University of Kentucky
UKnowledge

Theses and Dissertations--Electrical and Computer Engineering

Electrical and Computer Engineering

2016

# WEARABLE PRIVACY PROTECTION WITH VISUAL BUBBLE

Shaoqian Wang
*University of Kentucky*, shaoqian.wang@gmail.com
Author ORCID Identifier:
http://orcid.org/0000-0002-4912-4616

Digital Object Identifier: https://doi.org/10.13023/ETD.2016.514

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation
Wang, Shaoqian, "WEARABLE PRIVACY PROTECTION WITH VISUAL BUBBLE" (2016). *Theses and Dissertations--Electrical and Computer Engineering*. 97.
https://uknowledge.uky.edu/ece_etds/97

This Master's Thesis is brought to you for free and open access by the Electrical and Computer Engineering at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Electrical and Computer Engineering by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student’s advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student’s thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Shaoqian Wang, Student

Dr. Sen-Ching S. Cheung, Major Professor

Dr. Cai-Cheng Lu, Director of Graduate Studies

# WEARABLE PRIVACY PROTECTION WITH VISUAL BUBBLE

---

THESIS

---

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in Electrical Engineering
in the College of Engineering
at the University of Kentucky

By

Shaoqian Wang

Lexington, Kentucky

Director: Dr. Sen-Ching S. Cheung, Professor of Electrical and Computer

Engineering

Lexington, Kentucky

2016

Copyright© Shaoqian Wang 2016

# ABSTRACT OF THESIS

# WEARABLE PRIVACY PROTECTION WITH VISUAL BUBBLE

Wearable cameras are increasingly used in many different applications such as entertainment, security, law enforcement and healthcare. In this thesis, we focus on the application of the police worn body camera and behavioral recording using a wearable camera for one-on-one therapy with a child in a classroom or clinic. To protect the privacy of other individuals in the same environment, we introduce a new visual privacy protection technique called visual bubble. Visual bubble is a virtual zone centered around the camera for observation whereas the rest of the environment and people are obfuscated. In contrast to most existing visual privacy protection systems that rely on visual classifiers, visual bubble is based on depth estimation to determine the extent of privacy protection. To demonstrate this concept, we construct a wearable stereo camera for depth estimation on the Raspberry Pi platform. We also propose a novel framework to quantify the uncertainty in depth measurements so as to minimize a statistical privacy risk in constructing the depth-based privacy bubble. To evaluate our system, we have collected three datasets. The effectiveness of the proposed scheme is demonstrated with experimental results.

KEYWORDS: privacy protection, visual bubble, privacy bubble, wearable camera, depth uncertainty, stereo quantization

Author's signature: Shaoqian Wang

Date: December 9, 2016

# WEARABLE PRIVACY PROTECTION WITH VISUAL BUBBLE

By

Shaoqian Wang

Director of Thesis: Sen-Ching S. Cheung

Director of Graduate Studies: Cai-Cheng Lu

Date: December 9, 2016

# ACKNOWLEDGMENTS

I would like to thank my thesis advisor Dr. Samson Cheung for his guidance and support throughout my research project. He has provided me much valuable advice about research and scientific writing and is always ready to share new ideas. I would also like to thank my committee members, Dr. Yuming Zhang and Dr. Bruce Walcott, for their valuable support and suggestions.

Additionally, I would like to thank my friends and fellow collegues of the Multimedia Information Analysis Laboratory: Ying Luo, Zhaohong Wang, Nkiruka "Nikky" Uzuegbunam, Wanxin Xu, Po-Chang Su and Hasan Sajid. Specifically, I would like to thank them for the beneficial discussions about computer vision and for their willingness to be the guinea pigs for testing my algorithms.

Finally, I would like to thank my Ph.D. program advisor, Dr. T. M. Seigler, whose understanding and support has made all this work possible.

# Table of Contents

Acknowledgments . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . iii

Table of Contents . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . iv

List of Figures . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . vi

List of Tables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . viii

Chapter 1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . 1

Chapter 2 Related Work . . . . . . . . . . . . . . . . . . . . . . . . . 6
- 2.1 Review of visual privacy protection techniques . . . . . . . . . . . . 6
- 2.2 Review of stereo matching . . . . . . . . . . . . . . . . . . . . . 8

Chapter 3 Visual Bubble System . . . . . . . . . . . . . . . . . . . . . 11
- 3.1 Depth bubble and depth from stereo . . . . . . . . . . . . . . . . . 12
- 3.2 Uncertainty in stereo depth measurement . . . . . . . . . . . . . . . 13
  - 3.2.1 Uncertainty due to quantization . . . . . . . . . . . . . . . . 14
  - 3.2.2 Uncertainty from stereo correspondence ambiguity . . . . . . . 16
- 3.3 Uncertainty aware privacy bubble . . . . . . . . . . . . . . . . . . 17
- 3.4 Post processing with superpixel . . . . . . . . . . . . . . . . . . 18

Chapter 4 Hardware Implementation and Experiment . . . . . . . . . . . . 21
- 4.1 Hardware implementation . . . . . . . . . . . . . . . . . . . . . . 21
- 4.2 Dataset collection . . . . . . . . . . . . . . . . . . . . . . . . . 24
  - 4.2.1 Dataset collection using Kinect 2 . . . . . . . . . . . . . . . 26
  - 4.2.2 Ground truth depth calculation . . . . . . . . . . . . . . . . 27
  - 4.2.3 Validation of the ground truth depth . . . . . . . . . . . . . . 30
- 4.3 Experimental results . . . . . . . . . . . . . . . . . . . . . . . . 31
  - 4.3.1 Depth bubble results . . . . . . . . . . . . . . . . . . . . . 33

4.3.2 Privacy bubble results . . . . . . . . . . . . . . . . . . . . . . . . 38

Chapter 5 Conclusion and Future Work . . . . . . . . . . . . . . . . . . . . . . . . 43

Bibliography . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45

Vita . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52

# List of Figures

1.1 Wearable privacy camera . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
1.2 Privacy bubble implemented with Kinect 2 camera . . . . . . . . . . . . 4

3.1 Visual bubble system . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
3.2 Quantization effect . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
3.3 Stereo imaging model . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
3.4 Superpixel example . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

4.1 System Diagram . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
4.2 System components . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
4.3 Hardware implementation . . . . . . . . . . . . . . . . . . . . . . . . . 23
4.4 Pi camera mount CAD model . . . . . . . . . . . . . . . . . . . . . . . 23
4.5 System case STL model . . . . . . . . . . . . . . . . . . . . . . . . . . 24
4.6 Dataset 1 sample (left) images . . . . . . . . . . . . . . . . . . . . . . . 24
4.7 Dataset 2 sample (left) images . . . . . . . . . . . . . . . . . . . . . . . 25
4.8 Stereo kinect setup . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
4.9 Dataset 3 sample (left) images . . . . . . . . . . . . . . . . . . . . . . . 27
4.10 Stereo rectification . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
4.11 Dataset 3 sample depth images . . . . . . . . . . . . . . . . . . . . . . 30
4.12 Depth validation by comparing depth of matching templates . . . . . . . 31
4.13 Depth bubble results: dataset 1 . . . . . . . . . . . . . . . . . . . . . . 34
4.14 Depth bubble: dataset 2 . . . . . . . . . . . . . . . . . . . . . . . . . . 35
4.15 Histogram of disparity errors for dataset 3 . . . . . . . . . . . . . . . . 36
4.16 Depth bubble: dataset 3 . . . . . . . . . . . . . . . . . . . . . . . . . . 37
4.17 Privacy bubble generation . . . . . . . . . . . . . . . . . . . . . . . . . 39
4.18 Privacy bubble results: dataset 1 . . . . . . . . . . . . . . . . . . . . . 41
4.19 Privacy bubble results: dataset 2 . . . . . . . . . . . . . . . . . . . . . 42

5.1 Varying-depth visual bubble . . . . . . . . . . . . . . . . . . . . . . . . . 44

# List of Tables

4.1 Mean and standard deviation of the disparity errors for dataset 3 . . . . 36
4.2 Precision, recall, $F_1$ score and smoothness of depth bubble masks . . . . 38
4.3 Precision, recall, $F_1$ score and smoothness of visual bubble masks . . . . 42

# Chapter 1

# Introduction

The increasing computation power of small embedded platforms and affordable camera sensors have enabled many new and diverse applications ranging from entertainment, security and law enforcement to healthcare.

Some of these applications have significant privacy concerns. For example, in the past year, there have been strong calls for U.S. law enforcement officials to wear body cameras and record their interactions with the general public [1–3]. The police worn body cameras are intended to enhance police transparency and accountability, and they can also protect police officers from false complaints. Such videos, if shared, could offer a wealth of information to social scientists, journalists, and others. By now, body cameras have been adopted by the majority of the law enforcement officials in the UK and the US [4,5]. According to a nationwide survey conducted by the Major Cities Chiefs Association and Major County Sheriffs' Association, almost every large police department plans to move forward with body-worn cameras [5]. Among these police departments, 95 percent have either committed to body cameras or completed their implementation. However, privacy concerns have been raised because of the lack of guideline of the use for the cameras or the footage. The most important concern is in the avoidance of recording sensitive locations and situations. While it is entirely acceptable to record physical encounters between the police and individuals during law-enforcement related activities, sensitive background environments like bathrooms, clinics, and schools should not be recorded [6]. Such concerns have already begun to impede the popularity of body worn cameras. For example, a bill was introduced in Minnesota to prohibit the use of police body camearas for one year [7].

Another example is video recording the behaviors of special-need children, especially their interactions with others in naturalistic environments like schools and homes. They are highly valuable for diagnosis and treatment of various developmental disorders including autism and ADHD [8]. With the popularity of smartphone cameras and wearable cameras, videos can be recorded in almost any environment, capturing important intermittent behaviors that are difficult to observe during a brief clinical visit. By sharing such videos, they have become an effective tool to facilitate communication between families and professionals [9, 10]. However, their usages are governed by a myriad of privacy laws including HIPAA [11] and FERPA [12] in the US. Consent from bystanders is often difficult, if even possible, to obtain. Many studies have found that privacy is among the top concerns when setting up cameras in home and at school and sharing such videos online [13–16].

As a result, visual privacy protection has garnered a great deal of attention in the last few years. A recent survey paper has provided a comprehensive overview of different visual privacy protection technologies [17]. Most existing visual privacy protection schemes rely on intelligent classifiers to identify sensitive information such as faces or entire persons for protection. However, many of these classifiers are of questionable reliability. Furthermore, these techniques require additional selection mechanisms to differentiate target subjects, whose behaviors need to be recorded, from others whose privacy needs to be protected [18]. Any misidentification of target subjects can defeat the entire purpose of privacy protection.

The main problem considered in this thesis is how to design robust privacy protection schemes for wearable cameras to protect unintended bystanders and private environments. We propose a wearable privacy-enhanced camera that can be mounted on an adult observer for video recording. The preliminary design is shown in Figure 1.1. The novel contribution of our proposed system is its use of a “visual bubble” for privacy protection; the “visual bubble” defines an adjustable virtual zone around the camera for recording. A key advantage of visual bubble is that it does not rely on any human detector which can be unreliable. Instead, it uses pixel-based depth measurement which can be estimated with high enough fidelity for privacy protection. The popular Kinect 2 camera by Microsoft provides a very low-cost solution for such an application. Using a Kinect camera, we can easily demonstrate a visual bubble by selectively applying obfuscation on the color pixel based on its depth value. An example is shown in Figure 1.2.

On the other hand, the Kinect camera is not portable and does not work well in outdoor environments. Among all depth sensing technologies including time-of-flight, structured-light and stereo, the stereo camera provides the most versatile form of recording - it can cover a long distance and be used under a wide range of illumination from bright outdoor sunlight to dim indoor light. Also, its reliance on simple color cameras implies the highest resolution, the smallest size, and the lowest cost. The downside is that stereo cameras are not as accurate as other depth sensing technologies [19]. Thus, it is imperative to address the accuracy issues of stereo vision systems in order to use it for privacy protection. In this thesis, we propose

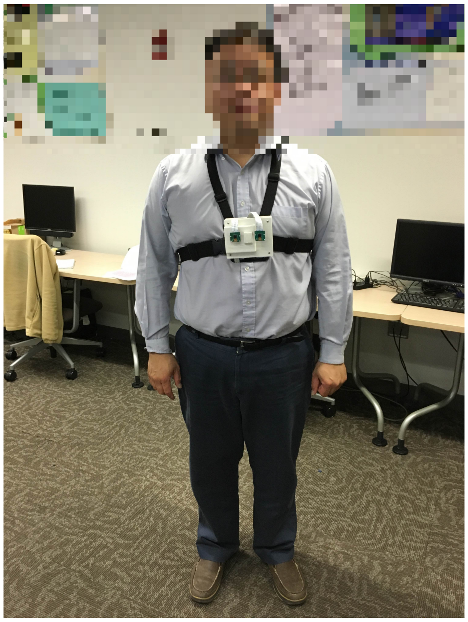


Figure 1.1: Wearable privacy camera

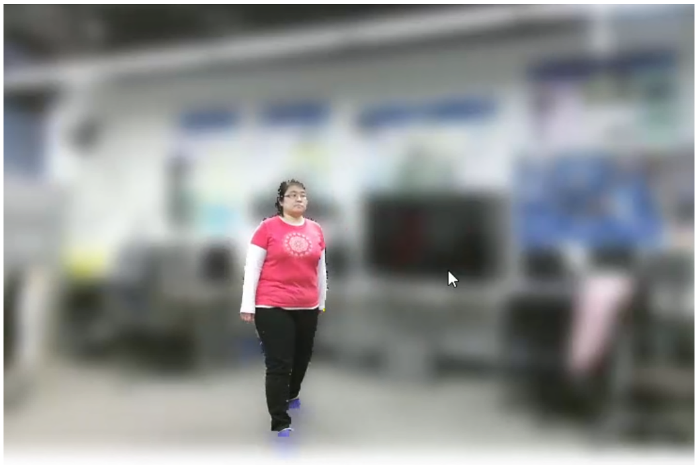


Figure 1.2: Privacy bubble implemented with Kinect 2 camera

an embedded design of privacy-enhanced wearable stereo cameras using embedded cameras on the popular Raspberry Pi platform [20]. In our design, the depth measurement is based on disparity estimated by a stereo matching algorithm and we have systematically compared different state-of-the-art stereo matching algorithms for our target application. Furthermore, we propose a statistical framework to quantify the uncertainty of the depth measurement and create the visual bubble by minimizing a statistical privacy risk so as to satisfy the more conservative requirement of privacy protection.

The rest of the thesis is organized as follows. In Chapter 2, we review related work about visual privacy protection. Additionally, traditional and some state-of-the-art stereo matching algorithms will be reviewed. We propose a distance based privacy protection technique in Chapter 3. We also develop the framework of analyzing the uncertainty in stereo-depth measurement and describe the privacy bubble system based on the probability framework. We present our hardware implementation and experimental results in Chapter 4. Chapter 5 concludes the thesis and discusses future work.

# Chapter 2

# Related Work

With the pervasiveness of surveillance and smartphone cameras, visual privacy has attracted much attention in recent years [17]. In this chapter, visual privacy protection techniques will be briefly reviewed. Since the visual privacy protection we are going to propose is based on distance estimated from stereo matching, traditional and some state-of-the-art stereo matching algorithms will also be reviewed.

## 2.1 Review of visual privacy protection techniques

Visual privacy protection means protecting the confidentiality of the private information in images and videos from being revealed to the general public. Private information or region of interest to protect varies based on different situations. It can be a person's face, facial expression, gait, credit card number, computer screen or printed document and so on. Visual privacy protection involves filtering out or scrambling the sensitive information; as with general privacy protection problems, visual privacy protection needs to strike a balance between privacy protection and utility. For example, if all of an image, not only the sensitive part, is scrambled, the image is of no utility any more; however, if no privacy related processing is done, private information will be revealed.

For visual privacy protection, there are techniques to prevent the private information from being captured in the first place. For example, the BlindSpots system designed by Patel *et. al.* is able to locate the retro-reflective CCD or CMOS camera lenses around a protected area and emits a pulsing light toward the detected lens, spoiling the images that may be captured [21].

After the image or the video is captured, visual privacy protection mainly consists of identifying the private information (or region of interest) and secure processing to hide the private information. The method of detecting sensitive regions can be (1) object-recognition based: e.g., facial recognition algorithms can be used to identify individuals; (2) visual-marker based: e.g. [22] makes use of colored markers, such as hats or vests, to identify the individuals who wish to remain anonymous; (3) gesture-based: e.g., the invidual who wishes to be protected (or to be revealed) needs to perform a specific gesture; (4) others: e.g., RFID [23] and biometric signals [18] can be used to identify sensitive information.

Measures taken to hide private information include (1) blanking: totally removing the region of interest; (2) filtering: e.g., applying Gaussian blurring or pixelation to the region of interest; (3) encryption: regions of interest of an encrypted image cannot be viewed by persons who do not have the decryption key; (4) abstraction: sensitive image regions are replaced by bounding boxes; individuals to be protected are replaced by avatars, silhouettes or edges from an edge detector; (5) others: e.g., video inpainting is used to fill the privacy region with the background; face-deindentification can be done to change the face in a way such that the identity is consealed, but gender and facial expression are preserved. There is a vast literature about visual privacy protection; interested readers are referred to [17, 24, 25].

## 2.2 Review of stereo matching

Stereo matching is one of the earliest approaches for depth measurements [26]. In computer vison, the problem of stereo matching is well studied [26] and it continues to be an active research area because of the challenges in finding stereo correspondence, e.g., lack of texture, repetitive patterns, reflective surface and so on. There are quite a few of review papers on stereo vision disparity algorithms as well [26–32].

Stereo matching algorithms can be categorized into sparse stereo matching and dense stereo matching. Sparse stereo matching usually extracts feature patterns first and then matches the feature correspondences. It generates only a sparse disparity map with the advantages of more reliable match and less running time. However, with the development of computing hardware and the need of new applications, most of the modern stereo matching algorithms fall into the second category which is generating a dense disparity map. Dense stereo matching algorithms generally perform (subsets of) the following four steps [26]: (1) matching cost computation; (2) cost (support) aggregation; (3) disparity computation/optimization; (4) disparity refinement. Most of the algorithms take two or more already rectified stereo images as input, since rectification as a preprocessing step restricts the search range for correspondence to epipolar lines which greatly saves the computation effort.

In the step of computing matching cost, lots of different metrics have been used. Among them, the most common ones are absolute intensity difference, the squared intensity difference and the normalized cross correlation. According to different features in the second and following steps, stereo matching algorithms are classified into the local approach and the global approach.

In a local approach, also called window based approach, matching costs in a support window are aggregated to reduce the impact of noises and get a more reliable result. The support window can be a fixed size square or rectangle centered at the pixel of interest. However, the support window of fixed size fails near disparity discontinuity because the underlying assumption that pixels in the support window have constant disparity is violated. Adaptive support window is able to overcome this problem. In the third step of a local approach, usally a winner-take-all strategy is used to determine the disparity of a pixel. Finally, refinements, such as interpolation, subpixel enhancement, median filtering, and bilateral filtering, are done to improve the disparity map.

On the other hand, a global approach treats the disparity assignment problem in an energy optimization framework. The energy function to be minimized usually consists of a data term and a smoothness term. Popular optimization methods are dynamic programming, Markov random field, graph cut, belief propagation and so on.

Real time stereo matching can be achieved with the help of Graphics Processing Unit (GPU) and/or additional hardware such as field-programmable gate arrays (FPGA) and application-specific integrated circuits (ASIC). Real time stereo matching algorithms are mostly based on the local approach. On one hand, regular and simple operations such as basic filtering can be easily and efficiently implemented with computing hardware; on the other hand, parallelism can be better taken advantage of in a window based approach on a GPU. Although these algorithms can achieve real time speed, their performance usually suffers from limited computation time. Global approaches are usually not implementable in real time because of its iterative nature. Besides, the complicated optimization algorithm of a global approach also prevents it from being efficiently implemented with parallel structure.

Readers can find a list of recently proposed stereo matching algorithms and their performances from Middlebury website (`http://vision.middlebury.edu/stereo/eval3/`) and KITTI stereo vision website (`http://www.cvlibs.net/datasets/kitti/eval_stereo_flow.php?benchmark=stereo`). This thesis will look more carefully at [33] and [34] because the these two algorithms are among the top runners of the Middlebury and KITTI stereo vision score board, and their implementations are publicly available.

In [33], the authors developed a slanted plane model to jointly recover an image segmentation, a dense depth map, as well as boundary labels. Their slanted plane model algorithm is between 2 to 3 orders of magnitude faster than earlier slanted models [35,36]. Besides, their model is shown to be insensitive of choice of parameters.

In [34], Zbontar *et al.* proposed to compute the matching cost by training a convolutional neural network to learn a similarity measure on small image patches. The authors point out that the learning transfers well in the sense that the neural network can be trained using one dataset and predict the results on a different dataset. As an example, the validation error on KITTI 2012 is even lower when using Middlebury traning set than when using KITTI 2015 training set, although the KITTI 2015 dataset is closer to the KITTI 2012 dataset.

# Chapter 3

## Visual Bubble System

Up to now, most work on visual privacy protection is done for single cameras. The drawback of the aforementioned approaches of detecting sensitive regions in Chapter 2 is their reliance on image segmentation and subject and/or gesture identification algorithms that may not be reliable enough for privacy protection. However, for some applications concerning privacy, for example, police worn body cameras and behaviorial/educational observation of children with special need, subject identification is not crucial.

This thesis proposes a distance based visual bubble for privacy protection. This approach relies on stereo cameras as the capturing device. To the author's best knowledge, this is a novel approach for visual privacy protection in spite of its simplicity.

Visual bubble can be a depth bubble which is purely distance based, or a privacy bubble which also considers the uncertainty of the distance/depth measurement. In this chapter, we will first describe the depth bubble. We use stereo matching to measure distance, or depth. In a privacy-concerned application, the reliability of the depth measurement should be considered. Therefore, we develop a probability framework for quantifying the uncertainty of depth measurement and describe the privacy bubble system based on the framework. Finally, we describe the superpixel technique that is used as a denoising post-processing in generating the visual bubble. The functional block diagram of the visual bubble system is shown in Fig. 3.1.

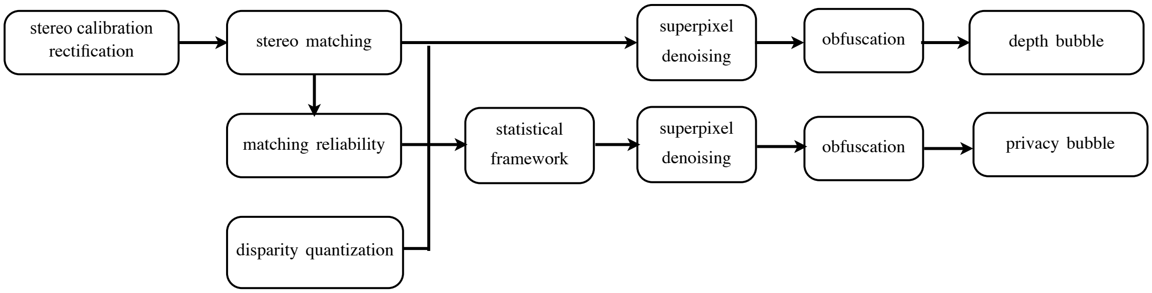


Figure 3.1: Visual bubble system

## 3.1 Depth bubble and depth from stereo

Depth bubble is a distance based visual privacy protection technique. It only shows pixels within a prescribed depth range from the camera, and whatever falls outside of the range will be filtered out. The effect of the depth bubble is like the spotlight on a dark stage. The advantages of this approach are that pixel-based depth measurement from stereo image pairs can be estimated with high enough fidelity for privacy protection and the cost is low enough for the general public. This approach relies on the assumption that the subject of interest is usually the person closest to the observer and therefore, falls within the bubble.

The creation of depth bubble depends on estimating the depth $Z$ for each pixel, which is inversely proportional to its disparity value $d$, given the camera focal length $f$ and the stereo baseline $B$, that is,

$$Z = \frac{fB}{d}. \tag{3.1}$$

Therefore, determining the disparity from stereo matching is a key step.

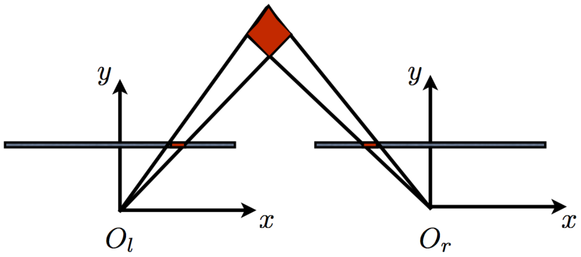


Figure 3.2: Quantization effect

## 3.2 Uncertainty in stereo depth measurement

Although stereo matching algorithms keep improving, a pixel's disparity, or its depth $z$ estimation, inevitably involves errors and measurement uncertainty. For a stereo matching algorithm which yields integer disparity values, the uncertainty of the depth estimate mainly comes from two sources: (1) the uncertainty in the stereo matching process due to matching ambiguity; (2) the uncertainty of disparity value itself due to the digital nature of the imaging system. The quantization effect is illustrated in Fig. 3.2. Assume the two red blocks on the image planes correspond to the image pixel, any spacial point in the red region forms the same images on the two image planes and has the same disparity. In other words, a pixel's disparity error is partly because of stereo correspondence mismatch due to the defect of the stereo matching algorithm; even if the disparity value from the stereo matching algorithm is correct, the disparity value also suffers from quantization. Therefore, this chapter is mainly dedicated to proposing a framework to quantify the uncertainty in depth measurement and developing an uncertainty-aware privacy bubble.

Based on the uncertainty sources of the depth measurement, our goal in this section is to characterize the conditional probability density function (pdf) $f(z|d)$

(the probability density function of depth $z$ given measured disparity $d$) in order to determine how reliable the depth estimate is. We model $f(z|d)$ based on its relationship with two other pdf's: $f(z|d_k)$ and $P(d_k|d)$ where $d_k$ with $k = 0, 1, 2, ...$ represents the ideal but unknown disparity, quantized due to the discrete nature of the system. Using Bayes' rule, these three pdf's are related by the following relationship:

$$f(z|d) = \sum_k f(z|d_k)P(d_k|d). \tag{3.2}$$

### 3.2.1 Uncertainty due to quantization

Quantization error in stereo imaging system is analyzed in [37, 38]. This thesis follows [37].

For standard stereo pinhole camera setup where the two camera image planes are coplanar as shown in Fig. 3.3, assume $f$ is the focal length of both cameras, $B$ is the baseline, $\delta$ is the image sampling interval. Assume a spatial point with world coordinate $(X, Y, Z)$ forms images on both image planes at $P_L$ and $P_R$, whose x-coordiates are $x_L$ and $x_R$, respectively.

Due to the discrete nature of the imaging system, $x_L$ and $x_R$ suffer quantization error up to $\pm 1/2\delta$. With slight abuse of notation, let the random variable $\bar{x}_L$ and $\bar{x}_R$ denote the unquantized x-coordiates of $P_L$ and $P_R$. Define the random variable $\bar{d}$ and $z$ as follows,

$$\begin{aligned} \bar{d} &:= \bar{x}_L - \bar{x}_R, \\ z &:= \frac{fB}{\bar{d}}. \end{aligned} \tag{3.3}$$

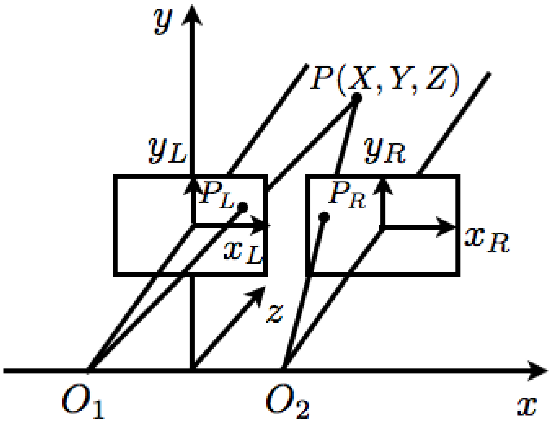


Figure 3.3: Stereo imaging model

Now assume $\bar{x}_L$ and $\bar{x}_R$ are independent and uniformly distributed. Given the quantized image coordinates $x_L$ and $x_R$, their conditional pdf's are provided as follows:

$$f(\bar{x}_L|x_L) = 1/\delta, \quad \text{for} \quad x_L - \frac{\delta}{2} \leq \bar{x}_L \leq x_L + \frac{\delta}{2},$$
$$f(\bar{x}_R|x_R) = 1/\delta, \quad \text{for} \quad x_R - \frac{\delta}{2} \leq \bar{x}_R \leq x_R + \frac{\delta}{2}.$$

Note that $d$ is only available to us as a quantized value. As such, for consecutive quantized disparity values $d_k := k\delta$, where $k = 0, 1, 2, \cdots, m$, $\bar{d}$ are confined in the interval $[d_{k-1}, d_{k+1}]$, with a triangular-shaped pdf,

$$f(\bar{d}|d_k) = \begin{cases} \frac{1}{\delta^2}(\bar{d} - d_k) + \frac{1}{\delta}, & \text{for} \quad d_{k-1} \leq \bar{d} \leq d_k, \\ -\frac{1}{\delta^2}(\bar{d} - d_k) + \frac{1}{\delta}, & \text{for} \quad d_k < \bar{d} \leq d_{k+1}. \end{cases} \tag{3.4}$$

Let

$$z_k := \frac{fB}{d_k}, \tag{3.5}$$

where $k = 1, 2, \cdots, m$.

It follows from (3.3), the conditional pdf of $z$ given $d_k$ is

$$f_Z(z|d_k) = \frac{f_{\bar{D}}(\bar{d}|d_k)}{fB/\bar{d}^2}\Big|_{\bar{d}=\frac{fB}{z}}.$$

Substituting in (3.4), we obtain

$$f_Z(z|d_k) = \begin{cases} (\frac{1}{\delta^2}(\frac{fB}{z} - d_k) + \frac{1}{\delta})\frac{fB}{z^2}, & \text{for } z_k \leq z \leq z_{k-1}, \\ (-\frac{1}{\delta^2}(\frac{fB}{z} - d_k) + \frac{1}{\delta})\frac{fB}{z^2}, & \text{for } z_{k+1} \leq z < z_k. \end{cases} \tag{3.6}$$

Since the real depth $z$ is confined within the range $[z_{k+1},\ z_{k-1}]$, we could use the length $\Delta_k$ of this interval to quantify the uncertainty of true depth:

$$\Delta_k := z_{k-1} - z_{k+1} = \frac{2}{\frac{fB}{z_k^2\delta} - \frac{\delta}{fB}}. \tag{3.7}$$

Note that the farther the point is from the camera, the bigger $\Delta_k$ is and the more uncertain its true depth becomes. Also, a smaller baseline $B$ means bigger depth uncertainty. This is important to the design of a wearable stereo camera as the baseline is highly constrained due to its compact size.

Now we have obtained the uncertainty of the depth measurement given the quantized true disparity. The disparity value is “true” in the sense that we have assumed perfect stereo matching in producing the disparity value $d$.

### 3.2.2 Uncertainty from stereo correspondence ambiguity

In a practical stereo matching system, false matches often occur due to varying illumination, lack of texture of the scene, reflected surface and camera distortion, etc. The uncertainty of the stereo matching process is modeled by $P(d_k|d)$, which is the conditional probability of the quantized disparity $d_k$ corresponding to the perfect disparity, given the measured disparity value $d$ obtained from the stereo matching algorithm. One way to model $P(d_k|d)$ is to assume $d_k$ takes values centered at $d$ with variance $\sigma^2$ proportional to the stereo matching cost.

The approach to estimate matching cost largely depends on the specific stereo matching algorithm itself. Some stereo matching algorithms will give stereo matching cost map as well as the disparity map. Simplistic error functions [39] and signal-to-noise ratio [40] were proposed to model the matching cost. More recent works estimate the matching cost using different machine learning techniques, ranging from linear discriminant analysis [41] to random forest [42] and convolutional neural network [34].

In the next chapter, we show one approach to estimate $P(d_k|d)$ for the popular semi-global matching algorithm [43]. As the approach is specific to the implementation of the algorithm, we defer the description to the next chapter.

### 3.3 Uncertainty aware privacy bubble

In this section, we show how the privacy bubble is generated using the estimated $f(z|d)$. In our target application, the subject that needs to be recorded is close to the wearable camera while we want to protect the privacy of the rest of the environment. Therefore, we could rely on the depth map and its uncertainty to segment the foreground subject and generate a privacy bubble by obfuscating other pixels. Assume we would like to generate a privacy bubble around the foreground subject within the depth of $z_p$. In order to generate the privacy bubble, we need to decide whether a pixel with depth $z$ should be shown or obfuscated. While the true $z$ is unknown, we have a measurement of disparity $d$. The conditional probability of the event $z < z_p$

given $d$ can be numerically computed as follows:

$$P(z < z_p|d) = \int_{z_{min}}^{z_p} f(z|d)\mathrm{d}z. \tag{3.8}$$

To determine whether this pixel should be revealed, we rely on the following likelihood test:

$$\frac{P(z < z_p|d)}{1 - P(z < z_p|d)} > S, \tag{3.9}$$

where $S > 0$ is the privacy protection threshold. If (3.9) is satisfied, the pixel is shown. Otherwise, it is obfuscated. The choice of threshold $S$ reflects how stringent the privacy requirement of the target application is. $S \gg 1$ will be very conservative but may wrongly obfuscate part of the subject of interest.

Now, we can apply the results from Section 3.2 to evaluate (3.8). It can be simplified with (3.2) and (3.6) as follows:

$$\begin{aligned} &P(z < z_p|d) \\ &= \int_{z_{min}}^{z_p} \sum_k f(z|d_k)P(d_k|d)\mathrm{d}z \\ &= \sum_{k=l+1}^{m} P(d_k|d) + P(d_{l-1}|d) \int_{z_l}^{z_p} f(z|d_{l-1})\mathrm{d}z \\ &\quad + P(d_l|d) \int_{z_{l+1}}^{z_p} f(z|d_l)\mathrm{d}z, \end{aligned} \tag{3.10}$$

where $z_l \leq z_p < z_{l-1}$ and $m$ is the upper bound of the disparity searching range.

## 3.4 Post processing with superpixel

The idea of superpixel was originally developed in [44]. By grouping pixels into perceptually meaningful clusters, superpixel technique has been widely used in depth

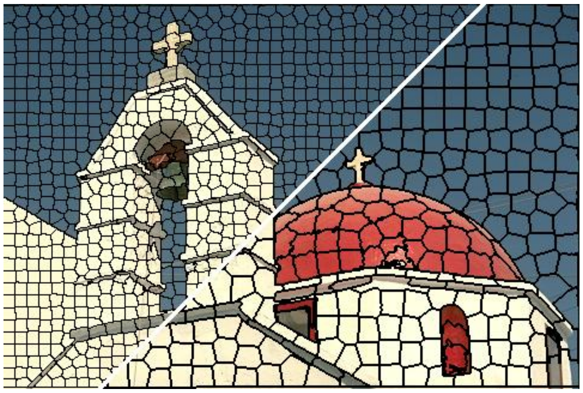


Figure 3.4: Superpixel example

estimation, segmentation, body model estimation and object localization [45]. One example of superpixel is shown in Fig. 3.4.

In our system, we use the ERS algorithm [46] to segment the input frame into $R$ superpixels, where $R$ is set large enough to ensure no semantic objects are lumped into a single segment. The superpixel segmentation is formulated as a graph partitioning problem. For a graph $G = (V, E)$ and $R$ superpixels, the goal is to find a subset of edges $A \subset E$ to approximate a graph $\bar{G} = (V, A)$ with $R$ connected sub-graphs. The vertex corresponds to a pixel in an image and an edge is formed by 4-connected neighborhood with weights computed based on color-similarity between connected vertices. The clustering objective function comprises of two terms: the entropy rate $H$ of the random walk on $A$ and a balancing term $B$:

$$\max_{A \subset E} H(A) + \lambda B(A) \text{ with } N_A \geq R \tag{3.11}$$

where $N_A$ is the number of connected components in $\bar{G}$. The entropy term encourages compact and homogeneous clusters, whereas the balancing term encourages clusters with similar size. Finally, to overcome exact optimization difficulty, a greedy algorithm with an approximation bound of $\frac{1}{2}$ is used to solve the problem.

Once the superpixels are computed, we use them to guide a simple denoising procedure of the input matrix $X$, where $X$ is the disparity map $d$ in depth bubble and the probability map $P(z < z_p|d)$ in privacy bubble. In the denoising step, we first compute the average of the inputs over each superpixel and then replace the individual input value at each pixel with the average of the superpixel to which the pixel belongs. This is based on the assumption that color-consistent neighboring pixels are more likely to have similar disparity or probability values. This assumption holds for small neighborhoods and greatly improves the accuracy of disparity along color edges. It is, however, important to set the number of superpixels $R$ large enough so that a large spatial object, which could have a large range of disparity values, will not be accidentally clustered into a single superpixel.

In summary, this chapter proposed a distance based privacy protection technique called visual bubble. Visual bubble can be either depth bubble or uncertainty aware privacy bubble. Depth bubble can be generated by thresholding the depth map with specified bubble range and depth is measured based on stereo matching. Furthermore, this chapter developed a framework to quantifying the uncertainty of depth measurement from stereo matching. Built on this framework, an uncertainty-aware privacy bubble was developed and finally the superpixel technique was proposed to be used in visual bubble post-processing.

## Chapter 4

# Hardware Implementation and Experiment

In this chapter, firstly, we describe the hardware implementation for the wearable visual bubble system. Secondly, we present the three datasets (image sequences) we have collected for evaluation purposes. The first two datasets are captured by our system. The third dataset is collected from Microsoft Kinect 2 RGB-D camera sensors and therefore, there is depth information enabling quantitative evaluation. Thirdly, we propose using semi-global (block) matching with superpixel (SGBM + SUP) to generate the depth bubble, which is evaluated and compared with the classic semi-global (block) matching (SGBM) [43] and other state-of-the-art stereo matching algorithms. Additionally, we compare the depth bubble with uncertainty-aware privacy bubble to illustrate the effectiveness of the uncertainty framework and our proposed privacy bubble scheme. Some of the earlier results can be seen in [47].

## 4.1 Hardware implementation

We have built the wearable visual bubble system using Raspberry Pi Compute Module (RPCM). The block diagram of the system is shown in Fig. 4.1. The wearable system consists of the RPCM, an I/O board, two Pi cameras and a WiFi dongle providing networking capability, as seen in Fig. 4.2. The RPCM is a small outline dual in-line memory module (SODIMM) sized (6.5cm by 3cm) Raspberry Pi board that contains the BCM2835 chip with 512MB RAM along with an onboard 4GB

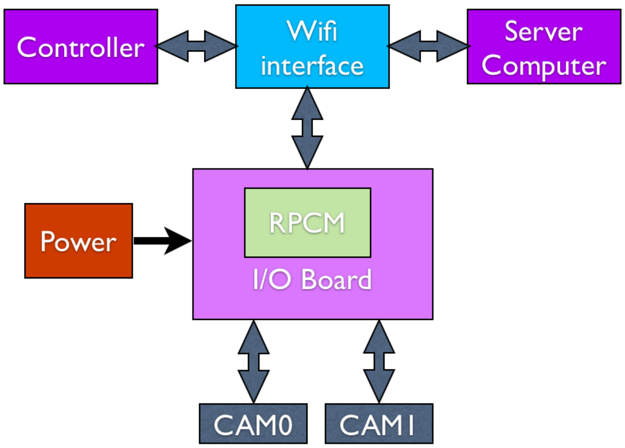


Figure 4.1: System Diagram

eMMC Flash memory. Integrated into the BCM2835 chip are a 700 MHz single-core ARM1176JZF-S CPU and a 250MHz Broadcom VideoCore IV GPU. The Pi camera has an image sensor with native resolution of 5 megapixel and is capable of capturing 2592 × 1944 static images and 1080p30, 720p60 and 480p60/90 videos. In our implementation, we will use the the resolution of 640 × 480. The onboard power supply provides 5V DC power boosted from a 3.7V 2500mAh Li-Po battery. The wireless interface features an Edimax 150 Mbps 802.11n WiFi USB adapter.

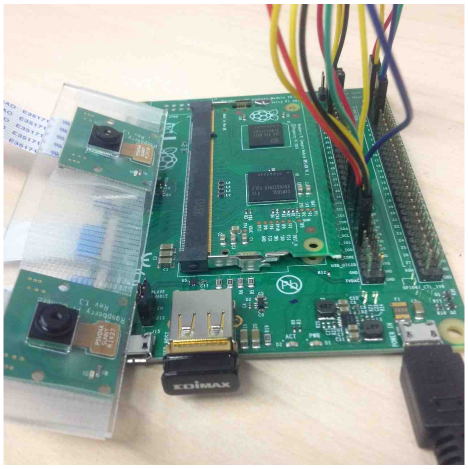


Figure 4.2: System components

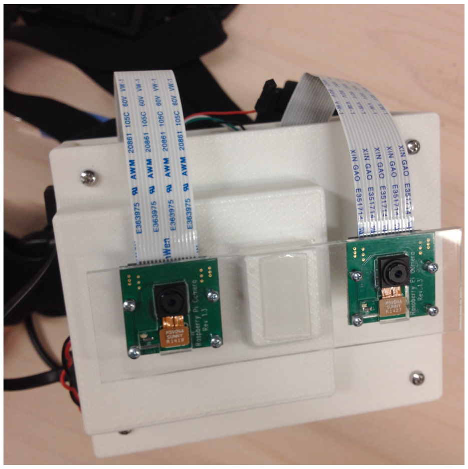


Figure 4.3: Hardware implementation

The prototype, housed in a 3D printed case and mounted on a chest strap harness, is shown in Fig. 4.3. While the current prototype is quite large (11.7cm by 9.7cm by 6cm), using a customized PC Board instead of the RPCM I/O board from the Raspberry Pi development kit would make the system much smaller. The stereo vision system can be controlled by a smart phone via an SSH connection.

The CAD model of the camera mount is shown in Fig. 4.4. The STL model of the case is shown in Fig. 4.5.

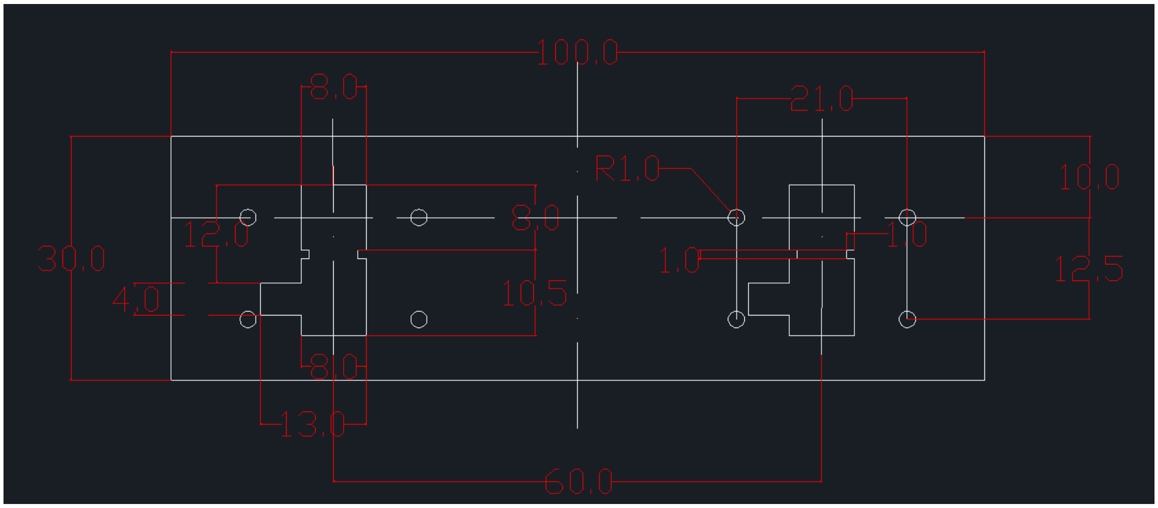


Figure 4.4: Pi camera mount CAD model

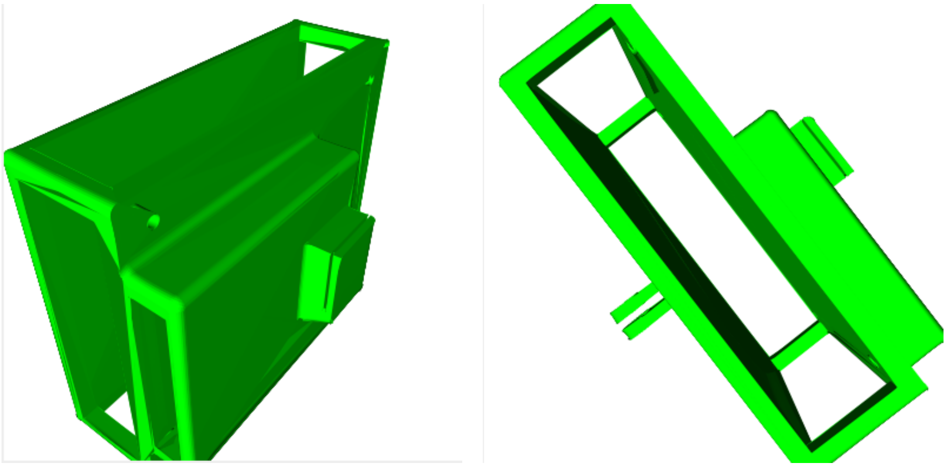


Figure 4.5: System case STL model

## 4.2 Dataset collection

We have collected 2 datasets using the system described in Section 4.1. No ground truth depth images are captured. The image pairs are then rectified using extrinsic parameters of the stereo camera acquired from camera calibration. Dataset 1 contains 76 pairs of stereo images (5-second video with 15 fps) in the outdoor environment. Sample images are shown in Fig. 4.6. Note that only the left images are shown.

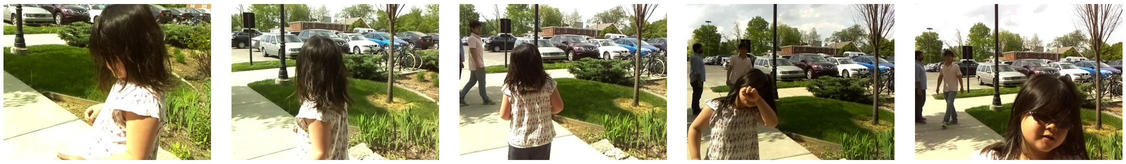


Figure 4.6: Dataset 1 sample (left) images

Dataset 2 contains 120 pairs of stereo images (8-second video with 15 fps) in the indoor environment. Sample images are shown in Fig. 4.7.

In order to evaluate our proposed algorithms quantitatively, datasets with available ground truth depth are required. Currently, there are some publicly available stereo datasets with ground truth disparity. The most notable ones are Middle-

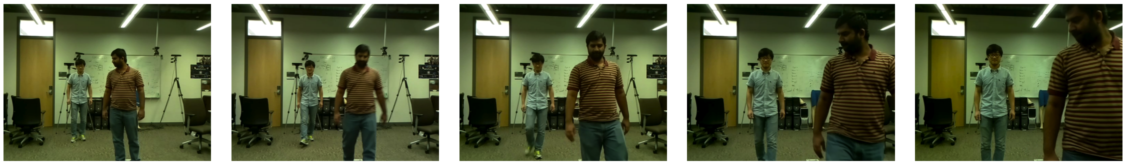


Figure 4.7: Dataset 2 sample (left) images

bury Stereo dataset (`http://vision.middlebury.edu/stereo/data/`) and KITTI stereo vision benchmark dataset (`http://www.cvlibs.net/datasets/kitti/eval_stereo.php`).

For a privacy protection scheme designed to be used in a privacy-concerned situation, it is better to be aware of the weakness and reliability of the protection technique. For our case, we should take into consideration the uncertainty coming from the stereo matching algorithms. This is especially true for wearable stereo cameras, because the small baseline of wearable stereo cameras will enlarge the uncertainty caused by disparity quantization. Since the aforementioned public datasets were captured by stereo cameras with much larger baselines (for example, the stereo camera baselines for Middlebury 2014 stereo dataset are around 150-240 mm while our baseline is only 60 mm), using these datasets will not be able to most effectively demonstrate the importance of the uncertainty analysis; therefore, we will generate our own dataset (with ground truth depth) to test the proposed privacy protection technique for our specific wearable camera system. Besides, it is always good to test a system in a real situation where it shall be put into use.

### 4.2.1 Dataset collection using Kinect 2

First, set up a pair of kinect 2 cameras as in Fig. 4.8 after calibrating each of them to align the color sensor and depth sensor. The color sensor image planes are set to be roughly coplaner. The basedline is set to about 60 mm, mimicking the baseline of the Pi-stereo system. We have captured 100 frames of color and depth images from both kinects simutaneously.

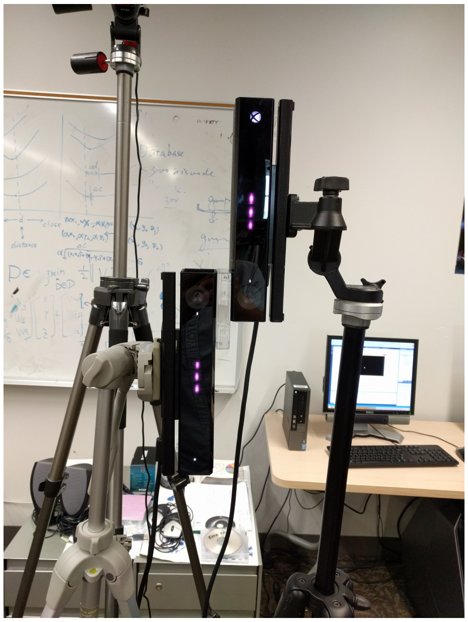


Figure 4.8: Stereo kinect setup

For Kinect 2, the resolutions for the captured color image and depth image cannot be modified. The resolutions are 1920 × 1080 for color images and 512 × 424 for depth images by default. Since the color images and the depth images have different resolution, we need to call the `MapColorFrameToDepthSpace()` function in order to generate a depth image whose resolution matches that of the color image. Then color images and depth images are rotated and resized to mimic the images capututred by Raspberry Pi cameras with 640 × 480 resolution. Finally, Bouguet's stereo rectification algorithm with opencv implementation is used to rectify the stereo color image pairs and depth images pairs. Sample color images (left image of the stereo pair) are shown in Fig. 4.9.

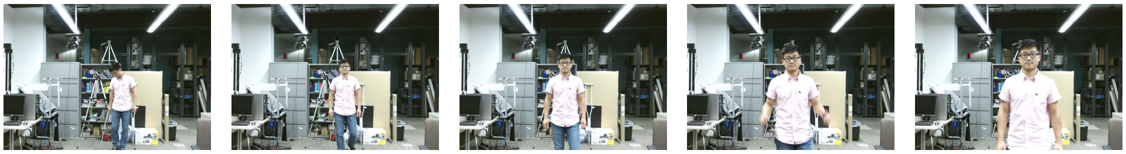


Figure 4.9: Dataset 3 sample (left) images

### 4.2.2 Ground truth depth calculation

Stereo rectification as the preprocessing step for stereo matching can greatly reduce the work for searching for stereo correspondence because corresponding pixels are warped to the same scan line of the images, i.e., they lie on the same row on the image pairs. The effect of rectification is illustrated in Fig. 4.10. Rectified image pairs can be treated as captured by a standard vitual stereo rig, as if in Fig. 3.3 in Chapter 3. Since for the virtual rig, the camera orientations are different from the original kinect depth sensor orientation, we need to calculate the ground truth depth for the rectified color images.

Assume pinhole camera model. Let the world frame coincide with the camera frame of the camera 1, it follows that for camera 1,

$$s \begin{bmatrix} u \\ v \\ w \end{bmatrix} = M_1 \left[ I \quad | \quad O \right] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{4.1}$$

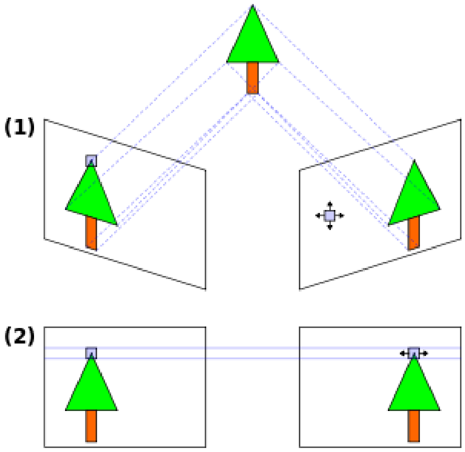


Figure 4.10: Stereo rectification

where $s$ is the scaling factor, $[u \quad v \quad w]^T$ is the homogenous pixel coordinates, and

$$M_1 := \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{4.2}$$

is the camera matrix for color camera 1.

For any spatial point $P$ on the original left image corresponding to pixel coordinates $(i, j)$, which is the pixel at the j-th row and i-th column, we would like to calculate its coordinates in the world frame. It follows that

$$s \begin{bmatrix} i \\ j \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \tag{4.3}$$

Thus,

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = s \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} i \\ j \\ 1 \end{bmatrix} = s \begin{bmatrix} \frac{i-c_x}{f_x} \\ \frac{j-c_y}{f_y} \\ 1 \end{bmatrix}. \tag{4.4}$$

Therefore, $s = Z$, $X = \frac{Z}{f_x}(i - c_x)$ and $Y = \frac{Z}{f_y}(j - c_y)$.

From the stereo rectification implementation of opencv reading as

```
stereoRectify(cameraMatrix[0], distCoeffs[0],
          cameraMatrix[1],  distCoeffs[1],
          imageSize, R, T, R1, R2, P1, P2, Q,
          CALIB_ZERO_DISPARITY, 0, imageSize,
          &validRoi[0], &validRoi[1]);
```

$R_1$ is the rectification transform (rotation matrix) for the first camera, denoted as

$$R_1 := \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}.$$

Assume in the rectified camera 1 frame, the coordinates of $P$ are $P_l' = \begin{bmatrix} X' & Y' & Z' \end{bmatrix}^T$, it follows that

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = R_1 \begin{bmatrix} \frac{Z}{f_x}(i - c_x) \\ \frac{Z}{f_y}(j - c_y) \\ Z \end{bmatrix}. \tag{4.5}$$

Thus,

$$Z' = Z(\frac{r_{31}}{f_x}(i - c_x) + \frac{r_{32}}{f_y}(j - c_y) + r_{33}). \tag{4.6}$$

Let $f'$ be the camera focal length and $B$ be the baseline of the virtual rig. The ground truth disparity of point P in the virtual rig is

$$d = \frac{f'B}{Z'} = \frac{f'B}{Z(\frac{r_{31}}{f_x}(i - c_x) + \frac{r_{32}}{f_y}(j - c_y) + r_{33})}. \tag{4.7}$$

Note that $f'$ is directly available in the $P^1$ matrix ($P^1$ is the projection matrix for the left image used in rectification, which is `P1` in the above code snippet), $f' = P^1_{11}$. Baseline $B$ can be calculated from $P^1$ matrix as

$$B = \frac{\|P^1_{14}\|}{P^1_{11}}, \tag{4.8}$$

where $\| \cdot \|$ denotes the absolute value of $(\cdot)$, $P^1_{ij}$ denotes the entry of matrix $P^1$ on the i-th row and j-th colomn.

Sample depth images corresponding to the images in Fig. 4.9 are shown in Fig. 4.11. The valid depth is assumed in the range of [500, 4500] (unit: mm) and the maximum value is set to be 5000 mm. The depth images are saved in `CV_8U` data format.

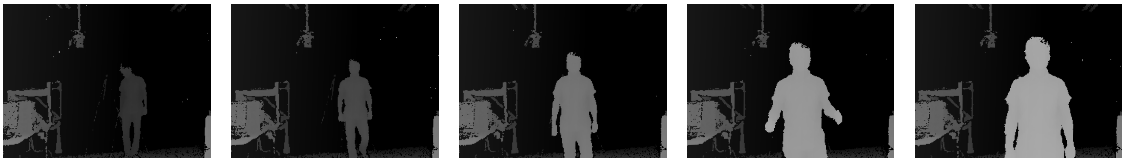
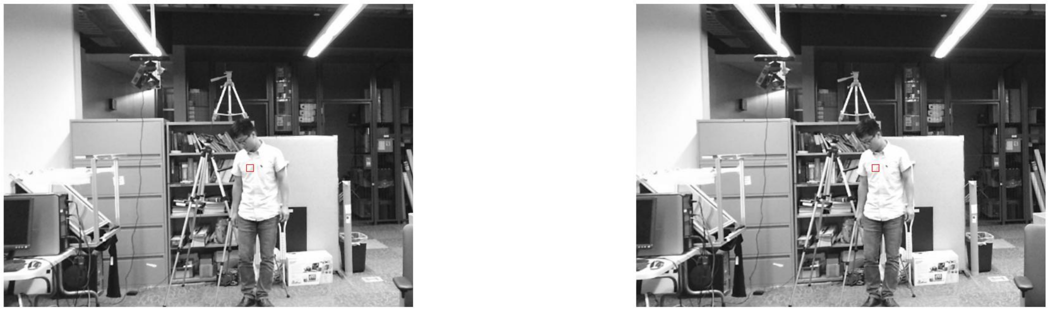


Figure 4.11: Dataset 3 sample depth images

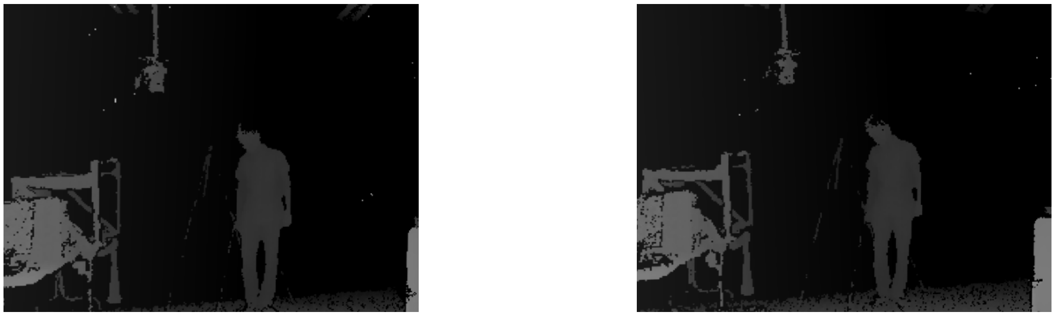
### 4.2.3 Validation of the ground truth depth

Method in the last subsection provides the ground truth depth, and therefore, ground truth disparity for the left image of the color image pairs. Since we can do the same thing and get the ground truth depth for the right image, the ground truth depth image pairs can be compared to check the quality of the depth image.

Shown in Fig. 4.12 are a pair of grey image pairs (converted from color images) and the corresponding depth images. A 11 x 11 pixel template is chosen in the left image, marked by a red box in the image. Then template matching (using grey images as input; therefore, grey images are shown in the figure) is done, the correspondence template is found in the right image and verifed by inspection. Now depth values from the two views at the same position marked with the red box are compared to evaluted the accuracy of the depth measurement. Depth values from 5 image pairs and 5 templates on each image pair (totally 25 templates) are used, that is, 25 × 121 depth value pairs. The average difference of the depth pairs is calculated to be 13.8324 mm, and the standard deviation of the depth difference of the depth pairs is 9.5411 mm. Since the depth resolution for depth data encoded in the `CV_8U` format is 19.6 mm (0-5000 mm depth encoded as 255-0), the depth images acquired from last subsection are accurate enough to be treated as ground truth depth.



Left image Right image



Left depth image Right depth image

Figure 4.12: Depth validation by comparing depth of matching templates

## 4.3 Experimental results

In our wearable visual bubble system, the stereo baseline $B = 6$ cm, the focal length of the Pi camera $f = 3.60$ mm, and image sampling interval $\delta = 6$ $\mu$m. With the wearable system, we have created two datasets, one for outdoor environment and one for indoor environment.

For depth bubble application, we propose, first, use semi-global block matching to generate disparity map; second, use superpixel technique as post-processing; and finally threshold the disparity map to generate the depth bubble. In order to evaluate our proposed algorithm, we compare it with 3 different stereo matching algorithms (the common last step is thresholding the disparity map to generate depth bubble) : (1) semi-global stereo matching (SGBM, without post-processing with superpixel); (2) slanted plane model based segmentation algorithm (SPSEG) [33]; (3) convolutional neural network (CNN) based learning algorithm [34].

For uncertainty aware privacy bubble, we propose, first, calculate the disparity map using semi-global stereo matching; second, estimate the matching reliability and generate the probability map; third, use superpixel to clean the probability map; and finally, threshold the probability map to generate privacy bubble. We compare results from this algorithm with our depth bubble algorithm, and show that it is beneficial to consider uncertainty in the stereo-matching.

In order to compare the generated bubble image (both depth bubble and privacy bubble) quantitatively, we compare the bubble mask image (recall that the mask image is a binary matrix, with 1 denoting that the corresponding pixel should be exposed, and 0 denoting that the corresponding pixel should be filtered out) with the ground truth bubble mask image. The metrics we have used are precision ($P$), recall ($R$), $F_1$ score and smoothness ($S$). Let $TP$ = True positive, $FP$ = False positive, $FN$ = False negative. (Note that, $TP$ is also called hit; $FP$ is also called false alarm, which is the Type I error; $FN$ is also called miss, which is the Type II error.)

Precision, recall and $F_1$ score are defined as:

$$P = \frac{TP}{TP+FP}, \quad R = \frac{TP}{TP+FN}, \quad F_1 = \frac{2P \cdot R}{P+R}.$$

$F_1$ score is the harmonic mean of precision and recall and it is commonly used in statistical analysis of binary classification as a measure of a test's accuracy.

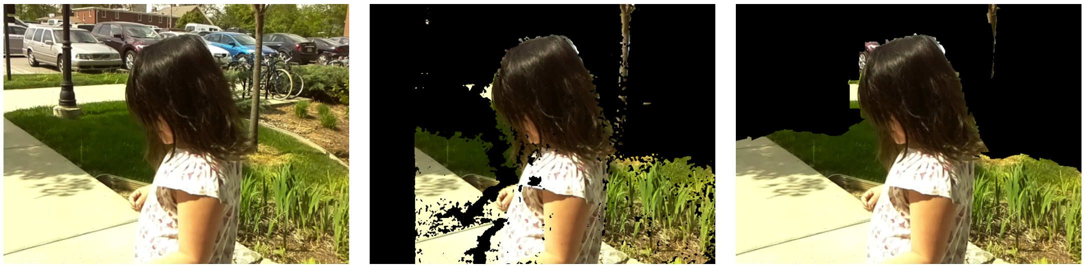
We have calculated the precision, recall and $F_1$ score for each frame and computed the arithmetic mean of precision, recall and $F_1$ score of 41 frames in dataset 3 in which frames a subject is in the visual bubble. Besides, we have also counted the number of 8-connected components with size larger than 10 pixels as a quantitative results for mask smoothness (S).

### 4.3.1 Depth bubble results

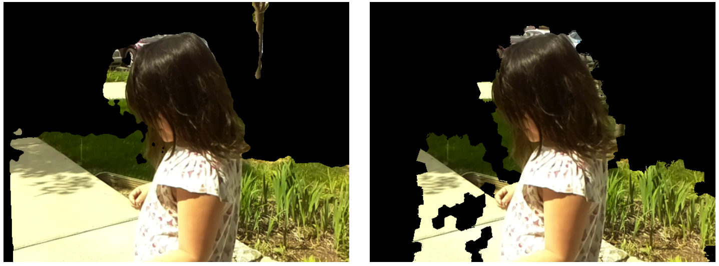
The depth bubble experimental results for two frames from dataset 1 are shown in Fig. 4.13. The specified bubble depth range is 3.2 m.

The depth bubble experimental results for dataset 2 are shown in Fig. 4.14. The specified bubble depth range is 3 m. In this set of results, the bubble images using SPSEG and SGBM+SUP are good. The depth bubble using CNN algorithm exposes lots of pixels which should not be shown, although it gives very smooth bubble boundary.

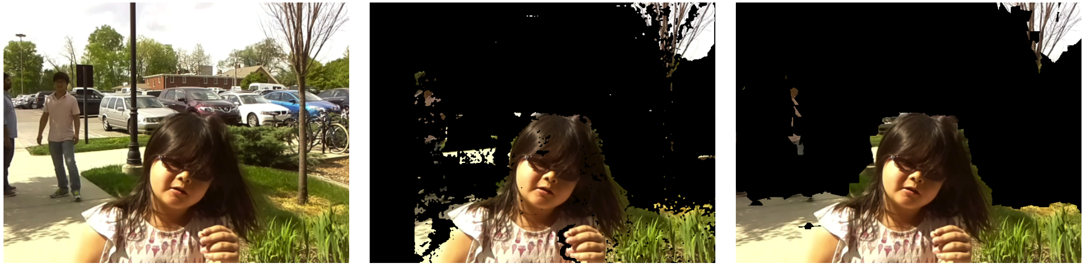
For dataset 3, since ground truth disparity is available, we can compare the disparity map generated from all the algorithms with the ground truth disparity. This comparison is not based on a specific bubble depth; therefore, it indicates the overall performance of these stereo matching algorithms.



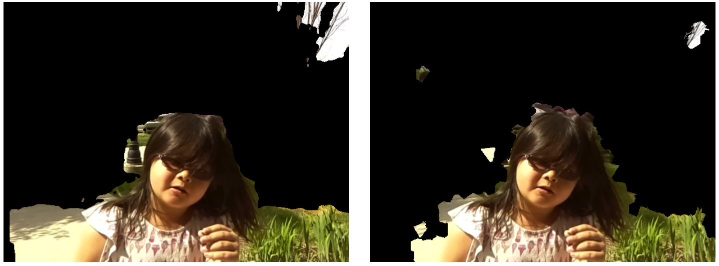
Original image 1 SGBM SPSEG



CNN SGBM+SUP



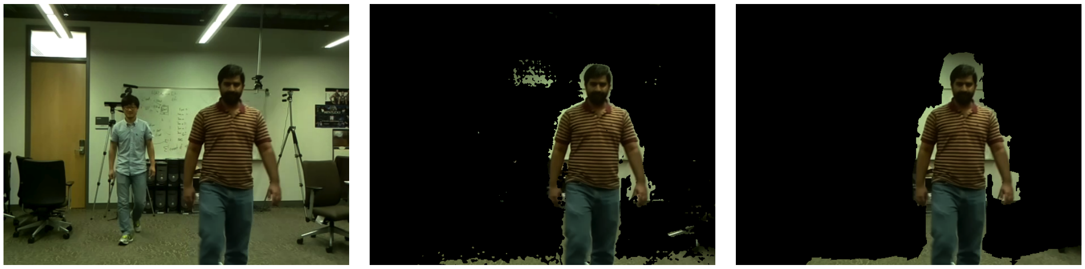
Original image 2 SGBM SPSEG



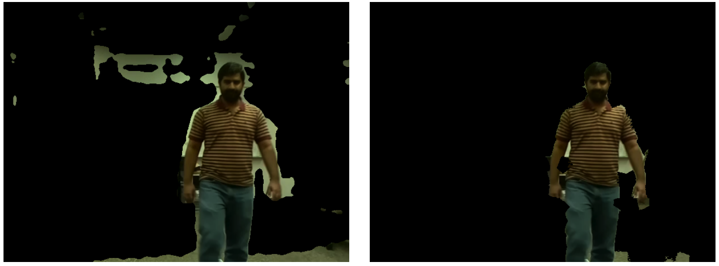
CNN SGBM+SUP

Figure 4.13: Depth bubble results: dataset 1

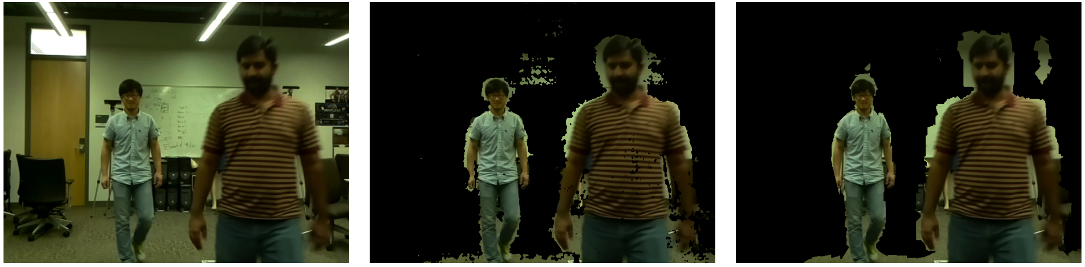
For each algorithm, we have computed the disparity for every frame in dataset 3. Disparity error is calculated as the difference between the computed disparity with the ground truth disparity. We then take the disparity errors for all frames and get the histogram of the disparity errors, as shown in Fig. 4.15. We also compute the mean and standard deviation of the disparity errors, as shown in Table. 4.1. From



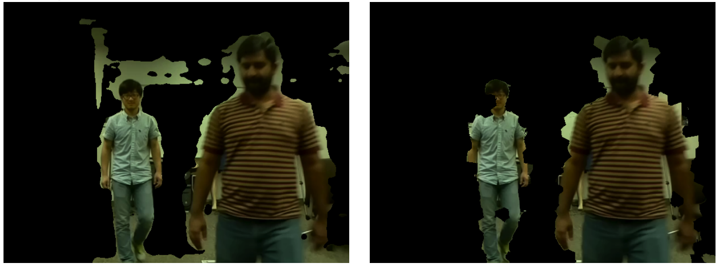
Original image 1 SGBM SPSEG



CNN SGBM+SEG



Original image 2 SGBM SPSEG



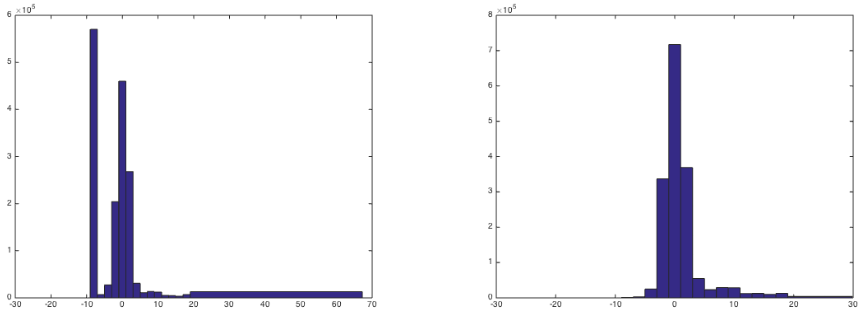
CNN SGBM+SUP

Figure 4.14: Depth bubble: dataset 2

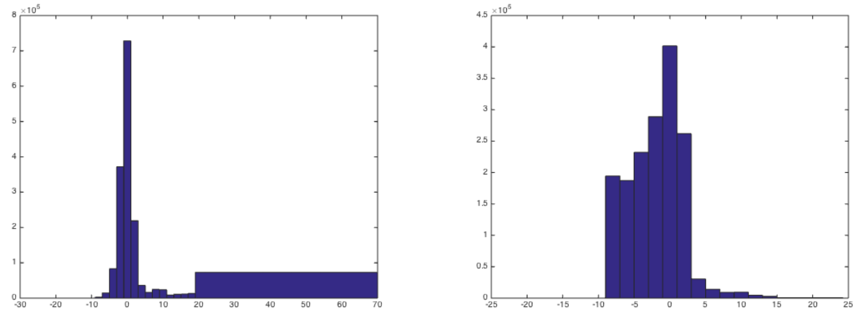
comparison, SPSEG and SBGM+SUP perform better.

The depth bubble experimental results for dataset 3 are shown in Fig. 4.16. The specified bubble depth range is 2.5 m. Quantitative results using precision, recall, $F_1$ score and smoothness as metrics are provided in Table 4.2.

As we can see from the qualitative and quantitative results, the SGBM algorithm



SGBM SPSEG



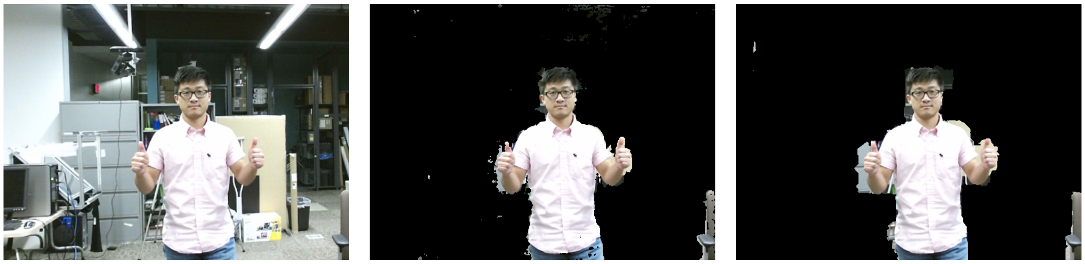
CNN SGBM+SUP

Figure 4.15: Histogram of disparity errors for dataset 3

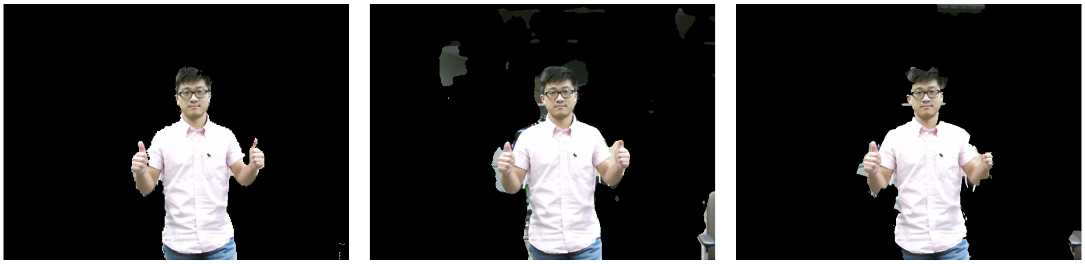
Table 4.1: Mean and standard deviation of the disparity errors for dataset 3

| Algorithm | mean (unit: pixel) | standard deviation (unit: pixel) |
|---|---|---|
| SGBM [43] | -2.5517 | 5.8845 |
| SPSEG [33] | 0.9699 | 3.5119 |
| CNN [34] | 1.9131 | 8.2517 |
| SGBM+SUP | -2.2258 | 3.8110 |

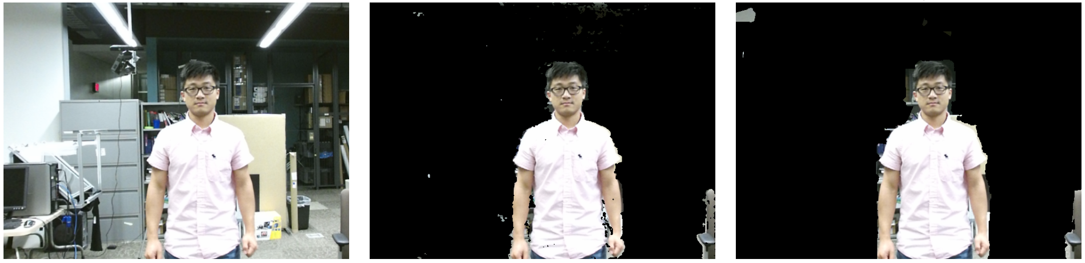
performs well, but the biggest problem is that the generated bubble image contains lots of 8-connected components ('small holes') rendering the bubble not clean nor smooth. This motivates our proposed depth bubble algorithm, which is based on SGBM but uses superpixel to clean up the bubble image. The slanted plane model based segmentation algorithm preserves connectedness of image segments very well and generates clean bubble images. The CNN based learning algorithm, although



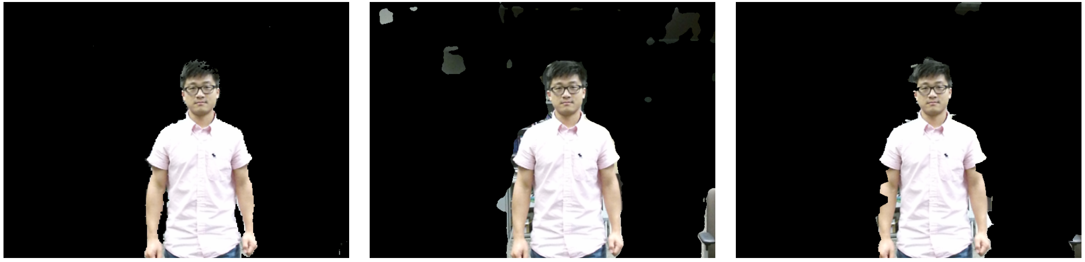
Original image 1 SGBM SPSEG



Ground Truth Bubble CNN SGBM+SUP



Original image 2 SGBM SPSEG



Ground Truth Bubble CNN SGBM+SUP

Figure 4.16: Depth bubble: dataset 3

shows the highest recall rate, but also has the lowest precision. This can be easily seen from the qualitative results where lots of pixels are falsely exposed, which may cause a serious problem as far as privacy is concerned. Although CNN based learning algorithm is among the top runners of stereo matching algorithms on the score board of Middlebury stereo website, and it claims to have good performance in the transfer

Table 4.2: Precision, recall, $F_1$ score and smoothness of depth bubble masks

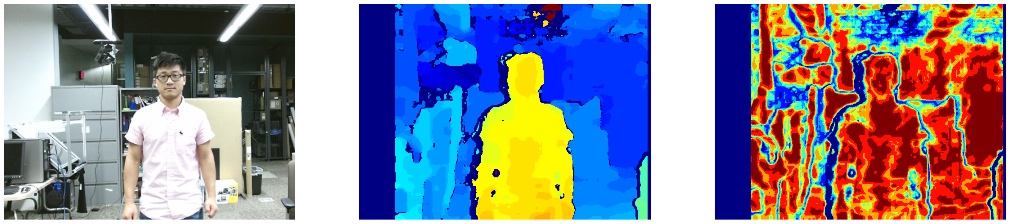
| Algorithm | $P$ | $R$ | $F_1$ | S |
|---|---|---|---|---|
| SGBM [43] | 0.8556 | 0.9539 | 0.8992 | 26.0488 |
| SPSEG [33] | 0.8225 | 0.9750 | 0.8896 | 3.5854 |
| CNN [34] | 0.7295 | 0.9786 | 0.8343 | 12.4390 |
| SGBM+SUP | 0.9100 | 0.9034 | 0.9019 | 3.5366 |

learning settings where the validation error is computed on a different data set than the one used for training, it has been shown in our experiment that its precision is the lowest among the four algorithms we are comparing. The reason is that the neural network we have used was trained using the Middlebury dataset which is greatly different from the dataset we use for testing. Overall, our proposed SGBM with superpixel algorithm has the highest $F_1$ score and the smallest average number of connected components.

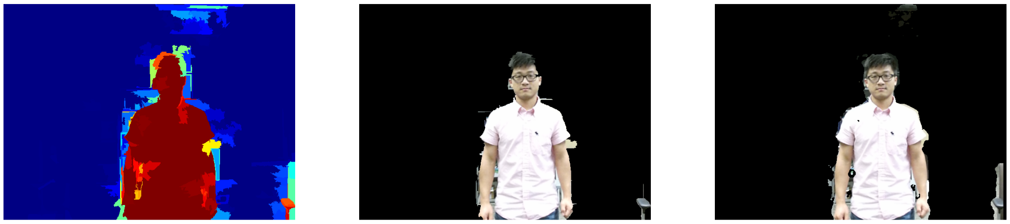
### 4.3.2 Privacy bubble results

As seen in Chapter 3, our privacy bubble generation scheme depends on the disparity uncertainty, or reliability estimate for a stereo matching algorithm. The estimation approach will mostly depend on the specific stereo matching algorithm. In this subsection, we first illustrate how we estimate the uncertainty of the disparity map and demonstrate a privacy bubble with $z_p = 3.6$m. Here, we choose privacy protection threshold $S = 4$. Through experimental results, we then show the benefit of adding uncertainty analysis in generating the visual bubble.

Using the Matlab implementation of the semi-global block matching algorithm [43], the parameter '`UniquenessThreshold`' indicates the uniqueness of a correspon-



(a) Original image (b) Disparity map (c) reliabilityMap



(d) Probability map (e) Privacy bubble (f) Depth bubble

Figure 4.17: Privacy bubble generation

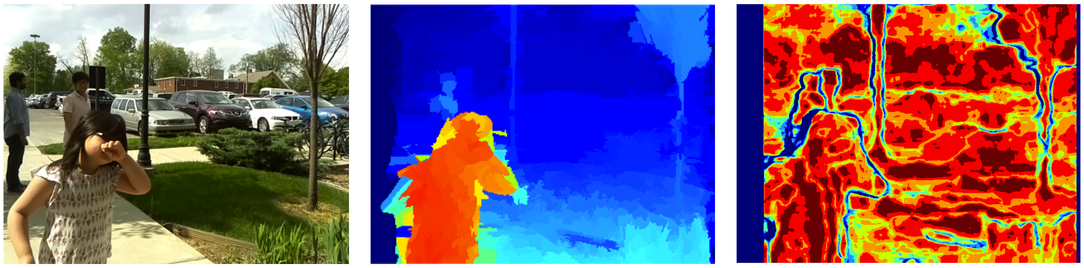
dence match. If the second smallest sum of absolute difference (SAD) value over the whole disparity range is not larger than the smallest SAD by the extent specified by this parameter, the estimated disparity will be marked as unreliable. We observed that when ‘`UniquenessThreshold`’ is set to 100, all of the stereo match will be labeled as unreliable. As such, we have run a series of tests by varying ‘`UniquenessThreshold`’ from 0 to 90 with a stepsize of 15. By counting how many times the computed disparity value is label as ‘reliable’, we could quantify the reliability of the disparity map of each pixel into seven levels, with 7 being the most reliable and 0 being not reliable at all. Figure 4.17(c) shows the reliability of the disparity map, with the red end being the most reliable.

Next, for reliability level $k$, $1 \leq k \leq 7$, assume the true disparity falls into one of $1 + 2(7 - k)$ disparity bins and the probability mass function forms a triangle shape with the given computed disparity value in the middle. We use (3.10) to calculate the overall probability of a spatial point being within the privacy bubble. Then, the probability map is post-processed by the superpixel algorithm where superpixels are formed based on original color image. The result is shown in Figure 4.17(d). Figure 4.17(e) shows the actual privacy bubble based on our probability calculations compared with the depth bubble generated by simply thresholding the depth map as shown in Figure 4.17(f). One can see that in the depth bubble, there is a 'hole' above the main subject because of the falsely computed disparity, as seen in Fig. 4.17(b). However, the reliability of that disparity value is low; thus, corresponding pixels don't have high enough probability to enter the privacy bubble. Therefore, the 'hole' has been filtered out in the privacy bubble.

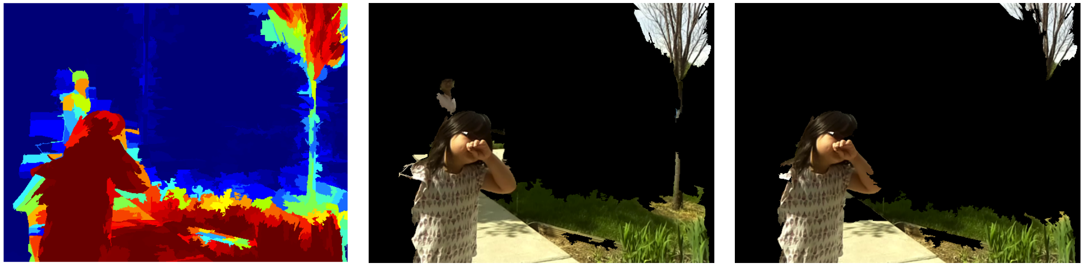
The privacy bubble experimental results for two frames from dataset 1 are shown in Fig. 4.18. The specified bubble depth range is 3.6 m and the privacy protection threshold is $S = 4$. From comparing the depth bubble and privacy bubble, it can be seen that privacy bubble with the uncertainty framework integrated can better protect the unintended bystander.

The privacy bubble experimental results for two frames from dataset 2 are shown in Fig. 4.19. The specified bubble depth range is 3 m and the privacy protection threshold is $S = 4$. It is shown in Fig. 4.19 that a threshold can be chosen to make the privacy bubble more conservative.

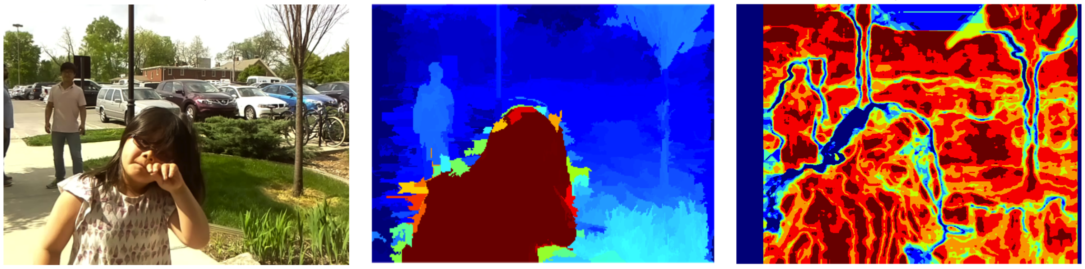
Table 4.3 provides the quantitative comparison between the 2.5 m depth bubble and 2.5 m privacy bubble using the metrics of precision, recall, $F_1$ score and smoothness. The privacy bubble is only slightly better than the depth bubble, because of the relative simple scene within the specified range.



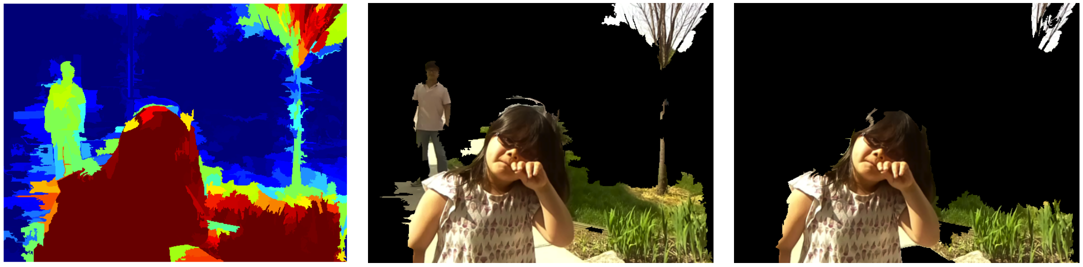
Original image 1
Disparity map
Reliability map



Probability map
Depth bubble
Privacy bubble



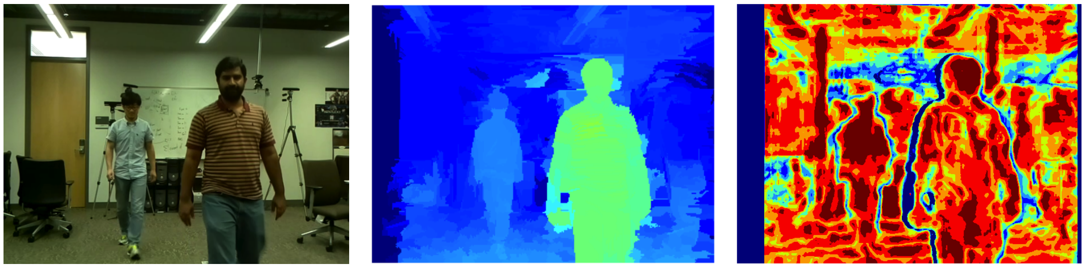
Original image 1
Disparity map
Reliability map



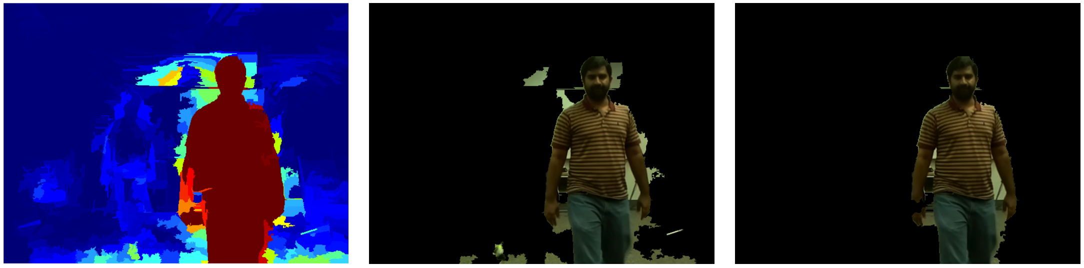
Probability map
Depth bubble
Privacy bubble

Figure 4.18: Privacy bubble results: dataset 1

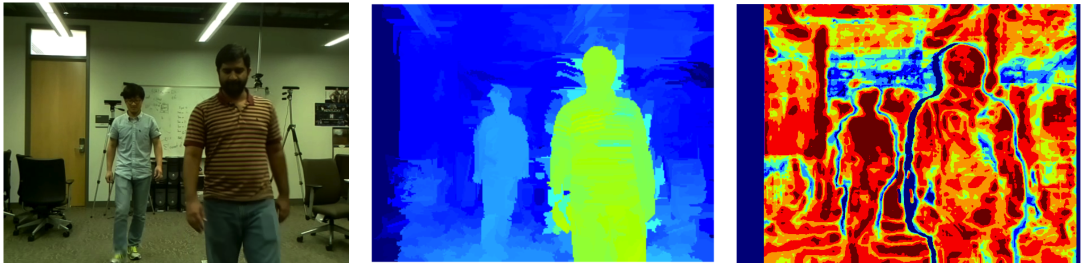
As we can see from the experimental results, the proposed visual bubble (depth bubble and privacy bubble) works equally well in both indoors and outdoors. Generally, the privacy bubble based on the probability framework outperforms the depth bubble in terms of privacy protection. More results can be found at `http://vis.uky.edu/nsf-autism/wearable-privacy-cam/`.



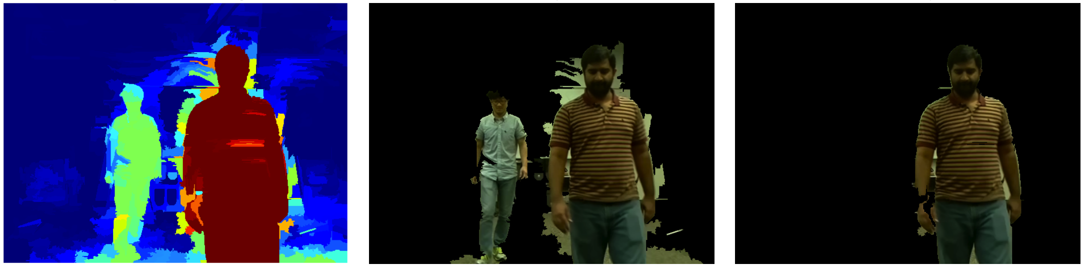
Original image 1 Disparity map Reliability map



Probability map Depth bubble Privacy bubble



Original image 1 Disparity map Reliability map



Probability map Depth bubble Privacy bubble

Figure 4.19: Privacy bubble results: dataset 2

Table 4.3: Precision, recall, $F_1$ score and smoothness of visual bubble masks

| Algorithm | $P$ | $R$ | $F_1$ | S |
|---|---|---|---|---|
| SGBM+SUP depth | 0.9432 | 0.9216 | 0.9267 | 2.6585 |
| SGBM+uncertainty privacy | 0.9303 | 0.9451 | 0.9326 | 2.2195 |

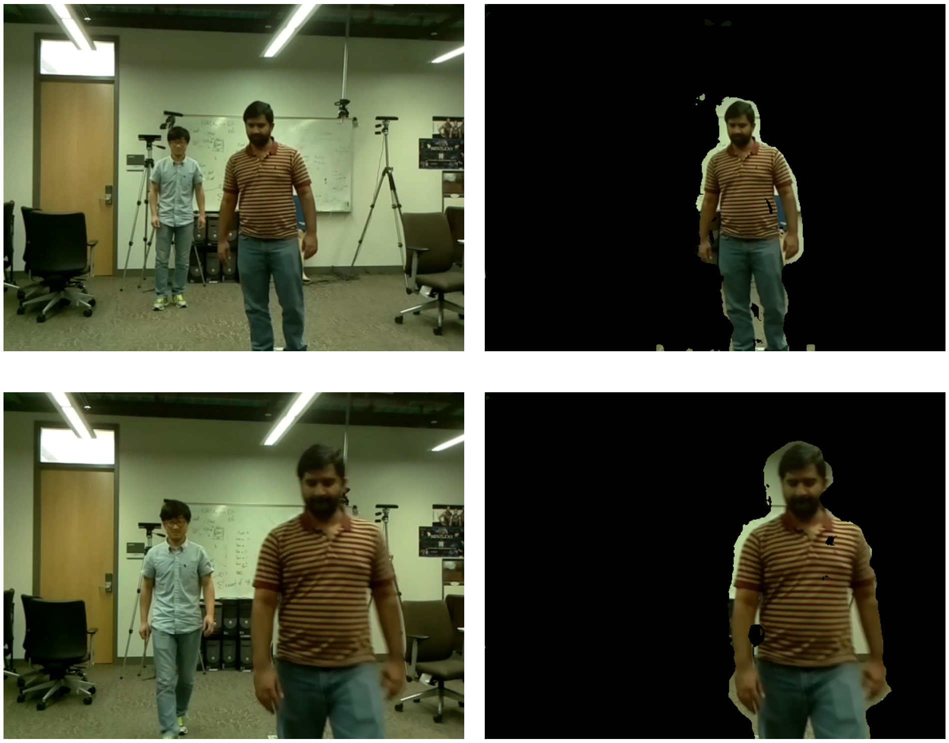
# Chapter 5

# Conclusion and Future Work

In this thesis, we have proposed a new visual privacy protection technique called visual bubble. A visual bubble can be a depth bubble which is purely distance based, or a privacy bubble which also considers the uncertainty of the distance/depth measurement. In a privacy-concerned application, the reliability of the depth measurement should be considered. To minimize the statistical privacy risk in constructing the depth-based visual bubble, stereo depth uncertainty has been considered in two aspects: uncertainty from quantization and from imperfect stereo matching. We develop a probability framework for quantifying the uncertainty of depth measurement and describe the privacy bubble system based on the framework.

An implementation of the wearable privacy camera using Raspberry Pi Compute Module has also been presented. In the software aspect, we propose using semi-global block matching with superpixel (SGBM + SUP) to generate the depth bubble, which is evaluated and compared with the classic semi-global block matching (SGBM) and other state-of-the-art stereo-matching algorithms. Besides, we have compared the depth bubble with uncertainty-aware privacy bubble. Experimental results have demonstrated the effectiveness of the uncertainty framework and our proposed privacy bubble scheme.

In addition to a visual bubble with a fixed radius, we have also experimented a privacy bubble with varying depth based on the closest individual. Figure 5.1 shows the preliminary results of determining the radius by clustering pixel depths using the $K$-means algorithm ($K = 3$) and assuming that the closest individual occupies the closest cluster. The preliminary results are reasonably good though additional work is needed to determine a more robust clustering scheme.



Original image sequences Visual bubble with varying depth

Figure 5.1: Varying-depth visual bubble

# Bibliography

[1] R. Baillon. *Mayor calls for 1,200 body cameras to be utilized by Milwaukee police officers, but how will they work?* Fox6Now.com, `http://tinyurl.com/zsx8zja`, September 3 2015.

[2] B. Keilar and D. Mercia. *Hillary Clinton calls for mandatory policy body cameras, end era of mass incarceration.* CNN.com, `http://tinyurl.com/goeuxsl`, April 29 2015.

[3] J. Stanley. *Police body-mounted cameras: With right policies in place, a win for all.* ACLU, `http://tinyurl.com/hx7y9hz`, 2013.

[4] Karson Kampfe. Police-worn body cameras: Balancing privacy and accountability through state and police department action. *Ohio St. LJ*, 76:1153, 2015.

[5] Mike Maciag. *Survey: almost all police departments plan to use body cameras.* `http://www.governing.com/topics/public-justice-safety/gov-police-body-camera-survey.html`, January 26 2016.

[6] Claire M Cochrane and Chiu C Tan. A study on the privacy and security of police body cameras deployments. In *Technologies for Homeland Security (HST), 2016 IEEE Symposium on*, pages 1–3. IEEE, 2016.

[7] Abby Simons. *Senator proposes one-year moratorium on police body cameras.* `http://www.startribune.com/`

`senator-proposes-one-year-moratorium-on-police-body-cameras/291572541/`, Febuary 11 2015.

[8] Richard Longabaugh. The systematic observation of behavior in naturalistic settings. *Handbook of cross-cultural psychology*, 2:57–126, 1980.

[9] Brooke Ingersoll and Anna Dvortcsak. Including parent training in the early childhood special education curriculum for children with autism spectrum disorders. *Journal of Positive Behavior Interventions*, 8(2):79–87, 2006.

[10] Carolyn Webster-Stratton. Advancing videotape parent training: A comparison study. *Journal of Consulting and Clinical Psychology*, 62(3):583, 1994.

[11] An Act. Health insurance portability and accountability act of 1996. *Public Law*, 104:191, 1996.

[12] Bobbye G Fry. The family educational rights and privacy act of 1974. *Student Records Management: A Handbook*, page 43, 1997.

[13] Gillian R Hayes and Khai N Truong. Selective archiving: a model for privacy sensitive capture and access technologies. In *Protecting Privacy in Video Surveillance*, pages 165–184. Springer, 2009.

[14] Julie A Kientz and Gregory D Abowd. Kidcam: toward an effective technology for the capture of childrens moments of interest. In *Pervasive Computing*, pages 115–132. Springer, 2009.

[15] Gabriela Marcu, Anind K Dey, and Sara Kiesler. Parent-driven use of wearable cameras for autism support: a field study with families. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 401–410. ACM, 2012.

[16] N Nazneen, Agata Rozga, Mario Romero, Addie J Findley, Nathan A Call, Gregory D Abowd, and Rosa I Arriaga. Supporting parents for in-home capture of problem behaviors of children with developmental disabilities. *Personal and Ubiquitous Computing*, 16(2):193–207, 2012.

[17] Thomas Winkler and Bernhard Rinner. Security and privacy protection in visual sensor networks: A survey. *ACM Computing Surveys (CSUR)*, 47(1):2, 2014.

[18] Ying Luo, Shuiming Ye, and Sen-ching S Cheung. Anonymous subject identification in privacy-aware video surveillance. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pages 83–88. IEEE, 2010.

[19] Fabrizio Pece, Jan Kautz, and Tim Weyrich. Three depth-camera technologies compared. In *Engineering in Medicine and Biology Society*, pages 1188–1193, 2012.

[20] Raspberry Pi Foundation. Teach, learn and make with Raspberry Pi. http://www.raspberrypi.org, 2015.

[21] Shwetak N Patel, Jay W Summet, and Khai N Truong. Blindspot: Creating capture-resistant spaces. In *Protecting Privacy in Video Surveillance*, pages 185–201. Springer, 2009.

[22] J. Schiff, M. Meingast, D. Mulligan, S. Sastry, and K. Goldberg. Respectful cameras: Detecting visual markers in real-time to address privacy concerns. In *International Conference on Intelligent Robots and Systems (IROS)*, pages 971–978. Springer, 2007.

[23] J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian. Privacy protecting data collection in media spaces. In *ACM International Conference on Multimedia*, New York, NY, October 2004.

[24] José Ramón Padilla-López, Alexandros Andre Chaaraoui, and Francisco Flórez-Revuelta. Visual privacy protection methods: A survey. *Expert Systems with Applications*, 42(9):4177–4195, 2015.

[25] Katharina Krombholz, Adrian Dabrowski, Matthew Smith, and Edgar Weippl. Ok glass, leave me alone: Towards a systematization of privacy enhancing technologies for wearable computing. In *International Conference on Financial Cryptography and Data Security*, pages 274–280. Springer, 2015.

[26] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1-3):7–42, 2002.

[27] Myron Z Brown, Darius Burschka, and Gregory D Hager. Advances in computational stereo. *IEEE Transactions on Pattern analysis and machine intelligence*, 25(8):993–1008, 2003.

[28] Federico Tombari, Stefano Mattoccia, Luigi Di Stefano, and Elisa Addimanda. Classification and evaluation of cost aggregation methods for stereo correspondence. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[29] Nalpantidis Lazaros, Georgios Christou Sirakoulis, and Antonios Gasteratos. Review of stereo vision algorithms: from software to hardware. *International Journal of Optomechatronics*, 2(4):435–462, 2008.

[30] Federico Tombari and Fabio Gori. Evaluation of stereo algorithms for 3d object recognition. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 990–997. IEEE, 2011.

[31] Beau Tippetts, Dah Jye Lee, Kirt Lillywhite, and James Archibald. Review of stereo vision algorithms and their suitability for resource-limited systems. *Journal of Real-Time Image Processing*, 11(1):5–25, 2016.

[32] Rostam Affendi Hamzah and Haidi Ibrahim. Literature survey on stereo vision disparity map algorithms. *Journal of Sensors*, 2016.

[33] Koichiro Yamaguchi, David McAllester, and Raquel Urtasun. Efficient joint segmentation, occlusion labeling, stereo and flow estimation. In *European Conference on Computer Vision*, pages 756–771. Springer, 2014.

[34] Jure Zbontar and Yann LeCun. Stereo matching by training a convolutional neural network to compare image patches. *Journal of Machine Learning Research*, 17:1–32, 2016.

[35] Christoph Vogel, Konrad Schindler, and Stefan Roth. Piecewise rigid scene flow. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1377–1384, 2013.

[36] Koichiro Yamaguchi, David McAllester, and Raquel Urtasun. Robust monocular epipolar flow estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1862–1869, 2013.

[37] Jeffrey J Rodriguez and JK Aggarwal. Quantization error in stereo imaging. In *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR'88., Computer Society Conference on*, pages 153–158. IEEE, 1988.

[38] Raman Balasubramanian, Sukhendu Das, S Udayabaskaran, and Krishnan Swaminathan. Quantization error in stereo imaging systems. *International journal of computer mathematics*, 79(6):671–691, 2002.

[39] Emanuele Trucco, Vito Roberto, S Tinonin, and M Corbatto. Ssd disparity estimation for dynamic stereo. In *BMVC*, pages 1–10, 1996.

[40] Andrea Fusiello, Vito Roberto, and Emanuele Trucco. Experiments with a new area-based stereo algorithm. In *Image Analysis and Processing*, pages 669–676. Springer, 1997.

[41] Martin Peris, Atsuto Maki, Sara Martull, Yoshihiro Ohkawa, and Kazuhiro Fukui. Towards a simulation driven stereo vision system. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 1038–1042. IEEE, 2012.

[42] Ralf Haeusler, Rahul Nair, and Daniel Kondermann. Ensemble learning for confidence measures in stereo vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 305–312, 2013.

[43] Heiko Hirschmüller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 807–814. IEEE, 2005.

[44] Xiaofeng Ren and Jitendra Malik. Learning a classification model for segmentation. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 10–17. IEEE, 2003.

[45] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.

[46] Ming-Yu Liu, Oncel Tuzel, Srikumar Ramalingam, and Rama Chellappa. Entropy rate superpixel segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2097–2104. IEEE, 2011.

[47] Shaoqian Wang, S Cheung Sen-ching, and Ying Luo. Wearable privacy protection with visual bubble. In *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*, pages 1–6. IEEE, 2016.

# Vita

Name: Shaoqian Wang

Bachelor of Science in Mechanical Engineering

Harbin Institute of Technology, Harbin, China

Place of birth: Shijiazhuang, Hebei, China