodate multiple instructions. The drawback of sound is the inability to provide directional cues without extensive infrastructure such as directional speakers.

**Visual**. Feedback can be provided through a display informing the user on the quality of the calibration. Advantages of using visual feedback include the possibility of using multiple representations (e.g. colour, graphs, numbers) to convey information and the relative ease of implementation. On the other hand, it may be challenging to design a display that is easily visible throughout the entire task as the user changes their position rapidly.

**Tactile**. Using tactile feedback (e.g. vibration motors) is a promising opportunity. This feedback form can be easily embedded in the calibration wand and it can be provided with equal quality regardless of the position of the user. However, tactile feedback may be ambiguous and its precise design may require a significant effort.

We rejected several other feedback possibilities. Notably, we dismissed the possibility of using augmented reality. While visualising the space to be calibrated in augmented reality seems like a tempting idea, this solution is implementation-heavy and raises the fundamental problem of establishing the position of the augmented reality device when the system is not calibrated and cannot provide positional information.

For an initial inquiry, we constructed a prototype augmented calibration wand device.. The device is to be slid onto the calibration wand. The prototype features two types of visual displays (a bar graph and a numerical display) and a set of vibration motors built into the handle of the wand. The components are integrated into a 3D-printed enclosure. Figure 2 provides an overview of the components of the device as well as evidence of the implementation.




(a) A design sketch for QualiWand (b) Prototype implementation

Figure 2: QualiWand is additional device designed to be slid onto the calibration wand. In our augmented version, the user holds the wand through QualiWand (b). This enables several design possibilities such as providing two feedback types — visual and sounds (a).

## 4 Study design

We are planning to conduct a user experiment that will hope us determine optimal feedback forms. We will compare how users perform the calibration task with or without the help of extra feedback. We will compare four conditions: no feedback (baseline), audio, visual and tactile. Participants will perform system calibration within a predefined time in all four conditions (a within-group study) and Latin squares will be used to minimise the role of task order. Performance will be measured in terms of quantity (percentage of volume calibrated) and quality (calibration residuals). We will then run ANOVA to look for significant effects.

## 5 Conclusions

In this work in progress report, we presented our initial insights into the design of a feedback system supporting the motion tracking system calibration task. We provided a detailed description of the task and outlined different design possibilities for augmenting the task. This report also discusses our initial prototype and an experiment plan. We hope to continue the research effort, perfom our planned studies and reach conclusions on what the optimal feedback form is.

## Acknowledgements

Zlatko Franjcic is an Early Stage Researcher in the ACT Marie Skłodowska-Curie ITN. Paweł Woźniak is an Early Stage Researcher in the DIVA Marie Skłodowska-Curie ITN (REA grant agreement nos. 289404 and 290227).

## References

[1] N. Alberto Borghese and P. Cerveri. Calibrating a video camera pair with a rigid bar. *Pattern Recognition*, 33(1):81–95, 2000.

[2] D. R. Flatla, C. Gutwin, L. E. Nacke, S. Bateman, and R. L. Mandryk. Calibration games: Making calibration tasks enjoyable by adding motivating game elements. In *Proceedings of UIST '11*, pages 403–412. ACM, 2011.

[3] S. Hofverberg. Theories for 3d reconstruction. Technical Report QMARK-TECH-1004, Qualisys AB, 2007.

[4] K. Pfeuffer, M. Vidal, J. Turner, A. Bulling, and H. Gellersen. Pursuit calibration: Making gaze calibration less tedious and more flexible. In *Proceedings of UIST '13*, pages 261–270. ACM, 2013.

[5] T. Pribanić, P. Sturm, and M. Cifrek. Calibration of 3d kinematic systems using orthogonality constraints. *Machine Vision and Applications*, 18(6):367–381, 2007.

[6] G.-Q. Wei and S. De Ma. Implicit and explicit camera calibration: Theory and experiments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(5):469–480, 1994.

[7] Z. Zhang. Camera calibration with one-dimensional objects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(7):892–899, 2004.

# Pen-based Error Detection with Supervised Machine Learning Algorithms

Afroza Sultana*
School of Information Studies, McGill University
Montreal, QC, Canada

Karyn Moffatt†
School of Information Studies, McGill University
Montreal, QC, Canada

**ABSTRACT**

Although interaction techniques have been extensively studied under controlled laboratory conditions, little is known about their capabilities during unconstrained free tasks. Understanding real world use is particularly important for older adults, as some may find computer input more challenging due to age-related declines in motor skill. As a step towards addressing this gap, we studied the feasibility of using machine-learning techniques to model in-situ interactions. Using pen-based data from older adults, we tested four supervised machine-learning algorithms: Decision Tree, Neural Network, Naïve Bayesian Network, and Rule Induction. Each yielded an accuracy rate above 90%, and true positive rates up to 64% for identifying errors from selection tasks. These findings indicate the potentials of accurately modeling real life unconstrained touch-based interaction data, using machine-learning algorithms.

**Keywords**: Target selection, pointing devices, error modeling, sub-movement analysis.

**Index Terms**: H.5.2. [Information Interfaces and Presentation]: User interfaces—*evaluation / methodology*.

## 1 INTRODUCTION

Many older adults (who are more than 65 years old) experience an age-related loss of motor function that can lead to frequent selection errors while aiming a target (e.g. a button or a link) on touch-screen. Conducting such errors may trigger frustration among these users [1], [3]. Understanding older adult users' input behavior can be helpful for designing accessible interfaces for them. Most work (see [16] for a review) targeted at understanding pointing behavior (both with older adults specifically, and users more generally) have relied on laboratory experiments that may not capture the full complexity of selection behavior [10]. Moreover, performance of the same user may vary significantly across different task sessions; hence data collected in a lab session may not reflect variation of the real word data [8].

Even though in-situ data offers the potential for greater understanding of selection behavior, studies conducted "in the wild", must grabble with the challenges of inferring the user's intent and teasing out the influence of external variables [2]. Furthermore, as the unconstrained user interaction data is collected over months, the multidimensional contents of such enormous datasets pose a challenge towards manual inspection of the performance measures. In recent years, some efforts have targeted to automatically model the in-situ pointing device input data. The Input Observer application analyzes the in-situ mouse pointing performance that can distinguish intentional and unintentional errors [4]. However, their system relies on human coders to identify errors that can be only visually identified from a still image. In another study, an unsupervised machine-learning algorithm was applied to model unobtrusive mouse input data [5]. However, this study focused on developing classifiers to identifying intentional and unintentional sub-movements. Their system does not attempt to identify errors.

It is clear from the existing pointing device performance analysis literature that there have been studies on analyzing in-situ user interactions with mouse, but no such studies were conducted on touch-based interaction devices. This research gap in the literature motivated us to explore unconstrained touch-based user interactions (particularly, from older adults), and apply machine-learning algorithms to model them. The work presented here represents a first step in this direction.

As mentioned before, the unconstrained real life user interaction data is multi-dimensional, and incorporated with the challenges of identifying unintended selection errors from the intentional selection errors. Therefore, before directly applying machine-learning algorithms on the unconstrained touch-based user interaction data, we intended to conduct a feasibility study on touch-based input data from older adult users that were collected in controlled lab experiments. The study reported here is the initial feasibility study, where we analyzed pen-based selection task data from controlled lab experiments, and applied the following four supervised machine-learning algorithms [6]: Decision Tree, Neural Network, Naïve Bayesian Network, and Rule Induction to model them. The objective behind choosing four different supervised algorithms was to identify an algorithm, if any, that best fits to model the touch-based input data.

Studies on touch-based error detection from the controlled lab experiment data emphasize only on selection errors and task completion time as error detection metrics [7], [15]. However, studies on the mouse movements of motor impaired users suggest that the sub-movement analysis of the cursor trajectories provides precise performance measures for error detection for such users [9], [17]. In our data modeling we also considered to use the sub-movement error detection measures, discussed in [9], [11], and [12]. Our experiment results showed that all of these supervised machine-learning algorithms successfully identified selection task errors generated by older adult users with above 90% accuracy rate, and at most 64.4% true positive rates.

## 2 EXPERIMENTAL METHODOLOGIES

For this initial study, we used an existing dataset from prior work [13]. This dataset contains target selection task data from 12 younger adults (19–29 years), and 12 older adults (65–86 years). Full details of the data collection procedures are available in [13].

During data analysis, we calculated the sub-movement analysis metrics that are described in [9], [11], [12], namely, movement direction change, orthogonal direction change, target axis crossing, sub-movement error (i.e., distance from the axis),

* afroza.sultana@mail.mcgill.ca
† karyn.moffatt@mcgill.ca

and target re-entry, number of sub-movements, number of pauses taken, and task completion time, from the x and y coordinates of each recorded sub-movements.

We designed our experiments to generate three sets of models from: data obtained from the older adult users only, data obtained from the younger adult users only, and data obtained from both older and younger adult users. We applied the RapidMiner implementation [14] of Decision Tree, Neural Network, Naïve Bayesian Network, and Rule Induction algorithms for our data modeling. For model construction, we provided two-third of the data from each dataset as the training samples and rest one-third of the data from the same datasets as the testing samples for all supervised machine-learning algorithms.

## 3 Results and Discussion

After constructing error prediction models from all four algorithms, we ran 3-x validation tests on RapidMiner to measure the correctness of our models.

For space constrains, here we only report the results from the older adult user data on all four algorithms (see Figure 1). All of our models demonstrated above 90% accuracy on error detection. However, the true positive rates (correct prediction of unsuccessful attempts, among all unsuccessful attempts) varied from 52.53%-64.40%, and the false positive rate (incorrect prediction of unsuccessful attempts, among all predicted unsuccessful attempts) varied from 10.60%-32.23%. Moreover, true negative rates (correct prediction of successful attempts, among all successful attempts) were above 95%, and false negative rates (incorrect prediction of successful attempts, among all predicted successful attempts) were below 7%, for all models. The running time for Decision tree, Naïve Bayesian network, and Rule induction algorithms were reported between 1-2 seconds. However, the Neural Network algorithm required 31-61 seconds to build each model. From Figure 1, it can be concluded that Naïve Bayesian Network and Rule Induction models outperform other algorithms in error prediction.

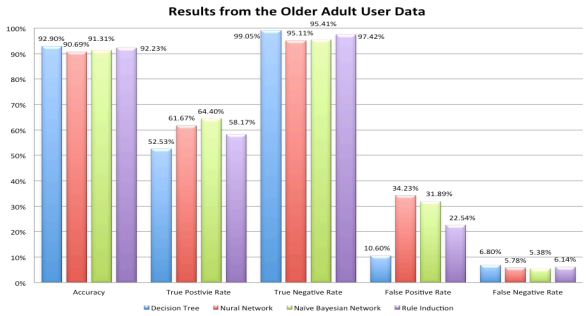



Figure 1: Results from the older adult user dataset.

## 4 Conclusions and Future Works

In this paper, we presented a feasibility study on applying supervised machine learning algorithms for modeling in-situ touch-based selection task data. Applying the selection task difficulty metrics and sub-movement distortion metrics, we observed that Naïve Bayesian network and Rule induction algorithms are the best-suited algorithms for modeling selection task errors among older adults.

Our findings indicate that machine-learning algorithms are a potential approach for modeling touch-based interaction errors, especially when real world data is collected from a large number of users. As a preliminary study, we used labeled data from a controlled study. Our future work will focus on extending our approach to in-situ tasks in uncontrolled environments.

We additionally plan to investigate how this approach can be extended to multi-touch selections. Given the prevalence of multi-touch devices, they are an ideal platform for this type of investigation. We will also focus on identifying alternative performance metrics to measure multi-touch movements.

## References


[1] K. S. Birdi and D. Zapf. Age differences in reactions to errors in computer-based work. In *Behaviour & Information Technology*, vol. 16 (6), p. 309-319, 1997.

[2] B. Brown, S. Reeves and S. Sherwood. Into the wild: challenges and opportunities for field trial methods. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, p. 1657-1666, 2011.

[3] S. J. Czaja and J. Sharit. Age differences in attitudes toward computers. In *The Jour. of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 53(5), p. 329-340, 1998.

[4] A. Evans and J. Wobbrock. Taming wild behavior: the input observer for text entry and mouse pointing measures from everyday computer use. In *Proc. of the 2012 ACM annual conf. on Human Factors in Computing Systems*, p. 1947-1956, 2012.

[5] K. Gajos, K. Reinecke and C. Herrmann. Accurate measurements of pointing performance from in situ observations. In *Proc. of the 2012 ACM annual conf. on Human Factors in Computing Systems*, p. 3157-3166, 2012.

[6] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*, 2nd edition, Elsevier, San Francisco, 2006.

[7] J. P. Hourcade and T. R. Berkel. Simple pen interaction performance of young and older adults using handheld computers. In *Interacting with Computers*, vol. 20(1), p. 166-183, 2008.

[8] A. Hurst, J. Mankoff and S. E. Hudson. Understanding pointing problems in real world computing environments. In *Proc. of the 10th international ACM SIGACCESS conf. on Computers and accessibility*, p. 43-50, 2008.

[9] F. Hwang, S. Keates, P. Langdon and P. J. Clarkson. A submovement analysis of cursor trajectories. In *Behaviour & Information Technology*, vol. 24(3), p. 205-217, 2005.

[10] A. Jansen, L. Findlater and J. O. Wobbrock. From the lab to the world: Lessons from extending a pointing technique for real-world use. In *Proc. of CHI'11 Extended Abstracts on Human Factors in Computing Systems*, p. 1867-1872, 2011.

[11] I. S. MacKenzie. Motor behaviour models for human-computer interaction. In *HCI models, theories, and frameworks: Toward a multidisciplinary science*, p. 27-54, 2003.

[12] I. S. MacKenzie, T. Kauppinen and M. Silfverberg. Accuracy measures for evaluating computer pointing devices. In *Proc. of the SIGCHI conf. on Human factors in computing systems*, p. 9-16, 2001.

[13] K. Moffatt and J. McGrenere. Steadied-Bubbles: Combining techniques to address pen-based errors for younger and older adults. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, p. 1125-1134, 2010.

[14] RapidMiner. www.rapidminer.com.

[15] X. Ren and S. Moriya. Improving selection performance on pen-based systems: A study of pen-based interaction for selection tasks. In *ACM Trans. on Computer-Human Interaction*, vol. 7(3), p. 384-416, 2000.

[16] R. W. Soukoreff and I. S. MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. In *International Journ. of Human-Computer Studies*, vol. 61(6), pages 751-789, 2004.

[17] J. O. Wobbrock and K. Z. Gajos. Goal crossing with mice and trackballs for people with motor impairments: Performance, submovements, and design directions. In *ACM Transactions on Accessible Computing*, 1(1), 4:1–4:37, 2008.

# Robot Arm Manipulation Using Depth Cameras and Inverse Kinematics

**Akhilesh Mishra**[1] **Dr. Oscar Meruvia-Pastor**[2]
**Department of Computer Science**
**Memorial University of Newfoundland.**

## 1 INTRODUCTION

An alternative approach to manipulate a robotic arm, using a depth camera to capture user input and inverse kinematics to define the joint motion of the robotic arm is presented. Existing manipulation techniques for a robotic arm use joystick or a controller/actuator designed specifically to manipulate the arm. The problem with the joystick or a controller is that they need a lot of training and the user needs to manipulate each joint to reach the target and this is typically done with Forward Kinematics. In addition, actuators are relatively expensive devices. This poster presents a new method in which the user just needs to point towards the target and the robotic arm will reach the target itself using the inverse kinematics algorithms. The advantage of this approach is that manipulation of the arm needs less training and is easy to learn. Also, simple speech commands were added to open and close the end-effector, which allows to pick and drop objects with the goal of making the robotic arm controller intuitive and easy to work with. We test our approach with a robot arm simulator similar to those used in commercial applications.

Key words: *Depth Cameras, Inverse Kinematics, CCD, Speech Commands, Gesture Recognition*.

### 1.1 MOTIVATION

Current robot arm simulators such as GRI Simulations Inc.'s Titan IV arm simulator [7] mainly work with Forward Kinematics, in which the user has to control a robotic arm using a joystick or a master controller. These controllers are generally expensive and require an ample amount of training before the user could perform the tasks efficiently using the simulator. With the presented approach the user just needs to point towards the target and the arm movement will be handled by the algorithm. Once the end-effector has reached a target position the user can manipulate the end-effector using the speech or gesture commands.

## 2 SYSTEM OVERVIEW

The block diagram of the overall system is shown in Figure 1. The user specifies the target position by moving his/her hands in front of a depth sensing camera. The depth camera returns the coordinates of the users' hand and the coordinates are passed as target position to the inverse kinematics module, where the joint angles for the arm simulator are calculated.

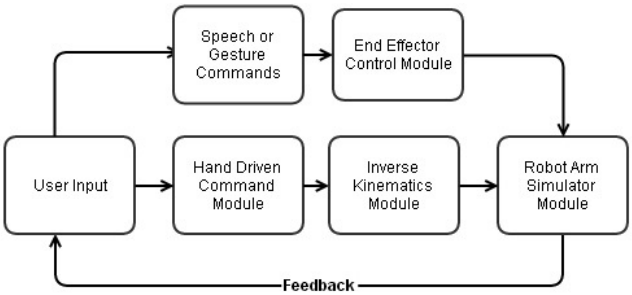



**Figure 1:** *System Overview*

After the joint angles have been calculated the robot arm simulator module applies the rotations calculated and the end-effector reaches the target. Once the target position has been reached the user can issue a voice or a gesture command to interact with the end-effector control module to pick and drop the objects.

### 2.1 HAND-DRIVEN COMMANDS USING DEPTH CAMERAS

Intel's Creative(tm) depth camera was used to get the user information. The camera SDK [8] returns some information about the user, such as hand position and its distance from the camera. The hand position is used to control a target ball which can move in 3D input space. The input space is a frustum in which the central position is the origin, where the robot is placed. As shown in Figure 2b, the camera field of view is mapped to a spherical region around the robot, whose radius is equal to the maximum length the robotic arm can stretch. The user can move the target ball within the sphere in X, Y, Z directions. Figure 2a shows how the ball moves into the scene with respect to the

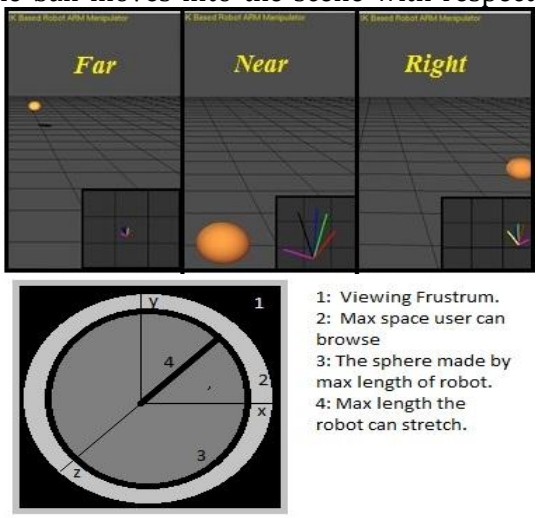



**Figure 2a**: (top) *Target ball positions w.r.t the user hand*, **2b)** *(bottom) Camera FOV mapping to the scene.*

[1]akm565-at-mun.ca

[2]oscar-at-mun.ca Assistant Professor, Department of Computer Science & Office of the Dean, Faculty of Science

user's hand. The 3D coordinates of the ball are passed as input to the IK module.

### 2.2 INVERSE KINEMATICS

Inverse Kinematics is a technique in which the user specifies the target position and the joint angles are computed by the algorithm, as shown in Figure 3a the user specifies the target position (x, y) and the joint angles $\theta_1$,$\theta_2$and $\theta_3$ *are* calculated using the algorithm.

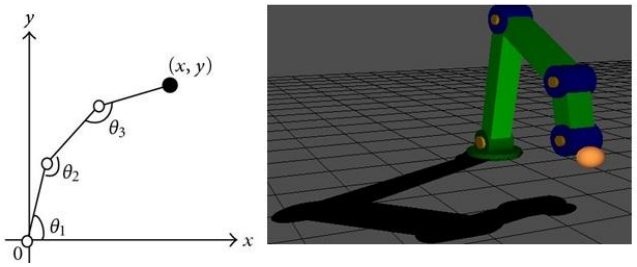



**Figure 3** *a) 3 Link Manipulator* **b**) *Robot Arm simulator.*

There are several algorithms for solving IK, coming originally from robotics applications. The most popular ones include *Cyclic Coordinate Descent* methods [1, 2], *Pseudoinverse* methods [3], *Jacobian Transpose* methods [5, 4] and *Triangulation* method [6]. Figure 3b shows the robotic arm simulator used for this research, developed in OpenGL. Cyclic Coordinate Descent algorithm was used to solve the IK problem, because of its simplicity and computational efficiency.

#### 2.2.1 CYCLIC COORDINATE DESCENT

CCD solves the IK problem through optimization. Looping through the joints from end to root, each joint gets optimized so as to get the end effector as close to the target as possible [1] [2].

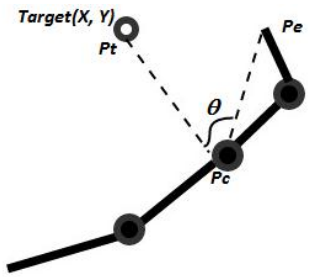



**Figure 4:** *CCD algorithm, the end-effector rotates to make* $\theta$ *zero*.

The Basic outline of the algorithm is as follows. The algorithm starts by measuring the difference between the two vectors formed between the effector position $P_e$ to $P_c$ and from $P_c$ to target position $P_t$. It then calculates the rotation and direction to reduce this difference to zero (see Figure 4). It does this for each joint, iterating from the end-effector to the root joint of the kinematic chain. The rotation is calculated by the dot product of two vectors and the direction is calculated by the cross product of two vectors [1]. To reach the target the equations (1) and (2) shown below are solved for each joint until the difference between the end-effector and target is zero or the number of iterations has reached its limit.

$$\cos(\theta) = \frac{p_e - p_c}{||p_e - p_c||} \bullet \frac{p_t - p_c}{||p_t - p_c||} \quad (1)$$

$$\overrightarrow{r} = \frac{p_e - p_c}{||p_e - p_c||} \times \frac{p_t - p_c}{||p_t - p_c||} \quad (2)$$

### 2.3 END-EFFECTOR CONTROL MODULE

This module uses the Intel SDK's[8] speech processing and image processing APIs to let the user interact with the end-effector using speech commands like "Pick", "Drop" for picking and dropping the objects from end-effector. The user can also use the "Thumbs up" or "Thumbs down" gestures for pick and drop commands.

## 3 EVALUATION

One of the goals of this research is to provide a mobile solution for manipulating the robotic arm, for that the camera will be mounted on the users' chest or head (also called first person manipulation) and the performance of the solution will be evaluated by user studies. Along with the first-person manipulation, presented method will also be tested in a third person perspective in which the camera is in front of the user. User studies will be conducted to compare the advantages and disadvantages of both approaches.

## 4 ACKNOWLEDGEMENTS

Research funded by the Research & Development Corporation (RDC) of Newfoundland & Labrador.

## 5 REFERENCES

[1] Wang, L. C., & Chen, C. C. (1991). A combined optimization method for solving the inverse kinematics problems of mechanical manipulators. *Robotics and Automation, IEEE Transactions on*, *7*(4), 489-499.

[2] Mukundan, R, & Member, S. (2008). A Fast Inverse Kinematics Solution for an n -link Joint Chain, (Icita), 349–354.

[3] D. E. Whitney, Resolved motion rate control of manipulators and human prostheses, IEEE Transactions on Man-Machine Systems, 10 (1969), pp. 47–53.

[4] W. A. Wolovich and H. Elliot, A computational technique for inverse kinematics, in Proc. 23rd IEEE Conference on Decision and Control, 1984, pp. 1359–1363.

[5] A. Balestrino, G. De Maria, and L. Sciavicco, Robust control of robotic manipulators, in Proceedings of the 9th IFAC World Congress, Vol. 5, 1984, pp. 2435–2440.

[6] Muller-Cajar, R., & Mukundan, R. (2007). Triangualation-a new algorithm for inverse kinematics.

[7] GRI Simulations Inc. In *Manipulator Trainer. Retrieved March 10, 2014, from* http://www.grisim.com/products.html#ManipTrainer

[8] Intel. In *Intel® Perceptual Computing SDK 2013, Retrieved March 10, 2014, from* http://software.intel.com/en-us/vcsource/tools/perceptual-computing-sdk

# Toward Scalable Digital Evidence Visualization

Serguei A. Mokhov* Concordia University
Miao Song† Concordia University
Peter Grogono‡ Concordia University
Joey Paquet§ Concordia University
Mourad Debbabi¶ Concordia University

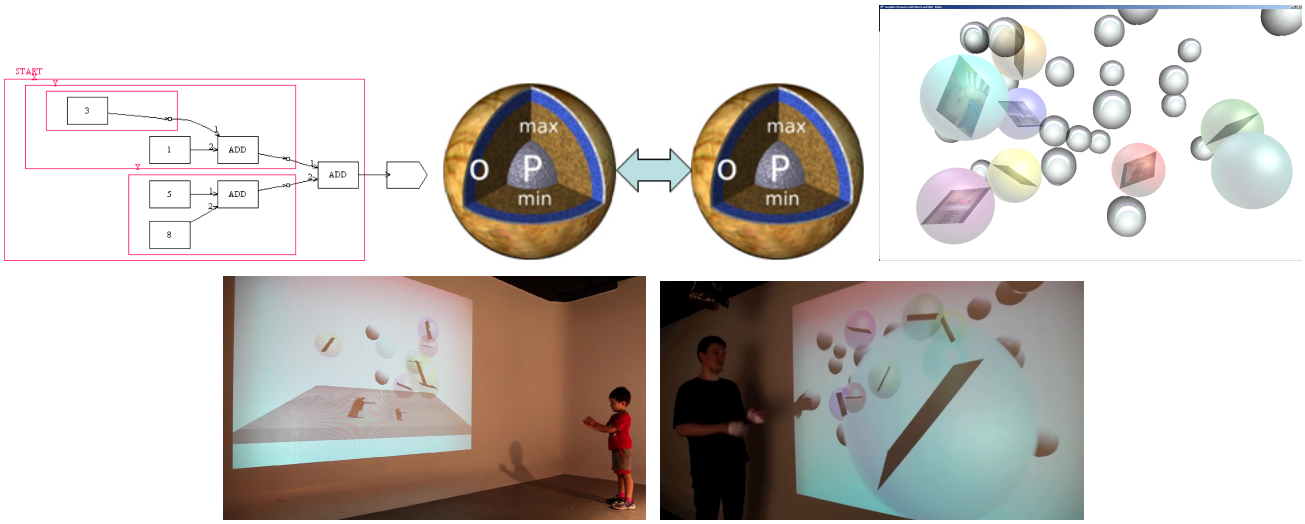



Figure 1: From data-flow graphs to scalable illimitable space visualization and management of digital forensics cases

## Abstract

Visualization requirements in FORENSIC LUCID have to do with different levels of case knowledge abstraction, representation, aggregation, as well as the operational aspects as the final long-term goal of this proposal. It encompasses anything from the finer detailed representation of hierarchical contexts to FORENSIC LUCID programs, to the documented evidence and its management, its linkage to programs, to evaluation, and to management of GIPSY software networks. This includes an ability to arbitrarily switch between those views combined with usable multimodal interaction. The purpose is to determine how the findings can be applied to FORENSIC LUCID and investigation case management. It is also natural to want a convenient and usable evidence visualization, its semantic linkage and the reasoning machinery for it. We present some of these deliberations as a future work item with the detailed related work review.

**Index Terms:** H.5.1 [Multimedia Information Systems]; H.5.2 [User Interfaces]: Group and Organization Interfaces—Evaluation/methodology; D.1.7 [User Interfaces]: Programming Techniques—Visual Programming; D.2.11 [Software Architectures]: Domain-specific architectures; Languages—[FORENSIC LUCID]

*e-mail: mokhov@cse.concordia.ca
†e-mail: m_song@cse.concordia.ca
‡e-mail: grogono@cse.concordia.ca
§e-mail: paquet@cse.concordia.ca
¶e-mail: debbabi@ciise.concordia.ca

## 1 Introduction

We propose a scalable management, visualization, and evaluation of digital evidence using the modified interactive 3D documentary system to represent, semantically link, and provide a usable interface to digital investigators.

Lucid programs are data-flow programs and can be visually represented as data flow graphs (DFGs) and composed visually. FORENSIC LUCID [8] a LUCID dialect, is a language to specify and reason about cyberforensic cases. It includes the encoding of the evidence (representing the context of evaluation) and the crime scene modeling in order to validate claims against the model and perform event reconstruction, potentially within large swaths of digital evidence. To aid investigators to model the scene and evaluate it, we propose to expand the design and implementation of the Lucid DFG programming onto FORENSIC LUCID case modeling and specification to enhance the usability of the language and the system and its behavior in 3D.

The concrete realization of the formal approach also has to be usable by a wider investigative audience who should be able to represent and visualize the voluminous case knowledge and reason about it efficiently. Thus, it is imperative to have usable scripting and visual aid tools to compose the case and import the digital evidence by human investigators. Additionally, the knowledge representation and case building and management should be friendlier to human investigators and take contextual meaning into the account. The subsequent case evaluation should be scalable and efficient at the same time due to the mere fact of a likely possibility to process a large amount of digital evidential data. There are a number of items and proposals in graph-based visualization and the corresponding languages. A discussion on possible visualization tools was carried out in [9] (see also references therein on the related work by several researchers on visualization of load balancing, configuration, formal systems for diagrammatic modeling and visual languages and the corresponding graph systems).

## 2 Related Work

There is a number of items and proposals in graph-based visualization and the corresponding languages. In 1982 Faustini proved that any INDEXICAL LUCID program can be represented as a DFG [5]. In 1995, Jagannathan defined various graphical intensional and extensional models for GLU programming—arguably one of the first practical graph-based visualization proposals for LUCID programs [6]. Paquet subsequently in 1999 expanded on this for multidimensional intensional programs as exemplified in [11]. Stankovic, Orgun, *et al*. proposed the idea of visual parallel programming in 2002 [13]. Ding provided the first implementation of Paquet's notion within the GIPSY project in 2004 [4] using Graphviz's `lefty`'s GUI and `dot` languages [2] along with bi-directional translation between GIPL's or INDEXICAL LUCID's abstract syntax trees (ASTs) and `dot`'s [9]. Related research work on visualization of load balancing, configuration, formal systems for diagrammatic modeling and visual languages and the corresponding graph systems was also presented by several authors in [16, 1, 3, 7]. More recently (2012), another very interesting work of relevance was proposed by Tao *et al*. on visual representation of event sequences, reasoning and visualization [14] (in their case of the EHR data) and around the same time Wang *et al*. proposed a temporal search algorithm for personal history event visualization [15]. Monore *et al*. note the challenges of specifying intervals and absences in temporal queries and approach those with the use of a graphical language [10]. This could be particular useful for no-observations [8] in our case. A multimodal case management interaction system was proposed for the German police called *Vispol Tangible Interface: An Interactive Scenario Visualization*.

## 3 Visualization of Forensic Lucid

The need to represent visually forensic cases, evidence, and other specification components is obvious for usability and other issues. Placing it in 3D helps to structure the "program" (specification) and the case in 3D space can help arrange and structure the case in a virtual environment better with the evidence items encapsulated in the nested 3D spheres akin to Russian dolls, and can be navigated in depth to any level of detail [9].

Illimitable Space System. We explore an idea of a scalable management, visualization, and evaluation of digital evidence with modifications to the interactive 3D documentary subsystem of the *Illimitable Space System* (ISS) [12] to represent, semantically link, and provide a usable interface to digital investigators. The ISS may scale when properly re-engineered and enhanced to act as an interactive "3D window" into the evidential knowledge base grouped into the semantically linked "bubbles" visually representing the documented evidence. By moving such a contextual window, or rather, navigating within the theoretically illimitable space an investigator can sort out and re-organize the knowledge items as needed prior launching the reasoning computation. The interaction design aspect would be of a particular usefulness to open up the documented case knowledge and link the relevant witness accounts and group the related knowledge together. This is a proposed solution to the large scale visualization problem of large volumes of "scrollable" evidence that does not need to be all visualized at once, but be like a snapshot of a storage depot. We propose to re-organize the latter into more structured spaces linked together by the investigators grouping the relevant evidence semantically.

In Figure 1, from left-to-right is an earlier 2D DFG rendition of a LUCID program, followed by the conceptual hierarchical nesting of the evidential statement *es* context elements, (observation sequences *os*, their individual observations *o* (consisting of the properties being observed $(P, \min, \max, w, t)$, details of which are discussed in [8]). These 2D conceptual visualizations are proposed to be renderable at least in 2D or in 3D via an interactive interface to allow modeling complex crime scenes and multidimensional evidence on demand. Then a screenshot from the actual ISS installation hosting multimedia data (documentary videos) users can call out by voice or gestures to examine the contents. We propose to organize the latter into more structured spaces linked together by the investigators grouping the relevant evidence together semantically instead of the data containing bubbles float around.

## References


[1] G. Allwein and J. Barwise, editors. *Logical reasoning with diagrams*. Oxford University Press, Inc., New York, NY, USA, 1996.

[2] AT&T Labs Research and Various Contributors. Graphviz – graph visualization software. [online], 1996–2012. `http://www.graphviz.org/`.

[3] R. Bardohl, M. Minas, G. Taentzer, and A. Schürr. Application of graph transformation to visual languages. In *Handbook of Graph Grammars and Computing by Graph Transformation: Applications, Languages, and Tools*, volume 2, pages 105–180. World Scientific, 1999.

[4] Y. Ding. Automated translation between graphical and textual representations of intensional programs in the GIPSY. Master's thesis, Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada, 2004.

[5] A. A. Faustini. *The Equivalence of a Denotational and an Operational Semantics of Pure Dataflow*. PhD thesis, University of Warwick, Computer Science Department, Coventry, United Kingdom, 1982.

[6] R. Jagannathan. Intensional and extensional graphical models for GLU programming. In *Intensional Programming I*, pages 63–75. World Scientific, 1995.

[7] N. G. Miller. *A Diagrammatic Formal System for Euclidean Geometry*. PhD thesis, Cornell University, U.S.A, 2001.

[8] S. A. Mokhov. *Intensional Cyberforensics*. PhD thesis, Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada, Sept. 2013. Online at `http://arxiv.org/abs/1312.0466`.

[9] S. A. Mokhov, J. Paquet, and M. Debbabi. On the need for data flow graph visualization of Forensic Lucid programs and forensic evidence, and their evaluation by GIPSY. In *Proceedings of PST'11*, pages 120–123. IEEE CS, 2011.

[10] M. Monroe, R. Lan, J. M. del Olmo, B. Shneiderman, C. Plaisant, and J. Millstein. The challenges of specifying intervals and absences in temporal queries: a graphical language approach. In *Proceedings of CHI'13*, pages 2349–2358. ACM, 2013.

[11] J. Paquet. *Scientific Intensional Programming*. PhD thesis, Department of Computer Science, Laval University, Sainte-Foy, Canada, 1999.

[12] M. Song. *Computer-Assisted Interactive Documentary and Performance Arts in Illimitable Space*. PhD thesis, Concordia University, 2012.

[13] N. Stankovic, M. A. Orgun, W. Cai, and K. Zhang. *Visual Parallel Programming*, chapter 6, pages 103—129. World Scientific Publishing Co., Inc., 2002.

[14] C. Tao, K. Wongsuphasawat, K. Clark, C. Plaisant, B. Shneiderman, and C. G. Chute. Towards event sequence representation, reasoning and visualization for EHR data. In *Proceedings of IHI'12*, pages 801–806. ACM, 2012.

[15] T. D. Wang, A. Deshpande, and B. Shneiderman. A temporal pattern search algorithm for personal history event visualization. *IEEE Trans. on Knowl. and Data Eng.*, 24(5):799–812, May 2012.

[16] C. Zheng and J. R. Heath. Simulation and visualization of resource allocation, control, and load balancing procedures for a multiprocessor architecture. In *MS'06*, pages 382–387. ACTA Press, 2006.

# Unified Terrain Synthesis with Large-Scale Structure and Fine-Scale Detail

Maryam Ariyan*
Carleton University

David Mould†
Carleton University

**ABSTRACT**

Sketch-based modeling and procedural modeling are complementary approaches to terrain synthesis.. Procedural modeling can provide detailed features whereas sketch-based modeling can provide user control over high level structures. We propose a framework that combines procedural and sketch-based modeling for geometry synthesis of terrains such as mountain ranges. The user sketches the terrain's silhouette and ridges as seen from ground level, and the system automatically completes a detailed terrain. In this report, we concentrate on the procedural synthesis based on an input set of ridge lines.

Direct solution of the Poisson equation can be used to create a smooth interpolation of input structure, and has been used for terrain synthesis in the past [2]. However, Poisson-based terrain synthesis produces smooth terrains; we plan to create rough terrains directly, using a path planning approach where a heightfield is created in the course of finding a least-cost traversal through a weighted graph [3]. Our approach is still gradient-based. We place a uniform grid over the region of interest and assign initial height values to some cells based on the input ridge heights. Then, in a first pass, we estimate gradient values near the ridge, propagating gradients outward until every cell has been assigned a gradient. In a second pass, we integrate gradients, using Dijkstra's algorithm to determine the integration sequences: cost is inverted height. Small local variations in the gradient values produces roughness in the output terrain.

Once the preliminary terrain has been constructed by the above process, we can add further detail. Relatively unstructured roughness can be added by perturbing some height values upwards and rerunning Dijkstra's algorithm, thus creating local maxima away from the original ridges. More structured features can be added by adding new ridges; we can do this procedurally by exploiting the order of node visits from Dijkstra's algorithm, perturbing all nodes in an uphill sequence back to the original ridge. Unlike the approach of Rusnell et al., we take into account the local arrangement of ridges when constructing our gradients. Unpleasant seams appear in the terrain when edge weights are assigned using the simple random weights or linear profiles of the previous approach.

The most prominent aspect of our terrains is the large-scale structure from the ridge lines. The landscape mainly slopes downhill from the ridges, gradually flattening out as distance from the ridges increases. In part, we average the slopes of nearby ridges to obtain local slope values, thus allowing the terrain to interpolate between ridges without a visible seam. The terrain is roughened by local modifications to the slopes, which can be done either in a structured or unstructured way. In both cases, the final integration pass ensures a unified terrain patch.

In the near future, we plan to complete our experiments with procedural terrain synthesis and editing. The next phase of the work will involve further exploration of the sketch-based aspect of the approach, in which 3D ridges are extracted from a 2D drawing. There are two major research problems on the sketching side. The first is assignment of ridge depth, underconstrained in the sketch; the second is automated completion of partially occluded ridges. Recent results in terrain sketching [1] should give some clues about how to resolve these issues.

*e-mail: maryamariyan@cmail.carleton.ca
†e-mail:mould@scs.carleton.ca

**REFERENCES**

[1] V. A. dos Passos and T. Igarashi. Landsketch: a first person point-of-view example-based terrain modeling approach. In *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling*, pages 61–68. ACM, 2013.

[2] H. Hnaidi, E. Guérin, S. Akkouche, A. Peytavie, and E. Galin. Feature based terrain generation using diffusion equation. In *Computer Graphics Forum*, volume 29, pages 2179–2186. Wiley Online Library, 2010.

[3] B. Rusnell, D. Mould, and M. Eramian. Feature-rich distance-based terrain synthesis. *The Visual Computer*, 25(5-7):573–579, 2009.

# Multi Layer Skin Simulation

Pengbo Li* Paul G.Kry†
McGill University

**ABSTRACT**

We introduce an ongoing approach extending embedded thin shells to multiple layers for physically simulating skin wrinkling deformations. We use adaptive meshes to represent each skin layer and combine them through frequency-based local constraints to form wrinkles with predictable wavelength matching skin physical properties.The multi layer model allows the simulation of minor wrinkles standing on the larger wrinkles, which is a typical pattern of human skin wrinkling. Dealing with the fact that skin has varying elasticity and thickness in different regions, we construct inhomogeneous constraints by sampling a property texture map, accounting for variations in wrinkle wavelength.

**Keywords:** multilayer, embedded shells, frequency-based constraint, inhomogeneous

## 1 INTRODUCTION

As common secondary motion effects, wrinkles on the skin are critical to tell the story of facial expressions or body movements. When people make different facial expressions, wrinkles with various patterns form around eyes or at mouth corners. The major reasons are layered structure of skin and varied thickness and elastic properties. Many techniques have been developed to produce realistic wrinkles for convincing animation. Rohmer et al. [6] created detailed wrinkles procedurally based on stretch tensors of coarse simulation. Larboulette and Cani [3] developed a art-directed system to design the wrinkle as part of character skin. Wrinkle maps are an efficient and most popular method for rendering animated wrinkles on faces and clothes proposed by Blinn et al. [1], but it brings a large amount of workload for designing. Therefore, we want to present a physically based technique to create expressive skin wrinkles automatically, allowing artists can easily control the generation of wrinkles.

Our technique is built upon the approach presented in the previous work by Remillard and Kry [5], which is a new technique for simulating deformations of composite objects where the skin is physically attached to its soft interior. They use coarse resolution lattices to simulate global deformation efficiently and high resolution embedded thin shells to generate the wrinkles under the frequency based local constraints matching the physical properties of composite objects. To accommodate the layered structure of human skin, a multi-layer model as the extension of single layer thin shell is naturally proposed to produce more realistic wrinkles.

Attaching more layers on the base meshes, we couple neighbor layers by frequency-based constraints, using bottom layers to simulate the larger deformation and using top layers to capture wrinkles in finer scale, correspondingly, the constraint density of top layers is higher than the lower layers. Similar in the spirit to composite signals, regarding real skin wrinkles as a complicated signal, we decompose it into a series of simple signals with specific frequency and amplitude in varying scale, then adds them together. Even when constraints are between the same layers, they are non uniform for skin that has varying thickness and elasticity. Given an input mesh, We sample elastic parameters and thickness for each vertex on a property map with respect to the texture coordinates, then run weighted clustering algorithm to construct inhomogeneous constraints. Instead of high resolution meshes, we adopt the approach of adaptive meshes presented by Narian et al. [4], which dynamically refines and coarsens triangle meshes to adjust the resolution locally and automatically. It overcomes the burdensome computation and memory costs for uniformly high-resolution simulation, because it only add small triangles and high resolution details in the regions that need to exhibit small wrinkles.

*pengbo.li@mail.mcgill.ca
†kry@cs.mcgill.ca

## 2 APPROACH

To obtain more realistic wrinkles, we enrich the embedded thin shells with additional layers based on the physiological multilayered structure of human skins, running internal elastic dynamic simulation for each layer and coupling them through frequency-based constraints with local support. Constraints are constructed based on critical wavelength that is computed by elastic parameters (Young's modulus and the Poisson ratio) and thickness of each layer. Finally, the tiny wrinkles are able to appear on the relatively larger wrinkles.

### 2.1 Multi-Layer Framework

Following the scheme of embedded thin shells, we build embedding relation between two layers using barycentric interpolation weights of the upper layer positions with respect to the lower layer positions, which is consistent with the situation that we use triangular meshes to simulate. The upper layer is forced to attach to the lower layer by constraints with local support. Given the position vector of lower layer $\mathbf{x}_{k-1} \in \mathbb{R}^m$, the upper layer position $\mathbf{x}_k \in \mathbb{R}^n$ is forced to match the embedded shape,

$$H_k \mathbf{x}_k = H_k B_k \mathbf{x}_{k-1} + t_k n_{k-1}, \tag{1}$$

where $B_k : \mathbb{R}^m \to \mathbb{R}^n$ is the linear embedding relation, $H_k : \mathbb{R}^n \to \mathbb{R}^c$ is our sparse constraint matrix. $k \in \{1,2,\cdots\}$, $\mathbf{x}_0$ represents the base mesh. Considering the thickness of each layer, we locally add the product of normal $n_{k-1}$ and thickness scalar $t_k$ to lower layer position.

### 2.2 Inhomogeneous Constraints

We create weak form constraints with local support. Within a cluster of vertices belonging to the upper layer, each constraint force the weighted position average of those vertices to match a corresponding weighted position average of embedded shape on the lower layer. For a given cluster of vertices $C$, the constraints is given by

$$\sum_{i \in C} \alpha_i x_i^k = \sum_{i \in C} \alpha_i [B x^{k-1}]_i, \tag{2}$$

where $\alpha_i$ is given by a truncated Gaussian function,

$$\alpha_i = \omega_C e^{-\frac{d_i^2}{2\sigma^2}} \text{ if } d_i < 2\sigma, \ 0 \text{ otherwise}, \tag{3}$$

where $\omega_C$ normalizes the sum of weights to one ensuring affine combination, and $d_i$ is the distance from vertex i to the cluster's center. The truncation $2\sigma$ determines the vertex presence in the

cluster, namely, is the radius of clusters. From the perspective of mesh reconstruction, our coupling process can be viewed as a filtered geometry reconstruction with limited frequency, where Gaussian weights play the role of local low-pass filter.

To reach the requirements of inhomogeneous constraints, we obtain distance weight for each vertex through sampling on predesigned property map with respect to texture coordinates over the entire mesh. Applying weighted clustering algorithm, the base mesh is partitioned into non uniform regions cross the surface (see Figure 2). Given the standard radius of clusters $\mathbf{r}_s$, we start by choosing a random vertex as the first center, then compute the edge traversal distance to all other vertices using breadth first search. When computing distance between two vertices, the length of edge is divided by the average weight of two vertices. Among the vertices that at least is $\frac{1}{2}\mathbf{r}_s$ distance from existing centers, we choose the farthest vertex as the next center and compute distances from all other vertices to this center. We repeat until all vertices have at most a distance of $\frac{1}{2}\mathbf{r}_s$ to a center.

### 2.3 Wrinkles Parameters

Embedded thin shells applied a simple model that has been used to describe the mechanical behavior of a thin film resting on top of a soft elastic foundation [7]. Among the alternative ways of obtaining the expression for the wrinkle wavelength, a general form of the wrinkle periodicity in human skin is given by Cerda and Mahadevan et al. [2]. Given the thickness of the upper layer $h_k$ and lower layer $h_{k-1}$, and the Young's modulus ($E_k$ and $E_{k-1}$ respectively), the critical wavelength is given by

$$\lambda_k \sim (h_k h_{k-1})^{1/2} \left(\frac{E_k}{E_{k-1}}\right)^{1/6}. \tag{4}$$

Once obtaining the critical wrinkling wavelength, we can compute the radius $\mathbf{r}$ of clusters and the standard deviation $\sigma$ using $\sigma = \lambda/\pi$ and $\mathbf{r} = 2\sigma$ according to Nyquist-Shannon sampling theorem [5].

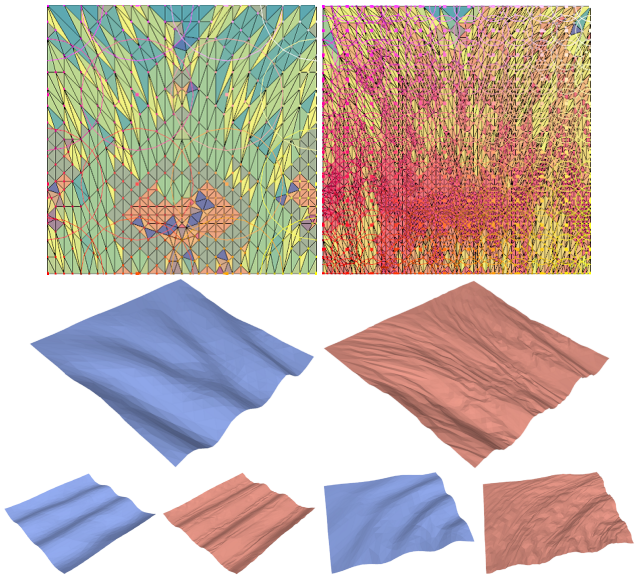


Figure 1: A flat skin with two layers undertake three types of compression. The top is the material space configuration of the example in the middle, the uniformly distributed circles indicate the local constraints with constant frequency. The blue one is the first layer, and the red one is the second layer.

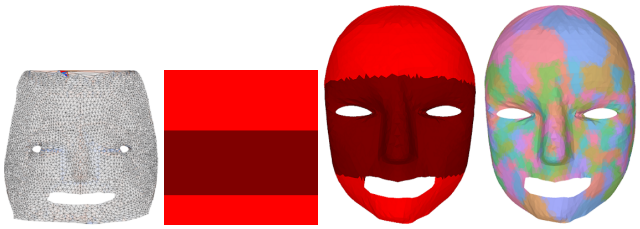


Figure 2: An example mesh with inhomogeneous elastic properties partitions into non uniform regions. From left to right, they are texture coordinates, property map, sampling results, clustering results. Given same elasticity and foundation thickness, the brightness of red color indicates the thickness of skin layer. The darker region is thinner than the bright region, resulting in the smaller critical wavelength and thus radius of clusters.

## 3 Results

Figure 1 shows preliminary results where multi layer model is implanted into a two layer flat example. The second layer is thinner and stiffer than first layer, leading to be more wrinkling and have smaller critical wavelength under same compression. Coupling two layers, we produce a complex wrinkle pattern that is easy to be found when you compress your skin or make expressions. Figure 2 shows the procedure to construct inhomogeneous constraints. After sampling on the property map, we compute critical wavelength for each vertex and then convert them to distance weights. The clustering results displays that the larger the value is, the larger radius a cluster has. In this case, we only sample the thickness of top layer. However, we are able to sample any other properties, or even directly get critical wavelength.

## 4 Conclusion and Future Work

We presented a approach applying multi layer framework to simulate skin wrinkles based on embedded thin shells. The model we proposed are able to capture detailed wrinkling features on the human skin and allow artists automatically to obtain visually realistic wrinkles. In the future work, we plan to incorporate this method into facial animations that are driven by blend shape deformation model, showing vivid and expressive animations. While our physically based technique is insufficient for simulation at the interactive rate, our approach can easily be extend to automate the creation of wrinkle maps for a real-time application. Furthermore, another potential approach to construct non uniform constraints in the friendly art style, determining the distribution of constraints according to the sketched lines made by artists on the base model.

## References


[1] J. F. Blinn. Simulation of wrinkled surfaces. *SIGGRAPH Comput. Graph.*, 12(3):286–292, Aug. 1978.

[2] E. Cerda and L. Mahadevan. Geometry and physics of wrinkling, 2003.

[3] C. Larboulette and M.-P. Cani. Real-time dynamic wrinkles. In *Proceedings of the Computer Graphics International*, CGI '04, pages 522–525, Washington, DC, USA, 2004. IEEE Computer Society.

[4] R. Narain, A. Samii, and J. F. O'Brien. Adaptive anisotropic remeshing for cloth simulation. *ACM Trans. Graph.*, 31(6):152:1–152:10, Nov. 2012.

[5] O. Rémillard and P. G. Kry. Embedded thin shells for wrinkle simulation. *ACM Trans. Graph.*, 32(4):50:1–50:8, July 2013.

[6] D. Rohmer, T. Popa, M.-P. Cani, S. Hahmann, and A. Sheffer. Animation Wrinkling: Augmenting Coarse Cloth Simulations with Realistic-Looking Wrinkles. *ACM Transactions on Graphics (TOG). Proceedings of ACM SIGGRAPH ASIA.*, 29(5), 2010.

[7] S. P. Timoshenko and J. M. Gere. *Theory of Elastic Stability*. Dover Civil and Mechanical Engineering Series, 2009.

# Contact Classification

Charles Bouchard* Paul G. Kry†

McGill University

**ABSTRACT**

During the contact simulation of a pile of objects, some contacts switch from being active to non-active. Contacts that switch configuration can be responsible for longer convergence time, more instability and possibly unpleasant artifacts. We use a classifier to identify those contacts before the solver step. This information could be used for a pre-solver step to accelerate the convergence time. We show that a machine learning approach including strong feature construction and a random tree learner can achieve a recall rate of 37.7% on the unstable contacts.

## 1 INTRODUCTION

Iterative solvers update the position or velocity of the bodies at each step. In a projected Gauss-Siedel (PGS) solver, this is done one contact at a time, with the update of one body affecting multiple other contacts. In particular, we look at when the Langrange multipliers meet the boundary of the constraints. This can be a contact that starts sliding, or two bodies that separate or interpenetrate at a contact point. During a simulation, certain arrangements of contacts and bodies might favour the apparition of unstable contacts. In the presence of unusually strong constraint forces, a contact very close to slipping or a dense cluster of bodies all in contact with each other could intuitively be labeled as a region where stability would be obtained after more iterations than elsewhere in the simulation. We use a machine learning approach to detect these situations by looking at some properties of each contact and their neighbourhood.

There are many methods for solving the contact problem, both direct and iterative methods. Compmlementarity formulations were introduced to the graphics community by Baraff [1] with acceleration-based updates. More recently, Erleben [2] uses PGS as an iterative approach with a velocity based shock propagation to improve the convergence rate. Kaufman et al. [3] use staggered projections to alleviate the lack of stability and accuracy that cripples other formulations. Recent work by Tonge [5] introduces the idea of mass-splitting to improve the convergence rate of the Jacobi-based method and is able to further accelerate the process by solving the contacts in blocks. Furthermore, the idea of trying to find ordering the contacts has been explored by Lacoursiere [4] to transform the problem into a block-wise PGS problem and distribute the computation on different CPUs.

In the present work, we present a novel approach to order the contacts based on their estimated stability using an algorithm that classifies the contacts into stable and unstable classes, where an unstable contact is defined to switch configuration two or more times during a PGS solve. Further work will involve using this information to improve a contact solver.

## 2 METHOD

Each contact is represented as a set of features that are computed before the solving step. The main idea is to collect the maximum amount of useful information about the contacts. Those features for a given contact $i$ are collected in a vector $X_i$. The features can be divided into two main categories, $X = [X_{i1}, X_{i2}]$. The first part includes all the information about a single contact, such as the mass of the two bodies in contact $(m_1, m_2)$, its position $(x, y)$ and the relative velocity of body 1 and 2 $(\dot{x}, \dot{y})$. We also want to use information coming from previous time step, such as the previous normal and frictional Lagrange multipliers computed for this contact $(\lambda_0, \lambda_1)$, if is this contact is new, and the angle from the friction cone $(\theta)$. This last feature quantifies how close the contact was to start sliding in the previous time step.

$$X_{i1} = \{m_1, m_2, \dot{x}, \dot{y}, new, reused, \lambda_0, \lambda_1, \theta\} \quad (1)$$

Note that some of these features are reals, some are integers and some are binary.

The second category includes information about the neighbourhood of the contact. The connectivity of all contacts between a collection of bodies can be represented by a graph in which the bodies are nodes and the groups of contacts between two given bodies form the edges. We weight the edge by the number of contacts between a pair of bodies. For example, two bodies that are in contact at 5 points will be linked by an edge of weight 5. We want to extract information from such a structure. We do it by creating features based on the various degrees of connectivity associated with a contact. Examples features include the total number of edges connected to body 1 and body 2 $(N_e)$ and the weight of the edge in which this contact lives $(W_e)$. Again, we exploit the information contained in the last time step by counting the number of unstable contacts in the vicinity of the contact $(U_c)$, and the number of edges which had unstable contacts $(UI_e)$. Figure 1 shows a visual interpretation of the neighbourhood of a set of contacts.

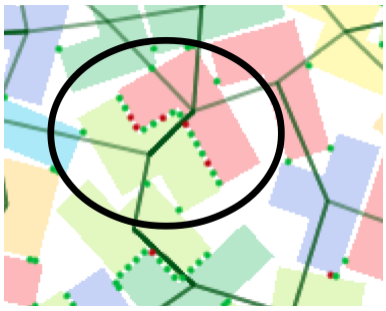


Figure 1: A small part of the contact graph. The circle represents the first degree neighbourhood of the contact, meaning that information comes only from the edges which are directly connected to body 1 and body 2.

$$X_{i2} = \{N_e, W_e, U_c, U_e\} \quad (2)$$

Finally, the number of times the contact switches between active and inactive is stored in $Y_i$. This is called the class of the sample.

*e-mail:charles.bouchard@mail.mcgill.ca

†e-mail: kry@cs.mcgill.ca

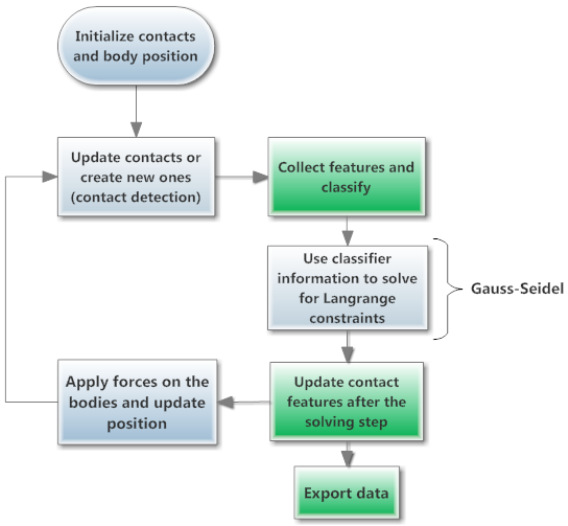



Figure 2: Illustration of the pipeline to collect and build the classifier. The green boxes are the additional steps introduces by our method.

| Features | Correlation |
|---|---|
| $U_n$ | 0.21 |
| reused | 0.19 |
| $\theta$ | 0.16 |
| $\lambda_0$ | 0.11 |
| $\lambda_1$ | 0.10 |

Table 1: Pearson correlation coefficients of the five most important features.

### 2.1 Pipeline

The implementation we use to perform the simulation to gather data for learning is a 2D implementation of the PGS solver described by Erleben [2]. The learner comes into play before and after the iterative solver at each time step, as shown by the green boxes in Figure 2.

A random tree classifier is built using those feature vectors. The data is initially accumulated by sampling 4 groups of 10 time steps. Given that there is around 800 contacts in our simulation, the process creates around 30 000 samples. The set of examples $[X_i, Y_i]$ are passed into Weka, a machine learning open-source tool to compute the parameters of the classifiers.

### 2.2 Contact Distribution and Feature Analysis

This problem inherently comes with an unbalanced sample population. There is a lot more stable contacts than unstable contacts. The histogram in Figure 3 shows that around 95% of the contacts are stable. Most classifier theory works under the assumption that the population is uniformly distributed between classes. Our case is at the other end of the spectrum. To fix alleviate this problem, we use oversampling to balance the population, by effectively throwing away the good contacts at random.

The main challenge is then to find the optimal combination of features that gives the best performance metrics for our classifier. We use the Pearson correlation coefficients of all the feature with the class label. The higher the correlation, the more information this feature will give us. Table 1 gives the five most powerful features to use for classification.

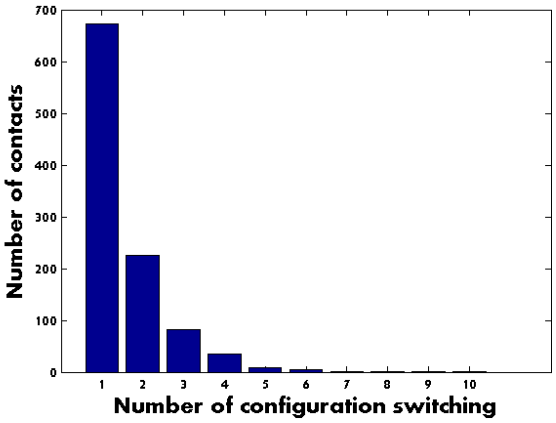



Figure 3: Histogram of the number of contacts for different number of configuration switches averaged over all the sampled time steps.

| | TP rate | FP rate | Recall | ROC |
|---|---|---|---|---|
| Stable | 0.856 | 0.623 | 0.856 | 0.856 |
| Unstable | 0.377 | 0.114 | 0.377 | 0.265 |
| Weighted Avg. | 0.767 | 0.534 | 0.768 | 0.747 |

Table 2: Performance metrics for the random tree classifier on 25507 examples. The recall value of 0.377 means that 37.7% of the unstable contacts were detected.

## 3 Results and Conclusion

Learning was done offline on 25507 examples of contacts fetched over 30 animation frames. The simulation included 250 2D Tetrix blocks piling up. Each frame included a 20 iterations Gauss-Siedel solve. The learning algorithm used was a random tree with 5 random features selected at each round. 10-fold cross-validation was used to prevent overfitting. The most important performance metric to augment was the recall rate of the unstable class, as it represents the number of unstable contact detected by the classifier, and therefore that we can fix. A list of performance metrics are included in Table 2.


## Acknowledgements

We thank Joelle Pineau for sharing her expertise in machine learning.


## References


[1] D. Baraff. Fast contact force computation for nonpenetrating rigid bodies. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '94, pages 23–34, New York, NY, USA, 1994. ACM.

[2] K. Erlenben. Velocity-based shock propagation for multibody dynamic animation. *ACM Trans. Graphics*, 26(2):12, 2007.

[3] D. M. Kaufman, S. Sueda, D. L. James, and D. K. Pai. Staggered projections for frictional contact in multibody systems. *ACM Trans. Graph.*, 27(5):164:1–164:11, Dec. 2008.

[4] C. Lacoursire. A parallel block iterative method for interactive contacting rigid multibody simulations on multicore pcs. In B. Kgstrm, E. Elmroth, J. Dongarra, and J. Waniewski, editors, *Applied Parallel Computing. State of the Art in Scientific Computing*, volume 4699 of *Lecture Notes in Computer Science*, pages 956–965. Springer Berlin Heidelberg, 2007.

[5] A. V. R. Tonge, F. Benevolenski. Mass splitting for jitter-free parallel rigid body simulation. *ACM Trans. Graphics*, 31(4), 2012.

# A New Approach for Evaluation of Stereo Correspondence Solutions in Augmented Reality

Bahar Pourazar* Oscar Meruvia-Pastor†

Department of Computer Science, Memorial University of Newfoundland, NL, Canada

## 1 Introduction

For many years, researchers have made great contributions in the fields of augmented reality (AR) and stereo vision. One of the most studied aspects of stereo vision since the 1980s has been *Stereo Correspondence*, which is the problem of finding the corresponding pixels in stereo images, and therefore, building a disparity map. As a result, many methods have been proposed and implemented in order to properly address this problem.

Due to the emergence of different techniques to solve the problem of stereo correspondence, having an evaluation scheme to assess these solutions is essential. Over the past few years, different evaluation schemes have been proposed by researchers in the field to provide a testbed for assessment of the solutions based on specific criteria. For instance, the Middlebury Stereo [8] and the Kitti Stereo benchmarks [2] are two of the most popular and widely used evaluation systems through which a solution can be evaluated and compared to others. However, both of these models take a general approach towards evaluating the methods; that is they have not been designed with an eye to the particular target application. In other words, they mainly focus on the fundamental aspects of designing a stereo algorithm as a solution per se to *efficiently* find the *best matches* of corresponding pixels in stereo pairs. However, this perspective may raise some questions for a punctilious researcher; for instance, "What actually is an *efficient* solution and on what basis is this *efficiency* defined?", or "What is a *best match* of corresponding pixels and how can it be defined?".

In fact, these questions lead to the motivation of evaluating the stereo matching solutions from a different point of view. In this approach, steps are taken towards an evaluation design based on the potential applications of stereo methods, which results in better definition and adjustment of the criteria for *efficiency* and *the best correspondence matches* while doing the evaluation. Since AR has attracted more attention in the past few years, the evaluation scheme proposed in this study is designed based on outdoor AR applications which take advantage of stereo vision techniques to obtain a depth map of the surrounding environment. This map will then be used to integrate virtual objects in the scene that respect the occlusion property and the depth of the real objects. In other words, the motivation of this research is to study the possibility of integrating stereo vision techniques in an AR system, while considering the most important constraints that AR systems normally encounter [5].

## 2 New Evaluation Scheme

In an AR system, there are certain factors that would affect the functionality and effectiveness of the system [5]. One of these factors, that have been the focus of this study, is human factors in AR.
Studies in binocular vision show that the visual system capability to distinguish two objects at different depths relative to each other is limited to a certain threshold [7]. This threshold, which is defined as the minimum detectable depth between two objects at different distances, is known as *stereoacuity*, which varies in different visual systems [7]. According to the geometry of binocular vision illustrated in figure 1, stereoacuity can be calculated using the following formula:

$$\theta = \beta - \alpha = a\Delta Z / Z^2 \tag{1}$$

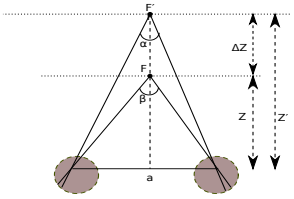



Figure 1: $\theta = \beta - \alpha$ is the stereoacuity, and $a$ is the interpupillary distance, i.e. the distance between the center of the two eyes. $Z$ is the distance of the fixated object ($F$) from the observer, and $\Delta Z$ is the relative depth between the fixated and moving object ($F'$).

In the proposed evaluation scheme, unlike the Middlebury or Kitti benchmarks, a pixel in the disparity results is labeled as an *outlier* if the corresponding angular measurement of the depth error between the ground truth depth and the depth value found by the stereo solution is more than the standard stereoacuity threshold for the human visual system (HVS) as determined by standard stereo tests [7, 1]. Moreover, the average stereoacuity for different age groups [1] is used to determine the performance of the algorithm for users at different ages; this makes the evaluation results more reliable and applicable to practical applications of AR.

To evaluate the performance of an algorithm to investigate whether it meets the requirements for being part of a real-time AR application [3], a module is integrated in the evaluation process that can report on the average execution time of the algorithm over the input dataset.

In addition, an approach is introduced in the model which can be of specific value to practical AR applications. This approach suggests that it is prudent to focus the evaluation procedure on particular regions of the disparity map rather than the whole image for AR applications. The main hypothesis is that salient edges caused by depth discontinuities, which can also represent the object boundaries and occlusion, are important depth cues in the HVS and can help the observer to better perceive the depth of different objects in the scene [9]. Consequently, in an AR application, finding more accurate corresponding matches in these regions can lead to a better alignment between the virtual and the real objects in the scene. This permits a higher quality combination of the depth map of the real world with the virtual depth of the synthetic objects that are part of the AR scene.

## 3 Validation

In order to assess the proposed evaluation model and investigate the validity of the hypotheses for its design, experiments have been

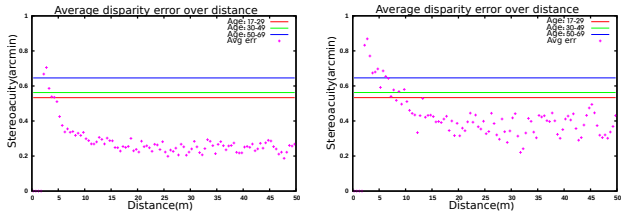
*e-mail: b.pourazar@mun.ca
†e-mail: oscar@mun.ca

conducted on two sample stereo matching algorithms: First, Semi-global block matching, also known as SGBM, which is a modified version of semi-global matching by Hirshmuller [4] and is now integrated within the OpenCV library. Second, an implementation of the solution proposed by Mei et al. [6], "On building an accurate stereo matching system on graphics hardware", known as ADCensus. It should be noted that the CPU implementation of both algorithms have been used in this study.

Experiments were carried out on a Linux platform with Intel Core(TM) i7 3.20GHz CPU. Fifty-two image pairs were chosen from the Kitti Stereo Dataset representing real outdoor scenes. A Canny edge detector was used for building the specified masks and a Dilation operation was used for expansion of the masked areas. Parameters corresponding to stereo algorithms, relative threshold in Canny, and the degree of Dilation were kept constant over all the images and experiments. Standard stereoacuities used for evaluation are based on the results published in [1]. These values for age ranges of 17-29, 30-49, 50-69 and 70-83 were set to 32, 33.75, 38.75 and 112.5 arcseconds, respectively.

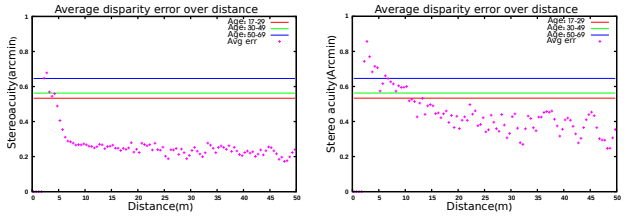
## 4 Main Results

Sample generated results by the evaluation model on both SGBM and ADCensus algorithms over the masked regions and the whole disparity image are shown in figures 2 and 3, respectively.




(a) Average disparity errors by SGBM (b) Average disparity errors by ADCensus

Figure 2: Average disparity errors over distance in masked regions




(a) Average disparity errors by SGBM (b) Average disparity errors by ADCensus

Figure 3: Average disparity errors over distance in the whole image

In these plots, a cross point below a stereoacuity threshold (straight lines) implies that the average error in the disparity values found by the stereo matching algorithm is imperceptible to the human visual system. However, a value higher than the threshold indicates that the error cannot be ignored and should be resolved to achieve a better alignment between the virtual and the real world in the AR application of interest, otherwise it will be noticed by the viewer.

As can be seen in the results, SGBM performs better in finding more accurate corresponding matches compared to ADCensus, as most of the error points fall below the standard stereoacuity lines. Moreover, the plots show that in both methods the significant amount of error corresponds to near field objects, within the first 5 meters. This range of the depth field can be considerably important in some applications, such as the ones involving certain manipulative tasks.

The average error over the masked regions, that is near the depth edges, is very similar to the results over the whole image in the experiments. This may imply that there is no additional benefit in the inspection of these regions. However, this might be merely an indication of the performance of the chosen algorithms, and can be better analyzed by evaluating more algorithms within this model. In either case, it is argued that, due to the importance of occlusion and areas near depth discontinuities to the HVS in AR applications, it is reasonable to focus more on the depth edges and their surroundings when designing or employing a stereo matching technique for an AR application. Furthermore, the average execution time over all the images for SGBM and ADCensus are estimated to be 0.54 and 272.82 seconds, respectively, during the evaluation. Considering the requirements of having an interactive real-time AR system [3], the processing time of each frame should not be more than 0.06-0.08 seconds. Although the current implementation of SGBM could be used when the real world scene remains stable for approximately one second, it can be safely concluded that none of these algorithms meet the requirements of a real-time interactive system. This suggests that GPU-based solutions are essential to obtain the processing speed required for real-time applications.

## 5 Conclusion

In this study, a hypothesis was presented stating that the schemes for evaluating stereo algorithms should be designed based on the specific requirements of the target application. This concept was then applied to the particular application of AR in outdoor environments. As a result, a practical analysis on the performance of the stereo algorithms, in terms of *accuracy* and *execution time* as perceived by the HVS in generating disparity results, was obtained. However, more experiments should be conducted in the proposed model to assess its benefits for other AR applications, such as underwater environments, and explore the other aspects which may be of great value to be considered in the evaluation process, such as the resolution of the display devices, the effect of contrast and brightness, and other relevant factors.


## Acknowledgements

Research funded by the Research and Development Corporation (RDC) of Newfoundland and Labrador.


## References


[1] L. Garnham and J. Sloper. Effect of age on adult stereoacuity as measured by different types of stereotest. *British journal of ophthalmology*, 90(1):91–95, 2006.

[2] A. Geiger. KITTI Vision. `http://www.cvlibs.net/datasets/kitti/eval_stereo_flow.php?benchmark=stereo`, 2012.

[3] A. Hertzmann and K. Perlin. Painterly rendering for video and interaction. In *Proceedings of the 1st International Symposium on Non-photorealistic Animation and Rendering*, NPAR '00, pages 7–12, New York, NY, USA, 2000. ACM.

[4] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):328–341, 2008.

[5] M. A. Livingston. Evaluating human factors in augmented reality systems. *Computer Graphics and Applications, IEEE*, 25(6):6–9, 2005.

[6] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang. On building an accurate stereo matching system on graphics hardware. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 467–474. IEEE, 2011.

[7] R. Reading. *Binocular vision: Foundations and applications*. Butterworths, 1983.

[8] D. Scharstein. MiddleBury Evaluation. `http://vision.middlebury.edu/stereo/eval/`, 2012.

[9] R. Szeliski. *Computer vision: algorithms and applications*. Springer, 2011.

# Real-Time Registration of Highly Variant Colour+Depth Image Pairs

Sahand Seifi* Afsaneh Rafighi† Dr. Oscar Meruvia-Pastor‡

Memorial University of Newfoundland
Computer Science Department

**ABSTRACT**

The focus of this research is to develop algorithms to align colour+depth image pairs taken from the same scene from different positions in real-time. Existing registration address this issue with image pairs that share most of the same scene and have small differences. Other algorithms can align image pairs with higher variation, but they do not perform in real-time. The registration technique proposed in this work uses a combination of 2D image feature detection algorithms and a false feature pair rejection method. It not only performs in real-time, but also supports large transformations with 6 degrees of freedom. Unlike the majority of available methods, the prototype of this technique also performs well when the image pairs have partial overlapping (50 percent or more).

**Index Terms:** I.4.3 [Computing Methodologies]: IMAGE PROCESSING AND COMPUTER VISION—Enhancement;

## 1 INTRODUCTION

3D cameras, a.k.a. depth cameras, provide us with 3D or depth images of a scene. This images consist of an array of pixels with $Z(depth)$ value. Depth cameras are usually accompanied by an auxiliary RGB camera to convey RGB information. Knowing horizontal and vertical field of view (FOV) of the camera, it is easy to convert the position $(x,y)$ of the pixels in the array to real-world position $(X,Y,Z)$. This results in a point cloud which is essentially a partial 3D model of the scene (Similar to Figure 2b).

In some depth sensing scenarios multiple cameras or multiple 3D images are involved. For example when used for indoor localization of robots (as in [3]), hand-held 3D scanning device [5] or motion capture. Given two point clouds of the same scene, registration or alignment methods estimate the transformation that transforms the coordinates of one of the clouds into the other one.

Efficient registration algorithms such as Iterative Closest Point (ICP) [2] and its variants [8] aim to solve this with greedy algorithms that aim to minimize the distance of two clouds (e.g. Sum of Absolute Distance of all points). They work well with small transformations, but easily fall into local optimums in more complex scenarios. On the other hand, algorithms like [6], are better at registering a pair of clouds with significant transformation, but they are costly and they also fall into local optimums when the shared portion of the scene in two images is only partial. This is due to the fact that this algorithms use metrics such as SAD or SSD as an indicator of distance between two clouds. When two clouds only share a portion of the scene (as in Figure 1), the non-overlapping portion of the image will only *increase* SAD and SSD when the images are aligned properly; while a wrong alignment that pushes the two clouds together produces less distance.

*e-mail: sahands (at) mun.ca
†e-mail: ar7107 (at) mun.ca
‡Assistant Professor, Department of Computer Science and Office of the Dean, Faculty of Science. e-mail: oscar (at) mun.ca

(a) Image 1 (b) Image 2

Figure 1: Partially overlapping image pair. Overlapped portion has been indicated with red rectangles.

In this work, a registration technique has been proposed that performs well regardless of the initial pose of the pairs. There is no restriction over the 6 degrees of freedom. It also performs well with partially overlapping pairs that cause most algorithms fall into local optimum, because it uses metrics that do not take all existing points into account. Finally, it runs in real-time ($>$ 20 fps).

## 2 METHODS

First SURF [1] or ORB [7] 2D feature detection algorithms are used to extract features from the 2D image provided by the depth cameras. Features from each image are matched together to derive feature pairs using brute force (Figure 2). OpenCV [4] GPU implementation of feature detection and brute force matcher algorithms are used for these steps.

In the second step, feature pairs from the 2D images are converted to their corresponding 3D points in the cloud. A noticeable number of features pairs are lost at this step since the point clouds acquired from depth cameras are incomplete due to missing information (e.g. shiny surfaces).

The purpose of the third step is to remove incorrect feature pairs. In this step 3 feature pairs are randomly selected: $(A1,B1,C1)$ from image 1 and their respective pairs $(A2,B2,C2)$ from image 2. A transformation $T$ is estimated that closely converts $(A1,B1,C1)$ to $(A1',B1',C1')$. Note that $(A1',B1',C1')$ and $(A2,B2,C2)$ would be identical if all the 3 pairs are precisely detected feature pairs, otherwise perfect matching of these pairs would be impossible.

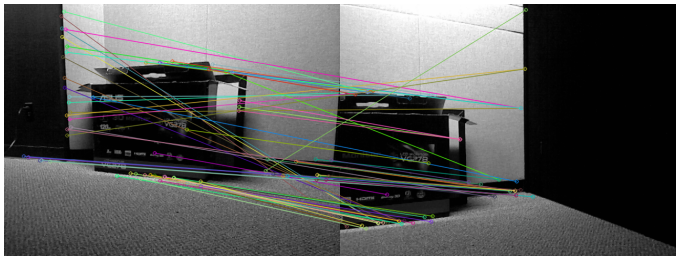
Next, a score is calculated for this random selection. The proposed score is the number all features that have a certain distance or less to their respective pairs after applying the transformation $T$. The distance threshold is selected in a way that any pair with satisfactory distance would not be deemed far apart visually.

If the estimated transformation $T$ is a correct transformation, this score is detecting and counting the correctly detected features. Assuming 50 percent of feature pairs are correct, there is a 12.5 percent chance that the randomly selected 3 features generate a good transformation. To ensure a good transformation $T_f$ is found, this step is repeated multiple times ($2 \times NumberOfFeatures$) and the random set that generates the best score is selected to estimate $T_f$.

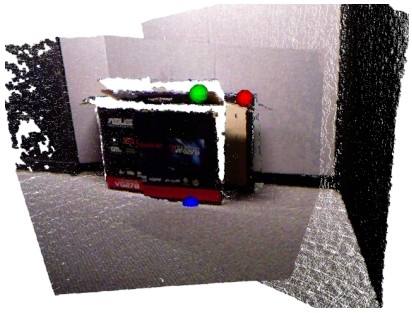
In step 4, based on $T_f$ all the feature pairs that have a distance larger than the distance threshold are removed from the feature set. Considering that $T_f$ is very close to the optimal transformation, incorrect features are removed and only features with minimal error

| Image Set | No. of Features | Accepted Features | Feature Detection Time | Registration Time (Step 2 to 5) | FPS |
|---|---|---|---|---|---|
| Box | 68 | 31 | 15.53 ms | 32.64 ms | 20.75 |
| Wall | 43 | 22 | 16.91 ms | 26.20 ms | 23.19 |
| Mug | 56 | 29 | 22.61 ms | 21.70 ms | 22.56 |

Table 1: Performance



(a) Features



(b) Final Alignment Result - The coloured spheres indicate the selected feature pairs.

Figure 2: Test Pair 1 - Box

remain. The goal of step 5 is to find the optimal transformation by following the same process as in step 3: selecting 3 random elements from the remaining features pairs, generating transformation for this set, calculating the score, repeat. The optimal transformation $T_{opt}$ is considered to be the transformation with the highest score. Considering the fact that all the incorrect feature pairs are removed from the pair set, $T_{opt}$ can be achieved by enough repetitions ($2 \times NumberOfFeatures$ in our case).

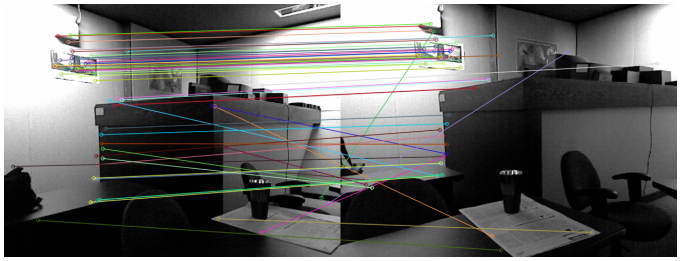
## 3 Results

To test this method, a Microsoft Kinect camera is used to capture colour+depth images. The workstation used for the tests is an Intel Core-i7 960 (3.2Ghz) machine with a NVidia Quadro K5000 GPU.

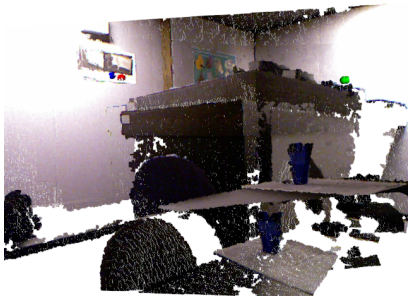
The tests are performed on three scenes: Box, Wall and Mug. The RGB images of the Box and the Mug scenes with detected features and final alignment result is depicted in Figure 2 and Figure 3. Table 1 outlines the performance with the number of features that were used for each test.

## 4 Conclusion

It is observed that optimal alignment can be achieved with this method in many cases such as in the Box and the Wall scene in real-time. As seen in Figure 3, this method does not always achieve perfect alignment. This is especially true for cases where the detected features are clustered together and are relatively far from the camera. The misalignment that happens in such cases is due to two reasons: 1) With the depth camera that was used, further points have less accuracy in their coordinates. 2) When features are clustered



(a) Features



(b) Final Alignment Result - The coloured spheres indicate the selected feature pairs.

Figure 3: Test Pair 3 - Mug

together, the smallest error with any of the 3 randomly selected features causes large rotation errors in the generated transformation because the features are very close to each other.

When selecting 3 random pairs, it is ideal that they are far apart from each other with maximum angle between them (almost equilateral). To avoid this issue, the coordinates of the features might be used for a weighted selection of the 3 feature pairs rather than a completely random selection. ICP or similar algorithms can be used with the feature pairs and their surrounding points to fine tune the alignment.


## Acknowledgements

The authors wish to thank Research and Development Corporation (RDC) and CREAIT for providing the funding and physical space for this research.


## References


[1] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. In *In ECCV*, pages 404–417, 2006.

[2] P. Besl and N. D. McKay. A method for registration of 3-d shapes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(2):239–256, Feb 1992.

[3] J. Biswas and M. Veloso. Depth camera based indoor mobile robot localization and navigation. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1697–1702, May 2012.

[4] G. Bradski. Opencv library. *Dr. Dobb's Journal of Software Tools*, 2000.

[5] D. Huber. Automatic 3d modeling using range images obtained from unknown viewpoints. In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pages 153–160, 2001.

[6] M. Magnusson, A. Lilienthal, and T. Duckett. Scan registration for autonomous mining vehicles using 3d-ndt. *Journal of Field Robotics*, pages 803–827, 2007.

[7] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2564–2571, Nov 2011.

[8] S. Rusinkiewicz and M. Levoy. Efficient variants of the icp algorithm. In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pages 145–152, 2001.