Vision Based Fall Detection and Localization on Construction Sites

Bingfei Zhang

A Thesis in

The Department of

Building, Civil and Environmental Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of

Master of Applied Science (in Building Engineering) at

Concordia University

Montreal, Quebec, Canada

September 2017

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Bingfei Zhang

Entitled: Vision Based Fall Detection and Localization on Construction Sites

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science in Building Engineering

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Chair
Dr. Osama Moselhi

_____ Examiner
Dr. Sang Hyeok Han

_____ Examiner (External)
Dr. Amin Hammad

_____ Supervisor
Dr. Zhenhua Zhu

Approved by _____
Chair of Department

_____
Dean,

Date _____

ABSTRACT

The fall accident is one of the major causes of fatal injuries and financial losses in the construction industry. Currently many methods have been studied to detect fall accidents on the construction sites to avoid fall injuries and reduce financial losses. The state-of-art fall detection methods include wearable sensor based fall detection, ambience based fall detection and vision based fall detection. However, these methods still have limitations in terms of accuracy and detection speed, when being used to detect and locate fall accidents on the construction sites in practice.

The main objective of this research is to propose a novel vision based framework to detect and locate fall accidents on the construction sites promptly and automatically. In order to achieve the main objective, three methods (worker localization, worker matching, and fall detection) are created under the proposed framework. The worker localization method acquires real-world map coordinates from video frames based on the perspective transformation. The worker matching method matches workers captured by different camera views based on their spatial relationship according to the construction sites. The fall detection method employs an artificial neural network. The neural network is trained with features extracted from videos to detect fall accidents automatically.

Experiments have been conducted both in lab and on real construction sites to test the performances of the methods under the proposed framework. The experiment results indicated that the average localization accuracy was 90%. The accuracy is similar to the

previous works; however, no attached sensors or tags are required with the proposed method. The matching accuracy was 93.01%. Compared with the method proposed by Lee et al. (2016), the proposed method is more accurate when cameras are set far from workers. The fall detection had an 83% precision and a 90% recall rate. The accuracy is similar to the previous works; however, the proposed method does not require subtle vision features of workers.

The main contribution of this research study is proposing a framework providing information about fall accidents on the construction sites promptly and automatically. Also, the methods created in this research study can be used to assist other automated construction processes including tracking, motion detection, etc. Future works will focus on improving the localization accuracy, matching workers under ultra-large baseline camera networks and implementing deep neural networks for fall detection.

ACKNOWLEDGEMENT

Foremost, I would like to express my sincere gratitude to my supervisor Dr. Zhenhua Zhu for his continuous support of my study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of the research and the writing of this thesis.

Besides my supervisor, my sincere thanks also go to my friends and colleagues Mr. Xiaoning Ren, Mr. Bo Xiao, Ms. Chen Chen and Ms. Wenjing Chu. They have helped, supported, truly criticized and corrected me in my research.

In addition, I would like to thank my examiners, for their time, advice and effort in reviewing my thesis.

Also, I would like to thank my parents and girlfriend for their endless support throughout my journey of graduate study.

Lastly, I would like to thank the GreenOwl Mobile and Mitacs for their direct and indirect financial support on this research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

This research seeks to demonstrate that the techniques in the area of visual tracking, detection, stereo vision and machine learning can be used to detect and locate worker's fall accidents on the construction sites from videos. Information about detected fall accidents can be further used by construction managers to identify unsafe areas on the construction sites. The following sections in this chapter introduce the research background, motivation, objectives, methodology, contributions, and the organization of this dissertation.

1.1 Background and Motivation

The construction industry is generally considered as one of the most dangerous industries (Kartam et al., 2000). Workers are exposed to the hazardous working condition and taking the risk of accidents like falls, equipment collisions and structure collapses when doing construction works. Those construction accidents lead to catastrophic results to the worker health and safety as well as the construction work progress. According to the Association of Workers' Compensation Boards of Canada (AWCBC), the construction accident was the third biggest cause of lost time claims and the biggest cause of fatalities in all industries (AWCBC, 2015). The data collected by AWCBC (Table 1.1) showed that the construction accident contributed about 25% of fatal accidents in all industries in from 2013 to 2015. Therefore, there is an urgent need to improve the construction safety.

Year	Construction	Total
2013	221	902
2014	232	919
2015	186	852
Total	639	2673

Table 1.1 Fatal occupational injuries of the construction industry (AWCBC, 2015)

The fall accidents are considered as one of the most common hazards and the major causes of serious injuries among different kinds of industrial accidents. Reported by AWCBC, from 2013 to 2015, 29.1% fatal construction accidents were fall accidents (Table 1.2) (AWCBC, 2015). In addition to deaths and injuries, the fall accidents also lead to tremendous financial losses. It was estimated that the unintentional fall accidents cost almost $80 million each year in U.S., combining public and private construction projects (Kendzior, 2010). In the construction industry, the fall accidents happen frequently and have become one of the leading causes of serious work-related fatalities and injuries, as well as time and financial losses (Simeonov et al. 2010; Kaskutas et al. 2010).

Year	Construction fatal falls	Construction fatalities from other events	Total
2013	65	156	221
2014	63	169	232
2015	58	128	186
Total	186	453	639

Table 1.2 Fatal workplace injuries in construction from falls and other events (AWCBC, 2015)

In order to avoid the fatal injuries and reduce the financial losses incurred due to fall accidents on the construction sites, it is important to detect fall accidents timely and promptly to handle them appropriately. The fall detection helps in two ways. First, the fall detection can contribute to avoiding the secondary damages caused by the delay of rescue. If a fall accident is detected promptly, the rescue actions for the victims of the fall accident could be initiated immediately, and the corresponding emergency response could be arranged in a timely manner. Second, the fall detection can provide plenty of data overlooked by humans for safety management. For example, it is common that workers experience stumbles or fall accidents on the construction sites but are not hurt. They may overlook those experiences and may not report them. Thus the safety managers are not aware of the existence of the potential risks until a disastrous fall accident finally happens. With the fall detection, all the stumbles and falls are detected and reported automatically, and then corresponding actions can be taken to eliminate the potential risks based on the collected data.

So far, one common category of fall detection methods mainly relies on the use of the wearable sensors, such as accelerometers, barometric pressure sensors, gyroscopic sensors or combination of them (Makantasis et al., 2015). However, it is always required to attach these wearable sensors on human bodies, in order to obtain useful information for the fall detection. Most construction workers are not willing to wear these sensors for the consideration of personal information privacy. Therefore, the popularity of this kind of methods is limited in practice. Another category of fall detection methods is based on the ambiance. Usually, the ambiance sensors are used to collect audio or vibration data when humans are close to the sensor (Mubashir et al., 2013). However, as the areas of construction sites are usually huge, it is hard to detect fall accidents with ambiance sensors on construction sites.

One new category of research studies focuses on the fall detection with computer vision techniques. Currently, it is common to set digital cameras on construction sites to monitor the working environments, due to the fast development of digital camera technology. Utilizing those cameras for fall detection can be economical and convenient. Compared with wearable sensors and acoustic sensors, cameras can acquire data more efficiently and more accurately. In this category of research studies, video cameras are set up at height on the construction sites to capture videos. Workers are extracted from the videos, and the workers' actions are classified by computer vision techniques. Then the fall accidents can be detected the instance they happen.

Although vision based fall detection methods are outstanding, they still have limitations. First, the location data of the fall accidents is not recorded by videos. That is, if a fall accident is detected by cameras, the location of the fall accident is required to be found manually. Second, most vision based fall detection methods are based on a single camera. A single camera is not enough to handle the occlusion. It is common that workers are partially or fully occluded in the camera views because the construction sites are complex due to the great number of materials, different heavy equipment, and various structural works (Cabonari et al., 2011). With only a single camera, it is hard to detect fall accidents when the workers are partially or fully occluded. Third, most vision based methods work with horizontal cameras (e.g., Han and Lee, 2013; Konstantinou and Brilakis, 2016), while in practice most cameras are set at the height to capture the whole scene of the construction sites.

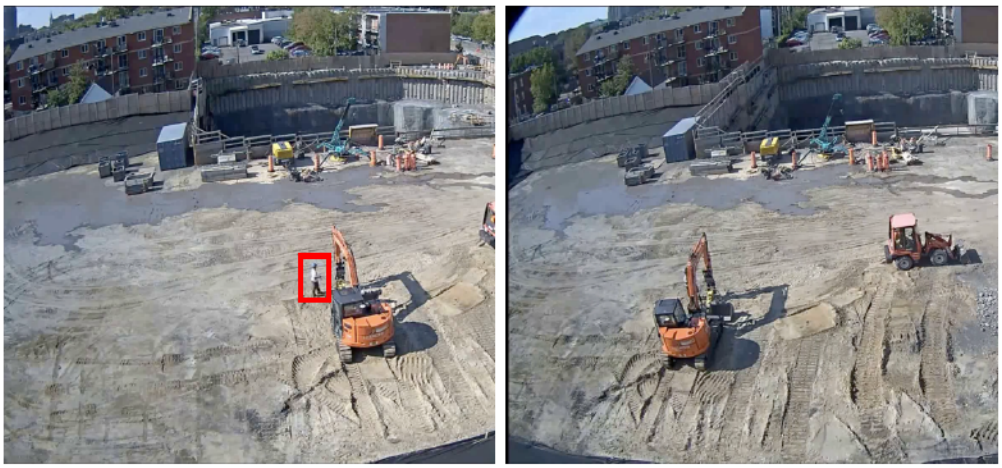

Figure 1.1 A worker occluded by an excavator

1.2 Research Objectives and Scope

The main objective of this research is to propose a framework using computer vision techniques including tracking, detection and machine learning for detecting and locating fall accidents on the construction sites. With the proposed framework, the fall accidents and nearly-fall accidents happened on the construction sites should be automatically reported to the construction manager or the safety manager.

To support the main objective, the research objective is divided into three sub-objectives specifically:

1. Create a method to locate the workers on the construction sites. The automatic worker localization method can help to find the location where fall accidents happen timely. The perspective transformation is used to retrieve the location data of workers from images. The workers are detected and tracked in the video frames with bounding boxes. The bounding box centers are projected on the map by perspective transformation to find the location of workers. The localization accuracy was tested on the real construction sites.

2. Create a method to match workers in different camera views. The matching method can help to identify the same worker in different camera views. The spatial relationship between workers in different camera views is used for matching. The potential matching candidates are searched based on the epipolar geometry constraint first, and then they are matched with combinatorial optimization. The matching accuracy was tested with

multiple workers on the real construction sites.

3. Create a method to detect fall accidents. The fall detection method can detect fall accident on the construction site automatically. The artificial neural network is chosen to identify fall accidents from other workers' routine actions. The height-width ratio of bounding boxes, the angle of bounding ellipses, the workers' actual height and their changes between frames are selected as features and extracted from video frames for identification. The artificial neural network is trained on the training set, and the accuracy of the identification is tested with the test set.

The main study object of this research is the worker on the construction sites, regardless the type of work. The scope of this research is fall detection of workers on a flat level surface (ground level, one floor). In this way, the workers' location can be described by a 2D coordinate containing the horizontal coordinates x and y, while the vertical coordinate z is not considered. Also, the algorithm for tracking and detection of construction workers are not studied in this research. In this research, mature detection and tracking methods are selected and integrated to generate the bounding boxes of workers in the videos. Those bounding boxes are used as the input of the proposed framework. In this research, all the methods proposed (i.e., localization, matching, and fall detection) are based on the assumption that all the workers are detected and tracked accurately.

1.3 Methodology and Results

In order to fulfill the mentioned research objectives, the research works in this study includes: (1) Retrieve location data of the workers based on the visual detection and tracking results. (2) Identify the same workers in different camera views. (3) Detect fall accidents using artificial neural networks. A framework was proposed to realize these three tasks.

The first part of the proposed framework is localization. The map coordinates are used to describe the workers' location on the construction sites. The perspective transformation matrix between the flat level surface and the video frames is calculated first as initialization. In the video, the workers are detected on the first frame and then tracked in the following frames. The detection and tracking results of the workers are mapped from images to the flat level surface by perspective transformation. In this way, the map coordinate of each worker in the videos is obtained.

The second part of the proposed framework is matching. A set of feature point pairs in different camera views are detected and matched at first. Then matched triangle mesh pairs are created based on the matched feature point pairs. The detected and tracked workers in different camera views then can be matched based on their relative position according to the matched triangle meshes.

The third part of the proposed framework is fall detection. Six features of workers

including the height-width ratio of the bounding box, the angle of the bounding ellipse, the actual height and their changes between frames are extracted from videos and used for fall detection. An artificial neural network is trained supervised on a training set with labels of fall or not fall. After training, the neural network can classify the actions of the workers as fall or not fall with an input of the features extracted from live videos.

The performance of each part was evaluated with proper metrics. For the first part, the construction sites were divided into small squares, and the accuracy was defined as the ratio between the number of points located in the correct squares and the total number of points. For the second part, the accuracy of the proposed matching method was defined as the ratio between the number of correct matched pairs and the total number of workers. The correct matched numbers were obtained by comparing the matching results with the manually matched ground truth. For the third part, both precision and recall rate were used to evaluate the proposed method. The precision was defined as the ratio between the correct detected falls and all detected falls, and the recall rate was defined as the ratio between the correct detected falls and all actual falls.

The test results validated the effectiveness of the proposed framework in this research. The localization accuracy is 90% The accuracy of matching is very high and can reach 93.01%. The precision of fall detection is 83%, and the recall rate is 90%. With parameter adjustments, the precision or recall rate can be even higher. All the test results confirmed the effectiveness of using computer vision and machine learning techniques for detecting

and locating fall accidents on the construction sites.

1.4 Contribution

The main goal of this research is to help the construction safety management by detecting fall accidents on the construction sites timely and automatically. The contributions of this research are listed as follows:

1. Provide the data about fall accidents without any wearable sensor or tag attached to the construction workers.
2. Reduce the waste of time and money of human monitoring by automatic localization and fall detection with video cameras.
3. Report accidents timely to avoid secondary damages caused by delayed rescue actions.
4. Detect fall accidents automatically for unsafe areas analysis.

The research work of vision based fall detection method can also be used to enhance other automated construction processes, including but not limited to:

1. Trajectory generation. The localization method can also be used for trajectory generation. The trajectories of workers or construction equipment can be generated from videos with the proposed method, then further analysis of the workers or equipment like the working efficiency analysis can be realized.
2. Object tracking. Usually, the construction sites are huge, so multiple cameras are

set up to capture the view of the whole sites. If one worker or construction equipment goes outside one camera view and enters another camera view, the tracking of the worker or equipment fails. With the proposed matching method, the worker or equipment can be matched in the overlap areas, and then tracking can be extended from one camera view to another camera view.

3. Action classification. The artificial neural network proposed in this research is not limited to fall detection. If trained with other appropriate features, the artificial neural network can also be used for classifying other workers' activities like sitting, running, lifting, etc.

1.5 Dissertation Organization

The background and motivation, objectives and scope, methodology and results, and contributions behind this research have been introduced in this chapter. The remaining chapters in the dissertation are organized as follows:

Chapter 2 is the background literature review. It first lists the current practices of localization on the construction sites, then followed by the overview of the state-of-art worker matching methods. These two works are the prerequisite works of the multi-view fall detection. After that, previous studies of fall detection methods are introduced. This chapter ends up with a summary of the limitations and issues of previous works this research going to solve.

Chapter 3 presents the methodology of this research. This research proposed a novel framework for detecting and locating fall accidents on the construction sites. The framework is separated into three main parts. The first part is the localization method for locating workers on the construction sites. The second part is the matching method for matching workers in different camera views. The third part is the fall detection method for classifying fall accidents with an artificial neural network. This chapter ends up with a summary of the proposed methods.

Chapter 4 describes the implementation and results of different parts of the proposed framework. The first and the second parts of the proposed framework are tested on the real construction sites. The third part was tested in an indoor environment simulating a real construction site. The result of each part is shown in this chapter and related factors are discussed. This chapter ends up with a brief of the contribution of each part of the proposed framework.

Chapter 5 is the conclusion of the whole research. First, the background, the methodology and the results are viewed. Then the whole research is discussed and a conclusion is made based on the test results and discussions. At the end of this chapter, future research directions about fall detection on construction sites are recommended.

CHAPTER 2: LITERATURE REVIEW

This chapter first introduces the current practice of two prerequisites of multi-view fall detection methods, localization and matching on the construction sites. It is then followed by the introduction of previous studies of different fall detection methods. The limitations and issues of existing methods are summarized at the end of this chapter.

2.1 Localization

Construction sites are well known for their complex and dynamic characteristic, thus employing Real-time Location Systems (RTLSs) to monitor the location of construction workers real-time and automatically is useful. By locating workers on the construction sites, the fall accidents victims can be found easily so that the emergency rescue work can be done promptly when accidents happen. Several methods are used for worker's localization on construction sites currently. Different kinds of RTLS are listed and introduced in the following sections.

2.1.1 Global Positioning System(GPS)

The GPS is a global navigation system operated by the U.S. Air Force that provides geolocation to GPS receivers anywhere on the earth (DoD, U.S., 2001). The GPS is already a mature system for tracking and locating on the construction sites (Park., 2012). The

system usually consists of a constellation of satellites, GPS sensors mounted on the targets and a central module for receiving real-time data from GPS sensors, as shown in Figure 2.1 (Park, 2012). The GPS sensors receive location information from satellites and send it to the central module over the network. The central module processes the information and makes it possible to visualize the location information of targets in real time.

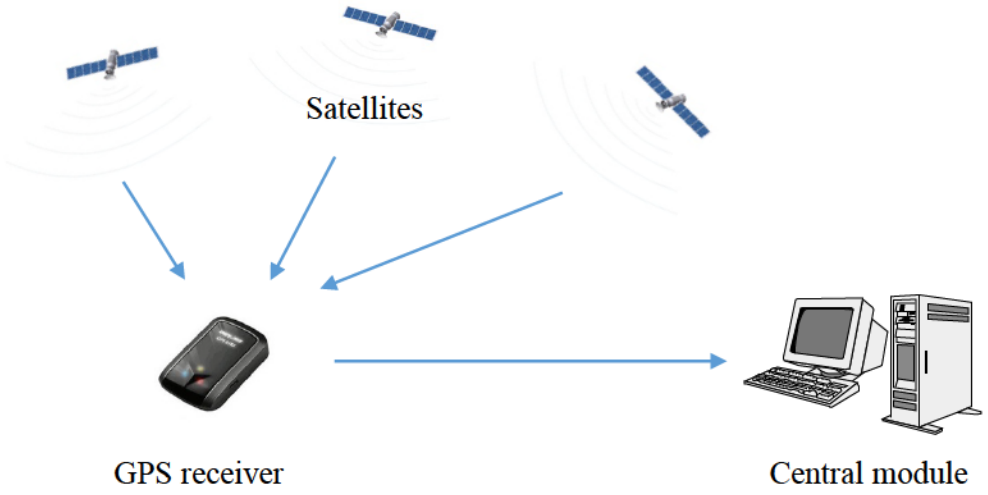

Figure 2.1 GPS localization system (Park, 2012)

Current applications of GPS on construction sites include continuously tracking the location of equipment to monitor their arrival and departure times (Hildreth et al., 2005) and to record the working cycle of the equipment for further analysis (Pradhanaga and Teizer, 2013). Also, construction materials like fabricated pipes can be located preciously with GPS receivers (Caldas et al., 2004). With the popularization of smartphones, it is now possible to track workers with smartphone built-in GPS modules (Kim and Park, 2013).

Although GPS is easy to use, the error of purely civilian GPS will be about 10 meters on a sunny day in an open area and the accuracy will be even worse when there are obstacles around or the weather is cloudy (Caldas et al., 2004). Besides, there is a report indicating that the widely used 4G-LTE open wireless broadband network that incorporates nationwide satellite coverage disrupts the operation of GPS on construction sites (ForConstructionPros.Com, 2011). Several research studies found that integrating GPS with other techniques will increase the localization accuracy obviously (Saeki and Hori, 2006). While in this way, the structures of the systems are more complicated. Another issue of tracking workers with GPS receivers is the workers' willingness. An independent survey showed that only 16% workers who haven't been tracked with GPS receivers before were positive to be tracked, and 38% workers had negative opinion because they worried about being tracked after work hours (Tsheets, n.d.).

2.1.2 Radio Frequency Identification(RFID)

The RFID is another mature localization method used on the construction sites. The RFID is a technology that stores data in tags with radio frequency (RF) compatible integrated circuits and transmits data with electromagnetic waves (Ni et al., 2004). The RFID system usually consists of RFID tags, a reader, and an information technology system, as shown in Figure 2.2 (Soltani, 2013). There are two major types of tags, passive tags and active tags. The passive ones receive energy from the electromagnetic field

generated by the reader and the active ones receive energy from its battery. When the reader requests, the tags will send data to the reader, and then the reader will send the data to the information technology system for further analysis (Goodrum et al., 2006).

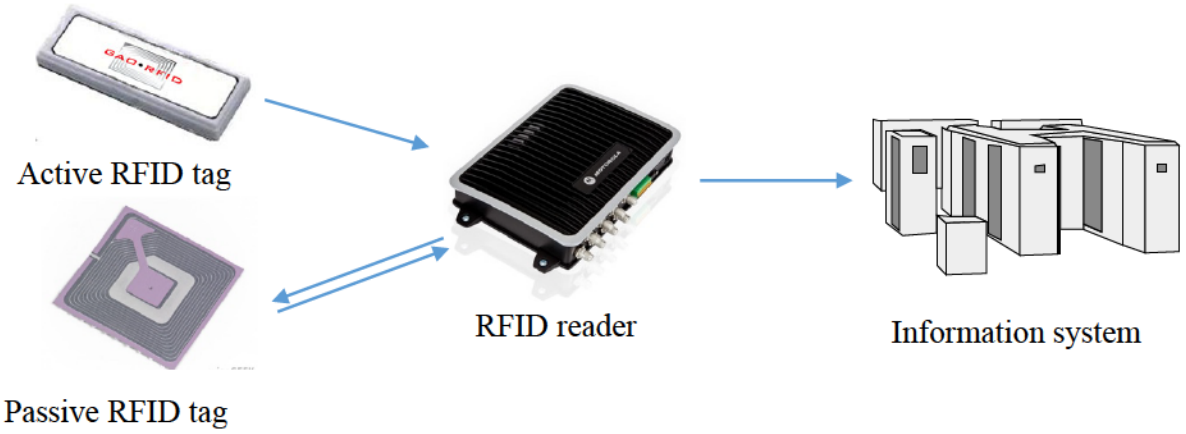

Figure 2.2 RFID localization system (Soltani, 2013)

Current applications of RFID localization systems are various. Ding et al. used RFID for tracking workers on an underground project for safety management (Costin et al.,2012). Wu et al. (2010) tracked workers and equipment with RFID system at the same time to detect near missed accidents. Also, RFID can be used for identification (Montaser and Moselhi, 2014; Chae and Yoshida, 2010; Sawyer, 2008), quality management (Wang, 2008), productivity management (Ergen et al., 2007; Grau et al., 2009), etc.

The RFID system is attractive as a mature localization system, while several limitations and issues still exist when applied to construction sites. RFID is not perfect due to the limited overlap areas (Li et al., 2016), inflexible networking capabilities, and the high cost of RFID readers (Skibniewski and Jang, 2007). Moreover, the data transmission

between tags and readers are easily interfered by metals, concrete, and moisture which are all common existences on the construction sites (Lu et al., 2007).

2.1.3 Vision Analysis

The vision analysis localization methods have also been studied for a long time and can reach an accuracy of 88% (Krumm et al., 2000). For vision based localization, the targets are not required to carry any receivers or tags. The only hardware required are different kinds of cameras such as ordinary cameras (Park and Brilakis, 2012; Park et al., 2011), range cameras (Ray and Teizer, 2012), stereo camera systems (Han and Lee, 2013; Park et al., 2011), etc. The construction sites are captured by cameras as videos or images. The captured videos or images are processed with different kinds of algorithms for localization.

The vision based localization methods are used for tracking workers and equipment currently. Yang et al. (2011) used the vision based systems to track the position of tower cranes and estimated the locations of tower cranes to track the ongoing activities. Han and Lee (2013) used vision based positioning to detect the unsafe worker behavior. The accuracy of vision based localization was tested by Brilikis et al. (2011) using a 65-mm truck model and by Park and Brilakis (2012) on a construction site.

The vision based localization methods can cover large areas of the construction sites and the implementation is easier than GPS and RFID. However, the application of vision

based localization is limited by the environment and condition of the construction sites. Reported by Park et al. (2011), occlusion, illumination, background color and other factors will influence the performance of vision based localization. Another limitation of vision based localization is the time consumption. Retrieve the location data by 3D triangulation from images (Park and Brilakis, 2012) usually takes a lot of time and can not meet the requirement of real-time processing.

2.1.4 Others

There are also many other kinds of methods for localization on construction sites, such as Ultrasound (Jang and Skibniewski, 2009; Priyantha, 2005; Priyantha et al., 2000), Ultra-wideband (UWB) (Ingram et al., 2004; Saidi et al., 2011; Shahi et al., 2012), Wireless local area network(WLAN) (Bahl and Padmanabhan, 2000; Khoury and Kamat,2009; Woo et al., 2011), etc. These methods are not mature and widely used currently as the three methods mentioned above.

2.2 Matching

Matching construction objects under different camera views is a challenging task, especially considering the fact that construction sites are typically complex, cluttered, and large. Numerous methods have been developed to improve the matching accuracy and robustness up to now. These matching methods could generally be classified into two

categories based on the matching strategy they employed. The methods in the first category relied on the visual features of construction objects in each camera view as the matching cues, while the other methods in the second category focused more on the spatial position of the objects according to the construction sites.

2.2.1 Visual Feature-based Matching

The point- and area- features are commonly employed for object matching between two camera views (Wu et al., 2011). In the visual feature-based matching methods, the visual appearance of the objects under different camera views are extracted as a set of local point or area features. Then, the objects' visual appearances in two camera views are assumed to be matched if they have the same point or area visual features. The matched objects' visual appearances signify that they are the same object captured by different cameras.

So far, several point feature detectors and descriptors are available, including SIFT (Lowe, 1999 and 2004) and SURF (Bay et al., 2006). The point features extracted by the SIFT are robust to the orientation changes of camera views, but it only detects the blob-like feature points, which might be sparse for the matching of objects' visual appearances in camera views. Compared with SIFT, SURF is detected faster, but SURF features are not fully affine invariant. It means that little feature points may be found when there is a significant change on the camera view orientations (Pang et al., 2016).

The area-feature based matching methods mainly treat the visual patterns in local image windows as the matching features. Typically, the methods in this category find seed points first and propagate to small image windows from these points. Then, the matching process can be conducted through finding cross-correlation of the local windows according to the patterns inside. For example, Pratt (2013) used the image intensities as the patterns to find the cross-correlation of local windows. Compared with the point-feature based matching methods, the area-feature based matching methods are able to produce denser matching results (Joglekar and Gedam, 2012) and are more robust to affine distortions. However, the matching with the area features might still fail, especially when the local image windows did not contain distinctive visual patterns or the patterns contained were deformed due to the complex image transformations (Chang and Gong, 2001).

The point-based and area-based feature matching methods have common limitations. First, local visual features describing the objects' visual appearance cannot always be found in captured camera views. This is especially true when the cameras are set up at the height and far away from construction sites, and the size of the object in the camera views is small. Also, the methods have difficulty to match objects with similar visual appearances in camera views. As the cameras are commonly set far from the construction site and the workers are usually wearing the similar safety vest and helmet, the visual appearances of workers in camera views are usually similar. As a result, the feature based matching

methods are easy to generate matching errors and fail to differentiate workers under different camera views.

2.2.2 Spatial Relationship-based Matching

In addition to the visual features based methods, the spatial position of the objects in camera views is also employed to conduct the object matching. The spatial information used for matching includes homography geometry and epipolar geometry (Chang et al., 2000, 2001).

The Homography geometry describes the relationship between two two-dimensional (2D) planes. When this relationship between two planes is specified, the correspondences of the points in one plane could be easily found, as long as their locations in the other plane are known (Lee et al., 2000). However, the Homography geometry could not be used for matching the objects lying in different planes.

Another commonly used spatial relationship for object matching is the epipolar geometry. According to the epipolar geometry, if the projection of a three-dimensional (3D) point X on the left view (XL) is known, then the corresponding epipolar line in the right view is decided. Meanwhile, the projection of the point X on the right view (XR) must lie on the corresponding epipolar line, as shown in Figure 2.3. Therefore, the search space for matching the objects' visual appearances under camera views is restricted from the whole image to a line (Papadimitriou and Dennis, 1996). Zhang et al. (1995) used the Least

Median of Squares (LMedS) to find the epipolar geometry between two camera views with an initial set of matched points. Based on the epipolar geometry, Lee et al. (2016) proposed a method to match onsite construction workers captured by a stereo camera system. Under this method, the location of each construction worker in the first camera view was used to find its corresponding epipolar line in another camera view. Then, the distances between the workers in the second camera view and the Epipolar line were calculated. The one closest to the line was considered to be the same worker in the first camera view (2016).

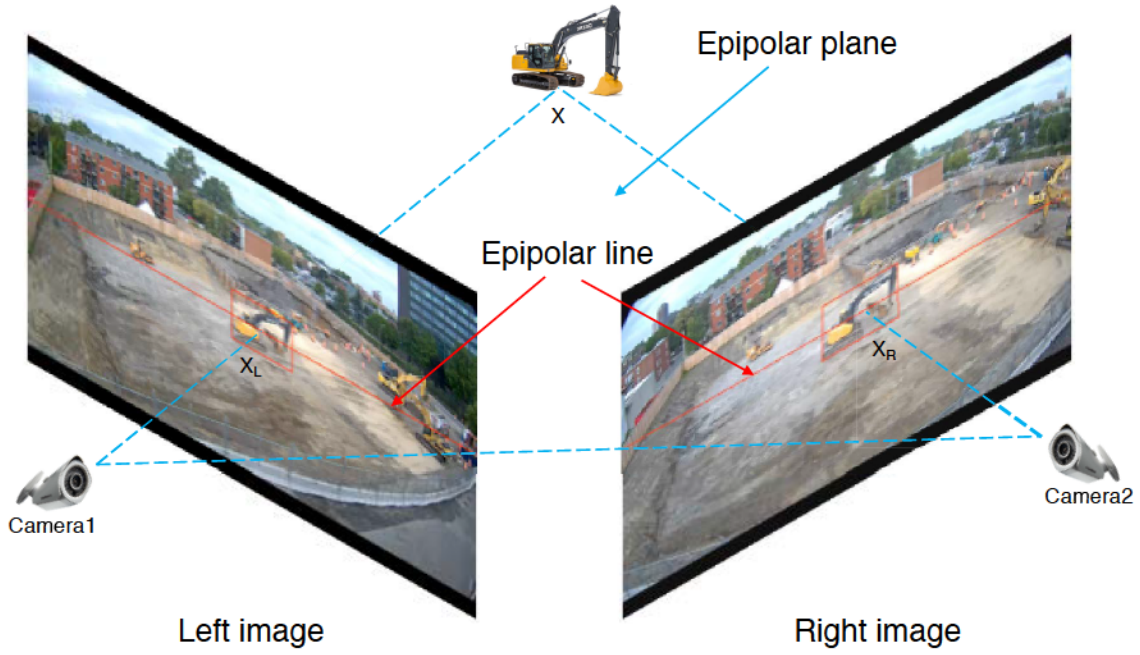

Figure 2.3 Epipolar geometry

The effectiveness of the method proposed by Lee et al. (2016) has been tested in real construction sites. However, the matching method might fail when there are two or more construction objects lie on the same epipolar line in a camera view, as shown in Figure 2.4, The failure was mainly due to the searching strategy of this method. The method used the

centroid of the bounding box of the object to represent its location and calculate the corresponding epipolar line in another camera view. The bounding boxes generated by the vision based tracking/detection methods usually have an offset, and thus the centroids do not always truly reflect the objects' location. This is especially apparent when the object in the first camera view is partially occluded. As a result, the deviations are introduced in the calculation of epipolar line and led to the potential matching error.

Figure 2.4 Object matching with epipolar lines

Konstantinou and Brilakis (2016) also proposed a method for matching workers on construction sites based on the epipolar geometry. Considering the shift of centroids, they also employed the visual features of workers and the epipolar lines of previous frames for a more accurate result. However, the cameras are set up near workers in their study thus the visual features are easily to be extracted, while in practice the cameras are set far away to capture the whole scene of the construction sites. Also, as frame drop is common when transmitting videos, features extracted from previous frames are not appropriate for real-time matching.

2.3 Fall detection

The currently widely studied automatic fall detection methods can be grouped into three major categories. The first category is wearable sensor based fall detection. Sensors are attached to the worker bodies to detect fall accidents by detecting abnormal motions. The second category is ambient sensor-based fall detection. Ambience features like audio, vibration, and pressure are collected by sensors set on the construction sites to detect fall. The third category is vision based fall detection. Workers are captured by cameras, and their features are extracted from videos or images for fall detection. The previous research studies on fall detection are introduced and discussed in the following sections:

2.3.1 Wearable sensors based fall detection

Wearable sensors based fall detection methods usually rely on the clothes embedded sensors to detect the position and monition of the body parts (Delahoz andabrador, 2014). Acceleration and postures are commonly recorded as fall detection features. Those features are employed by different kinds of classifiers for fall detection.

The accelerometer is one of the most commonly used wearable sensors (Mubashir et al., 2013). The accelerometers can be mounted on different parts of human body for fall detection. Mathie et al. (2004) used an accelerometer mounted on the waist to detect negative acceleration for fall detection. Luo et al. (2004) implemented a group of sensors

on the belt to analyze the body acceleration and posture for fall classification. To acquire more accurate results, different kinds of accelerometers are used. For example, tri-axial accelerometers are used to detect the acceleration in three axial directions (Zhang et al., 2006; Wu and Xue, 2008; Lai et al., 2011). Expect accelerometers， sensors for detecting physiological information like rate gyroscopes, photodiodes or barometric pressure sensors are also used to assist. These sensors rely on the fact that physical activities and body motions usually result in the changes in heart rate and blood pressure (Schwickert et al.,2013).

Although the wearable sensors are easy to set up and operate (Patel et al., 2012), their disadvantages are also evident. Usually, this kind of method assumes a fixed relation between the sensor and the wearer (Yu, 2008). That relation heavily relies on the workers' working environment or the type of work. Thus this kind of method is prone to fail when the working environment or the type of work are changed. Also, the workers' unwillingness of being tracked and recorded is an unavoidable issue on the implementation of wearable sensors.

2.3.2 Ambience based fall detection

The ambient sensors detect fall accidents based on the ambience changes. When a fall accident happens, changes of the surrounding ambiance are not avoidable, like the voice, the vibration and the pressure change on the ground. As the ambiance changes are usually

tiny, the ambience based fall detection methods are commonly used to detect fall accidents happen close to the sensors.

Most ambiance based fall detection methods use pressure sensors to detect and locate the fall accident (Mubashir et al., 2013). Apart from the pressure change, vibration and audio signal are also used to detect fall. In order to detect the vibration caused by fall, Alwan et al. (2006) mounted a vibration sensor on the floor. The sensor is capable of providing the location data of the fall accident based on the vibration features as well.

As the detection range of ambient sensors is usually tiny (Rashidi and Mihailidis, 2013), the ambiance based fall detection methods are not good options for large construction sites. Also, it is hard for sensors to discern the source of the ambient information. Thus the accuracy of ambience based fall detection on the construction sites is not high (Yu, 2008).

2.3.3 Vision based fall detection

Many vision based fall detection methods have been proposed in recent years. These methods adopt different kinds of camera systems, including stereo-camera systems, multi-camera systems, monocular camera systems, and depth camera systems (Sathyanarayana et al., 2015).

Multi-camera systems are mainly used to acquire the 3D features of fall detection. The 3D model of a person can be generated by the multi-camera systems, and the

distribution of the 3D model is then used to decide whether a fall accident happens or not (Auvient et al., 2011). Apart from the 3D shape, the principal component and variance ratio of the 3D human silhouette are also calculated from multi-view images and used for detecting fall accidents (Hazelhoff et al., 2008). In most multi-camera systems, the features extracted or processed from each single camera are combined with a fusion unit, so that these features could be complementary with each other to conduct the fall detection (Sathyanarayana et al., 2015). The advantage of multi-camera systems is that the detailed 3D information for the fall detection could be acquired from the multiple camera views. However, the accurate calibration and synchronized video sequences are required in order to get the reliable data. Also, it might be difficult to guarantee the real-time processing with the affordable camera hardware configurations.

Monocular camera systems are also used for fall detection. These systems, unlike the multi-camera systems, focus on the 2D features for the fall detection. These features, for example, include but are not limited to the height-width ratio of the bounding box, the velocity of the center of the bounding box and the angle of bounding ellipse (Foroughi et al., 2008). The problem of the 2D features is that the distance between the camera and the human would influence the reliability of the extracted features. In order to address such a problem, several methods for generating the 3D feature with a monocular camera system are proposed. In doing so, the camera calibration and inverse perspective mapping are used (Makantasis et al., 2012).

In order to get 3D features for the fall detection with one single camera, the idea of using the depth cameras is also proposed. The depth cameras utilized the time-of-flight principle. This way, the actual vertical velocity (Mastorakis and Makris, 2012) and 3D motion history (Dubey et al., 2012) obtained by the depth camera could be used for the fall detection. Although the depth cameras are able to get the 3D information easily and fast, the cameras are usually equipped with short-range sensors. Therefore, they are not capable of providing a wide field of view and monitoring a large area like construction sites.

2.4 Gaps in Body of Knowledge

Detecting fall accidents automatically and timely can help to improve the safety of the construction sites. Currently, numerous research has been done about fall detection and various methods have been proposed. However, several limitations and issues still exist on the state-of-art fall detection methods. For wearable sensors, the main difficulty is persuading every worker to work with attached sensors. For ambient sensors, the main difficulty is covering large construction sites with limited sensors. For vision based methods, the main difficulty is finding a fast, robust and accurate method.

In this research study, multi-camera system is employed to detect fall accidents on the construction site. In order to detect fall accidents with multiple cameras, there are still two gaps to fill. The first gap is locating the fall accident. Localization on construction sites is usually done by GPS or RFID systems, in which receivers or tags are required to be

attached on the target. For the purpose of simplifying localization process, a vision based localization method is proposed in this research and integrated with vision based object detection, object tracking and fall detection. These functions form the whole framework of the propose multi-camera fall detection method. The second gap is objective matching. In order to detect fall accidents with multiple cameras, matching the same worker in different cameras is required. The existing matching methods are mainly based on visual features and spatial relationships. Due to the complexity of construction sites, the visual features are hard to extract. The spatial relationship is also not accurate enough for matching numerous workers on large construction sites. To fill this gap, this paper proposes a novel method for matching multiple workers on the construction sites based on triangle meshes. With the support of the matching method, the multi-camera fall detection framework is implemented.

CHAPTER 3: METHODOLOGY

This chapter proposes a framework for detecting and locating fall accidents on the construction sites. The framework consists of three main parts. The first part is workers' localization. The second part is workers' matching in different camera views. The third part is vision based fall detection. The whole framework is shown in Figure 3.1. The details of the proposed framework are described in the following three sections.

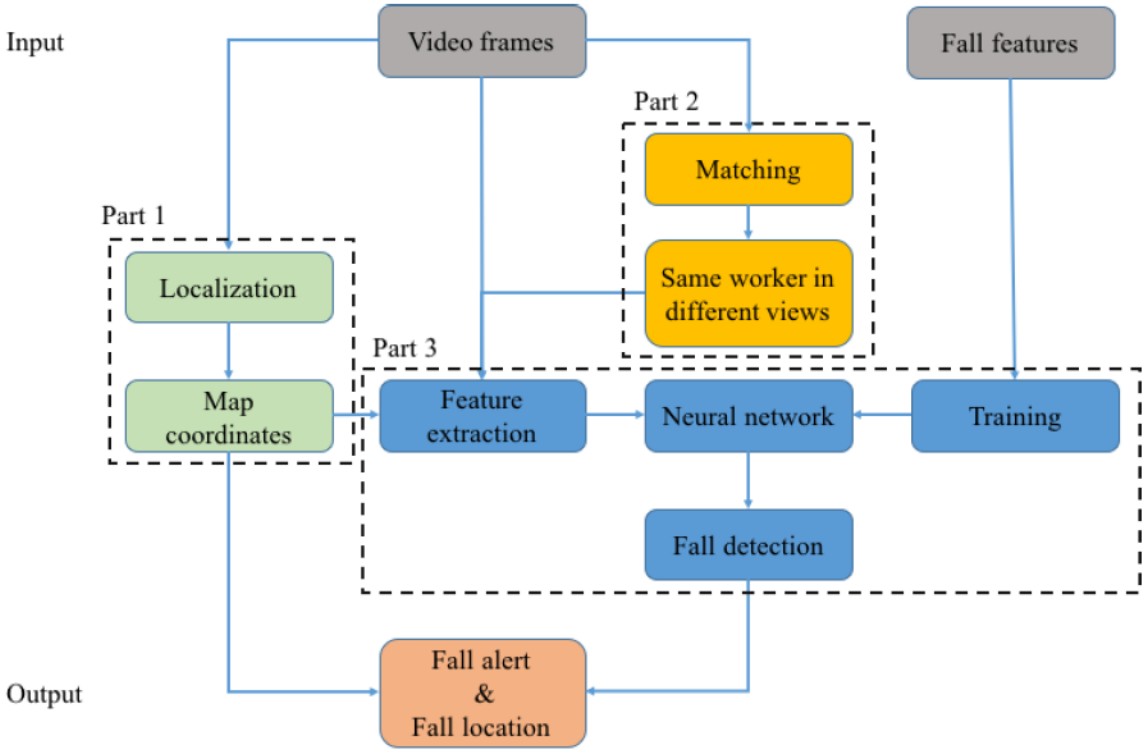

Figure 3.1 Fall detection and localization framework

3.1 Localization

This research proposes a fast localization method for workers on construction sites. It could retrieve workers' working square from videos automatically after initialization. For initialization, the correspondence between the real world map and video frames is found. The method detects workers in the video frames at first, then tracks workers based on the detection results. To improve the tracking accuracy, the detection method is implemented every few frames. With the tracking/detection results, the workers are mapped from video frames to the Google Map and the working square of workers are visualized. The framework of the localization method is shown in Figure 3.2.

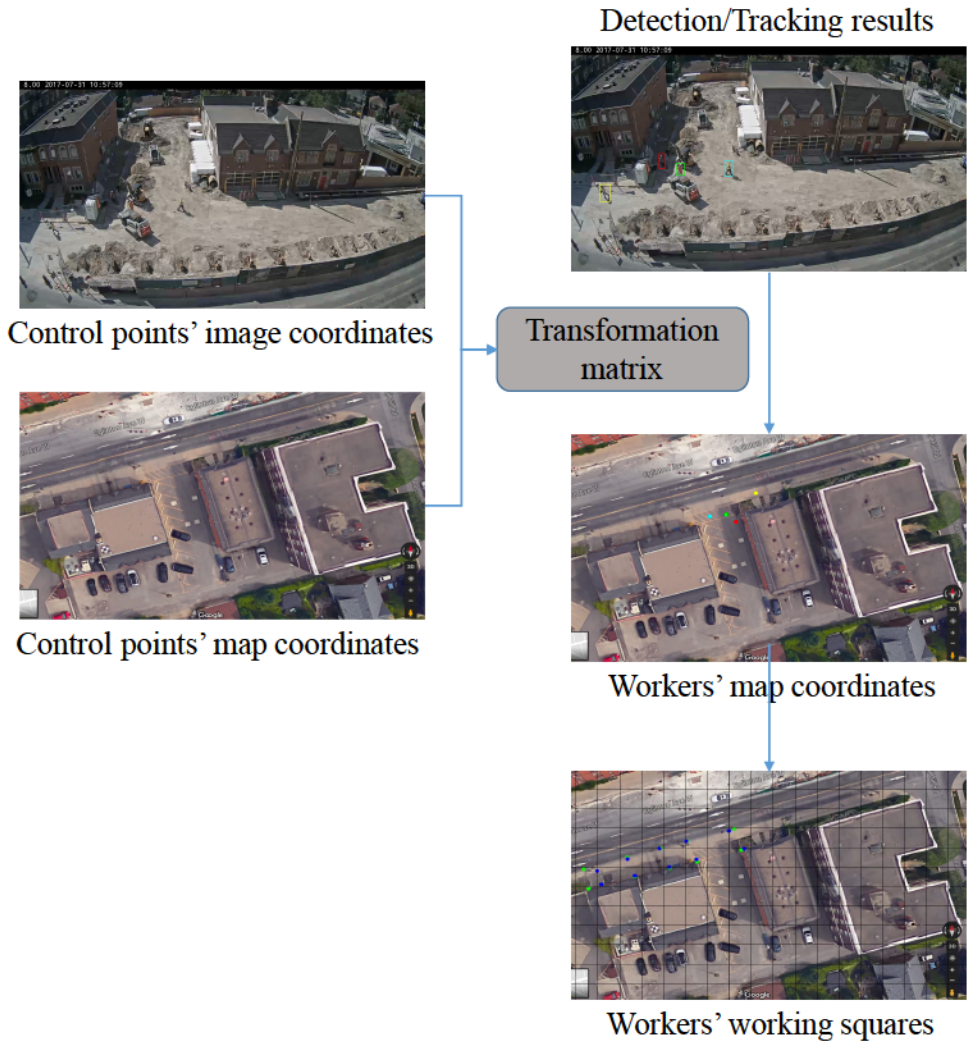

Figure 3.2 Framework of the localization method

3.1.1 Initialization

The purpose of initialization is to find the correspondence between the map of the construction site and the video frames. The locations of workers are expressed by their map coordinates and measured with working squares in the map. The working squares are shown in the following Figure 3.3 and each side of squares is 50 pixels, about 6 meters in

practice.

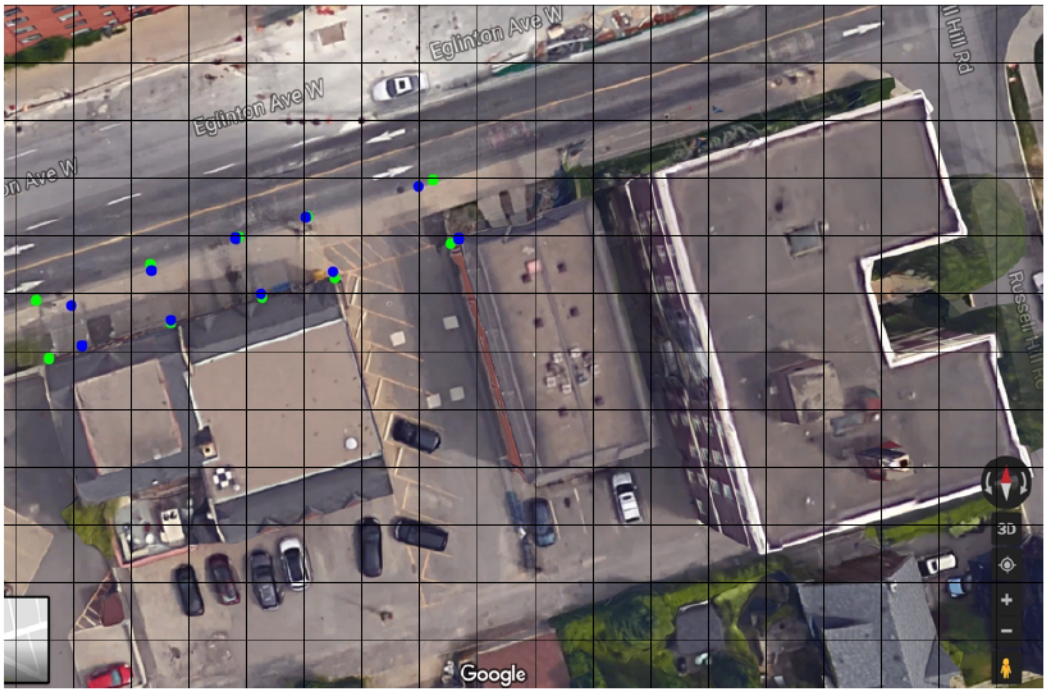

Figure 3.3 Example of working squares

It is assumed that the transformation between the map coordinates and the image coordinates is perspective transformation. Thus the relation between the coordinates in two systems is:

$$\begin{bmatrix} tx \\ ty \\ t \end{bmatrix} = M \cdot \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{3.1}$$

where (x, y) is the image coordinate of the point, M is the 3×3 perspective transformation matrix and t is a random constant. To calculate M, 4 pairs of matched map and image coordinates are required. In this research, the map coordinates and the image coordinates both are collected manually in Adobe Photoshop CC.

3.1.2 Detection and tracking

In order to locate the workers on the map, it is necessary to locate workers in the images at first. For automatic locating workers in the video frames, Single Shot multi-box Detection (SSD) detection algorithm (Liu et al., 2016) and Kernelized Correlation Filters (KCF) tracking algorithm (Henriques et al., 2015) are employed.

In the first frame of the video, the workers are detected by the SSD algorithm. The detection algorithm generated a bounding box of each detected worker and passed the bounding boxes to the KCF algorithm. The KCF algorithm then tracked workers based on the received bounding boxes. The KCF tracker is very fast that can track workers real-time, but it is not robust to occlusion. When the tracking object is occluded, the tracker will fail. Thus SSD is called to initialize the input bounding boxes of KCF tracker during the tracking process. As the speed of SSD is relative slow, it is called once every 24 frames in this research, which means the detection is performed once per second.

In this research, all the proposed methods including localization, matching and fall detection are based on the correct worker bounding boxes generated by the proposed integration of SSD and KCF. That is to say, all the workers in the videos have a bounding box showing their location during the whole length of the videos. The performance of KCF and SSD can be referred in *Ssd: Single shot multibox detector* (Liu et al., 2016) and *High-speed tracking with kernelized correlation filters* (Henriques et al., 2015).

3.1.3 Transformation

After the localization on the image, the workers' position on the map can be obtained by perspective transformation. The coordinate of the right-bottom point of a worker's bounding box is selected as the location of the worker on the image. Then the image coordinate is transformed to map coordinate with Eq. 3.1. The worker's location then can be described with the working square containing the map coordinate.

3.2 Matching

The research proposes a novel method that could be used to match construction objects (e.g., equipment, worker, and temporary facility) captured from onsite camera views in an automatic manner. The proposed method includes two main parts. The first part searches the potential matching candidates and the second part matched them with combinatorial optimization. In the first part, the visual feature points under different camera views are detected and matched at first. Then, the epipolar geometry between different camera views is established based on the matched feature points to search the potential matching candidates. Also, a dynamic matched triangle mesh pair is generated in different camera views using the matched visual feature points. Based on the locations of the potential matching candidates in the corresponding triangle meshes, their triangle coordinates are further calculated. The difference in their triangle coordinates is defined as

the matching cost. This way, the matching of multiple construction objects in different camera views can be solved by finding the minimum matching cost through the combinational optimization in the second part. Figure 3.4 illustrates the overall framework of the proposed method, and its details are described in the following sections.

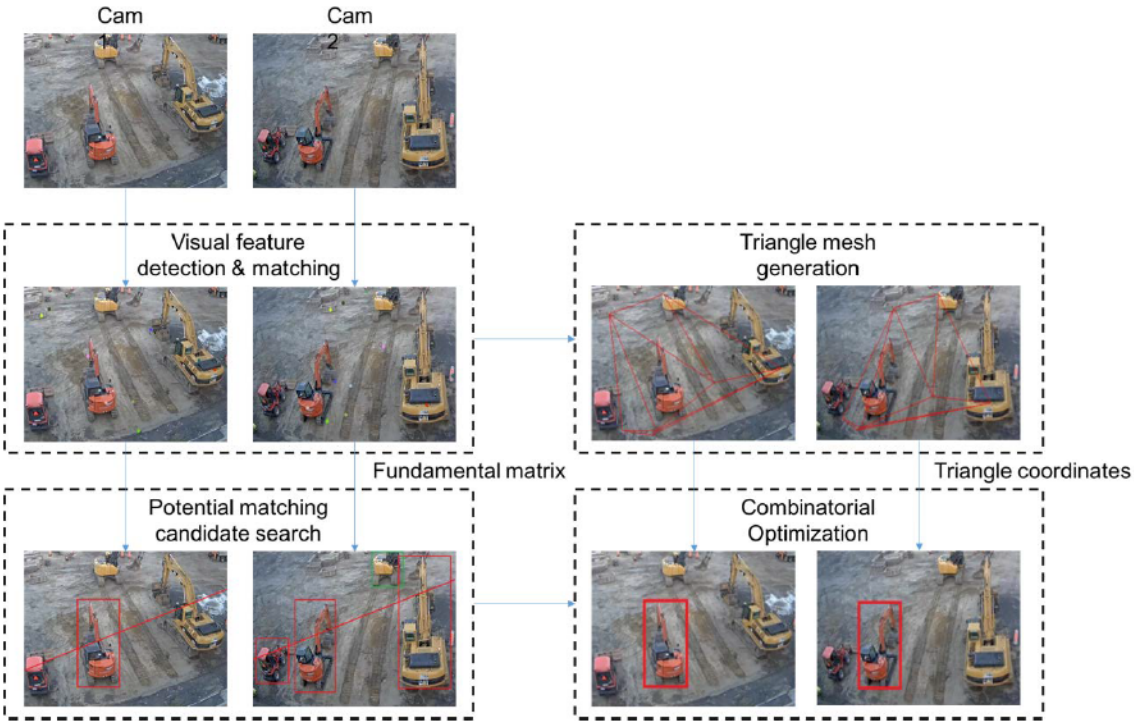

Figure 3.4 Framework of the matching method

3.2.1 Potential Matching Candidates Search

The potential candidates matching work under each pair of camera views is similar to the previous research study proposed by Lee et al (2016). Suppose two camera views, i.e., Cam1 and Cam2, are under investigation. The SIFT detector/descriptor (Lowe, 1999, 2004) is first employed to find an initial set of matched feature points in both camera views,

considering the robustness of SIFT to large perspective and scale changes (Zhu et al., 2007). Then, the RANdom Sample Consensus (RANSAC) method (Fischler and Bolles, 1981) is adopted to remove those wrong matched point pairs, as suggested by Hartley and Zisserman (2003). Only correctly matched feature points are remained for the next part.

Based on those correctly matched feature points, a 3x3 fundamental matrix is calculated to establish the epipolar geometry, which describes the intrinsic projective geometry between any pair of camera views (e.g., Cam1 and Cam2). For each object of interest in Cam1, its corresponding epipolar line in Cam2 is determined with the fundamental matrix. The distances between the objects of the interest in Cam2 and the epipolar line are calculated. Only those objects whose distances to the epipolar line are equal to or smaller than their bounding box size is kept as the potential matching candidates of the object of interest in Cam1. The purpose of finding the potential matching candidates in Cam2 of the objects of interest in Cam1 is to reduce the computation complexity of the combinatorial optimization described in the next section.

3.2.2 Combinatorial Optimization

When all potential candidates in Cam2 are identified for matching the objects of interest in Cam1, the proposed method tries to address the multiple objects matching problems between two camera views with a combinatorial optimization. Suppose n objects $\{O_1, O_2, O_3, \ldots, O_n\}$ are found in Cam1, and m potential matching candidates $\{C_1, C_2,$

C_3, ..., C_m} are identified in Cam2. Then, an n by m cost matrix, M, is formulated, as shown in Eq. 3.2.

$$M = \begin{bmatrix} M_{11} & \cdots & M_{1m} \\ \vdots & \ddots & \vdots \\ M_{n1} & \cdots & M_{nm} \end{bmatrix} \tag{3.2}$$

Where the element, M_{ij}, in the matrix indicates the corresponding cost for considering the ith object (O_i) in Cam1 is matched to the jth candidate (C_j) in Cam2. If C_j in Cam2 is not in the list of the potential matching candidates of O_i identified in the previous part, M_{ij} is set as $+\infty$. Otherwise, the specific value of M_{ij} is calculated as follows. A triangle mesh (TM_1) is first generated based on the correctly matched feature points found in the last part in Cam1 using the Delaunay triangulation process (Lee and Schachter, 1980). The triangle mesh (TM_1) is then projected into CamView2 to form the corresponding triangle mesh (TM_2). Then, the triangle coordinates of the object (O_i) according to TM_1 and the candidate (C_j) according to TM_2 are calculated. The triangle coordinate of a point p according to a triangle ABC is defined by Eq. 3.3:

$$(p_a, p_b, p_c) = \left(\frac{S_{\Delta PBC}}{S_{\Delta ABC}}, \frac{S_{\Delta PAC}}{S_{\Delta ABC}}, \frac{S_{\Delta PAB}}{S_{\Delta ABC}}\right) \tag{3.3}$$

Where $S_{\Delta ABC}, S_{\Delta PBC}, S_{\Delta PAC}, S_{\Delta PAB}$, are areas of triangles.

The matching cost of the object and a corresponding candidate is defined as the Euclidean distance between their triangle coordinates as illustrated in Eq. 3.4:

$$M_{ij} = \sqrt{(O_{i1} - C_{j1})^2 + (O_{i2} - C_{j2})^2 + (O_{i3} - C_{j3})^2} \tag{3.4}$$

Where (O_{i1}, O_{i2}, O_{i3}) and (C_{j1}, C_{j2}, C_{j3}) are the triangle coordinates of the centroids of

the object (O_i) in *TM1* and the candidate (C_j) in *TM2*.

After determining the matching cost matrix *M*, the Hungarian algorithm (Jonker and Volgenant, 1986) is used to find a total minimum cost incurred to conduct the one-on-one matching between the objects of interests in Cam1 and the potential candidates in Cam2. It is worth to note that the number of the objects of interest in Cam1 does not have to be the same as the number of the potential candidates in Cam2. In other words, the Hungarian algorithm (Jonker and Volgenant, 1986) could still work even if *M* is not a square matrix (i.e., n≠m). The path with the smallest total cost generated by Hungarian algorithm indicates the best matching result.

3.3 Fall detection

This research proposes a novel method for fall detection with the help of an artificial neural network. The artificial neural network is trained with a training set at first and then can be used for fall detection on construction sites. The artificial neural network utilizes six features of workers to classify workers' action including the height-width ratio of bounding box, the bounding ellipse angle, the workers' actual height and the difference of them between video frames. After training, the artificial neural network can detect workers' fall accidents on construction sites automatically. The details of feature extraction and artificial neural network training are described in the following sections. The framework of the fall detection method is shown in Figure 3.5.

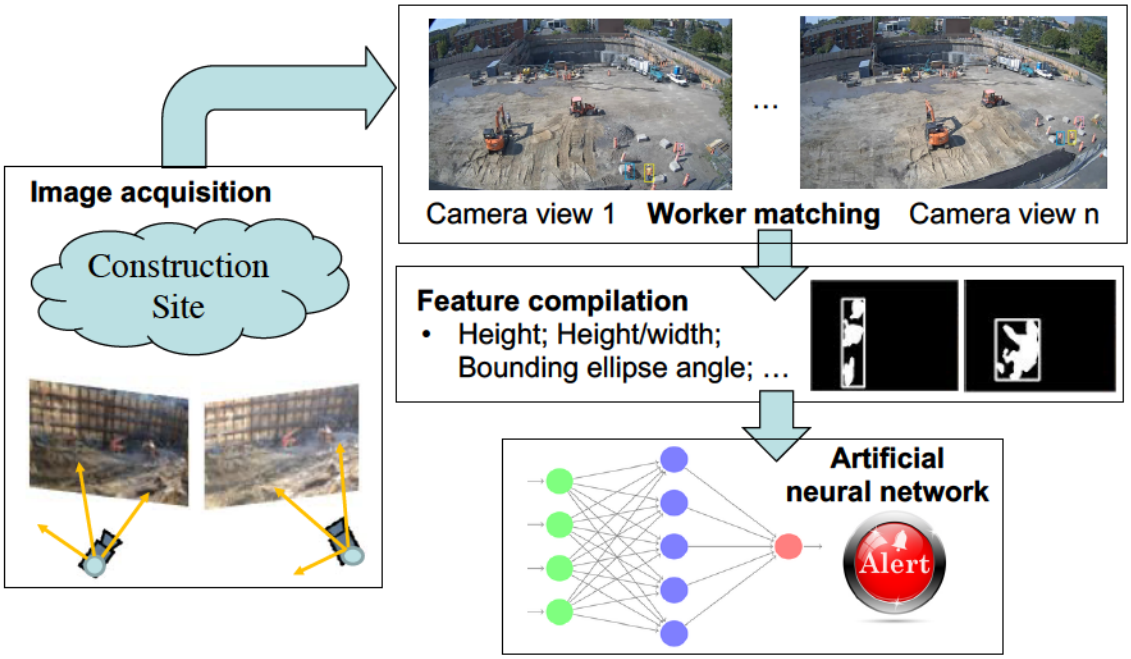

Figure 3.5 Framework of the fall detection method

3.3.1 Feature extraction

According to Makantasis et al.'s (2015) work, fall accidents could be described by motion features, including height-width ratio, bounding ellipse and actual height. When a worker is falling, the height-width ratio of his or her bounding box is always smaller than the one when he or she is standing. Also, the angle of the bounding ellipse is close to 90° when the worker is standing and close to 0° when fall. To differentiate fall from other actions like a bow or sit, the changing speed of height-width ratio and bounding ellipse are also considered. The actual height and vertical motion velocity can also reflect the motion of the worker. The actual height of the worker is very small when the worker falls on the

ground and the value of the vertical motion velocity would be larger when a fall accident occurs. Thus all these six features are used for detecting a fall accident.

In order to extract all these features, first of all, the workers' silhouette should be extracted from the video images. A foreground extraction algorithm proposed by KaewTraKulPong and Bowden (2002) was used. The workers are extracted from the background based on the K Gaussian distributions of pixels. The extracted workers' silhouettes are then transformed into the black-white format. At the same time, the minimum bounding box containing all the pixels of each extracted silhouette is calculated, as shown in Figure 3.6. The pixel coordinate of four corners (top-right, top-left, bottom-right and bottom-left) of the bounding box are recorded as P_{tr} (p_t, p_r), P_{tl} (p_t, p_l), P_{br} (p_b, p_r), P_{bl} (p_b, p_l).

Figure 3.6 Bounding box of extracted workers

The first feature to be calculated is the height-width ratio. The height-width ratio is defined as the ratio between the height and the width of the bounding box. As the coordinates of the four corners of the bounding box are known, the ratio could be expressed by the following equation:

$$R = \frac{h}{w} = \frac{p_t - p_b}{p_r - p_i} \tag{3.5}$$

where h is the height of the bounding box and w is the width of the bounding box. For a standing worker, the ratio is higher than a fell down worker. The difference of height-width ration between two frames ΔR is also calculated and recorded as a fall detection feature, and for the first frame the value is set to be zero.

In order to obtain the bounding ellipses of the workers, the image moments describing the extracted workers' silhouettes were used. An ellipse is defined by its centroid (x_c, y_c), its major and minor semi-axes a and b and its orientation θ. These parameters of an ellipse could be calculated with the image moments of a black-white silhouette. The image moment is a certain weighted average of the image pixels' intensities, and the image moment M_{ij} of a scalar image is defined as (Hu,1962):

$$M_{ij} = \sum_x \sum_y x^i y^j I(x, y) \quad \text{for i, j} = 0, 1, 2 \ldots \tag{3.6}$$

where $I(x, y)$ is the pixel intensity at the point (x, y).

The centroid of the ellipse (x_c, y_c) coincides with the mass center of the extracted silhouette, which could be calculated by (Hu,1962):

$$(x_c, y_c) = \left(\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}}\right) \tag{3.7}$$

After calculating the centroid of the ellipse, the orientation of the ellipse could be calculated by the second order central moments. The central moment μ_{ij} is defined by (Hu,1962):

$$\mu_{ij} = \sum_x \sum_y (x - x_c)^i (y - y_c)^j I(x,y) \quad \text{for i, j} = 0, 1, 2 \ldots \tag{3.8}$$

and its orientation θ is obtained by (Hu,1962):

$$\theta = \frac{1}{2} \arctan \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right) \tag{3.9}$$

The length of the major semi-axis a and the minor semi-axis b could also be calculated by the central moment (Hu,1962):

$$a = \left(\frac{4}{\pi} \right)^{\frac{1}{4}} \left(\frac{{I_{max}}^3}{I_{min}} \right)^{\frac{1}{8}} \tag{3.10}$$

$$b = \left(\frac{4}{\pi} \right)^{\frac{1}{4}} \left(\frac{{I_{min}}^3}{I_{max}} \right)^{\frac{1}{8}} \tag{3.11}$$

where I_{max} and I_{min} are moments of inertia and given by (Hu,1962):

$$I_{max} = \frac{1}{2} (\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4{\mu_{11}}^2}) \tag{3.12}$$

$$I_{min} = \frac{1}{2} \left(\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4{\mu_{11}}^2} \right) \tag{3.13}$$

The bounding ellipses are drawn with the three features calculated above and shown in Figure 3.7. The bounding ellipse angle θ and the difference of bounding ellipse angle between two frames $\Delta\theta$ are calculated and recorded as fall detection features. For the first frame, the value of $\Delta\theta$ is set to be zero.

Figure 3.7 Bounding ellipse of extracted workers

Instead of using worker's projection height, the worker's actual height is used in this method. It is because the actual height is irrelevant to the distance between the worker and the camera. In addition, the actual height could give information about the type of the moving object extracted from the background, and then the moving construction equipment won't be mistaken as a worker. What's more, with the actual height, the mounting position and the view of the camera would not be restricted.

In order to obtain the actual height of the worker, the camera location and the worker's location are required. In the first part of the framework, the worker's working square is determined. As the size of the working square are relative small compared with the distance between the working square and the camera, the center of the working square can be considered as the worker's location. Thus the distance between the camera and the worker can be estimated as the distance between the camera and the center of the working square containing the worker. As we are using a pinhole camera model, the worker actual height H is proportional to the worker image height h, and the relationship could be described as

$$H = \mathrm{Z}\frac{h}{f} \tag{3.14}$$

where f is the focal length of the camera. The relationship is shown in Figure 3.8. As h is calculated above and f could be obtained from the instruction of the camera, the actual height of the worker then can be obtained. The vertical motion velocity is calculated by the difference of height in two frames.

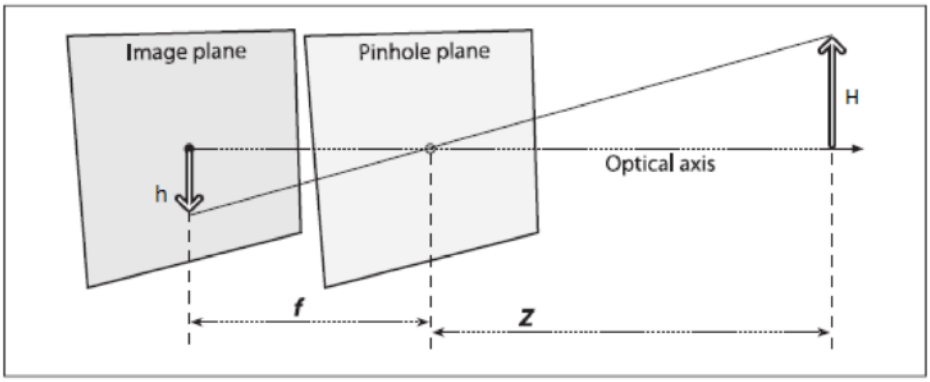

Figure 3.8 Pinhole camera model (Forsyth and Ponce, 2011)

3.3.2 Artificial neural network

When the six features of workers' motion are obtained, a Back Propagation Neural Network is trained to detect the fall. In order to avoid that the propagation fall to some local minima, a Genetic Algorithm (Whitley, 1994) is used to generate the initial weight and threshold values. Once the training of the neural network is done, the neural network could detect fall accidents from video automatically.

The neural network is trained with the supervised learning method. The inputs are 7x1 vectors. The former six elements are the above mentioned six features extracted from each frame and the seventh element is the ground truth label identifying the image as fall or not,

0 for not fall and 1 for fall. The video is extracted to frames of 5 fps, which means the time between two successive images is 200ms. This time is sufficient for detecting a fall and discriminating it from other activities. For the last element, the value will be set as 1 if the previous six features describe a fall and the value will be 0 oppositely.

The neural network is a 1-hidden layer neural network. It consists of three layers. The first layer is the input layer with six neurons, the second layer is the hidden layer and the third layer is the output layer with one neuron. The activation function of the hidden layer is Rectified Linear Unit function which is:

$$f(x) = \begin{cases} x & x > 0 \\ 0 & x < 0 \end{cases} \tag{3.15}$$

and the activation function of the output layer is the Sigmoid function which is:

$$f(x) = \frac{1}{1+e^{-x}} \tag{3.16}$$

The structure of the neural network is shown in Figure 3.9. During the training process, the neural network is initialized at first. Then the feature vectors are put into the initialed neural network and the identification results are calculated. The Loss is defined as the difference between the output of the neural network and the ground truth label of frame identifying fall or not. The parameter of the neural network including the weight w_1, w_2 and the bias b_1, b_2 are optimized by gradient descent to reach the global minimum of the Loss.

$$A = Relu(Z_1) \qquad Y = Sigmoid(Z_2)$$

$$Z_1 = w_1 X + b_1 \qquad Z_2 = w_2 X + b_2$$

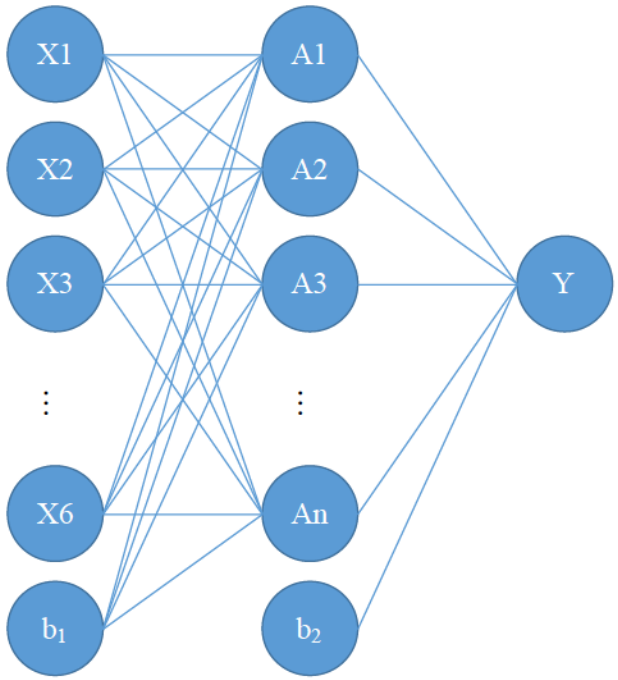

Figure 3.9 Structure of fall detection neural network

The neural network would work automatically after the training. The input of the network are the six features extracted from the videos and the output is the label 0 or 1, 1 for fall and 0 for other actions. Considering the precision, only if 6 in 10 successive frames are labelled with 1, a fall accident will be alerted by the algorithm. The time window of 6 frames is about 1.2s which is the average duration of a fall incident.

3.3.3 Multiple cameras fall detection

The fall detection algorithm is conducted by each camera separately and then the single camera fall detection results are integrated to reach a final result. Suppose that *m* cameras are employed for fall detection. If no less than *m/2* cameras reported a worker's fall

accident at the same time, a fall accident is detected by the proposed framework. In this way, if the worker is occluded in some camera views, the proposed framework can still detect the fall accident.

CHAPTER 4: IMPLEMENTATION AND RESULTS

In order to validate the effectiveness of the proposed framework, three parts of the framework were implemented and tested. For each part, the implementation environment and testbed are listed at first, then followed the test results. The results of each part of the framework are discussed separately as well in each section.

4.1 Localization

4.1.1 Implementation

The proposed method was implemented on the Python platform with the support of the OpenCV library (Beyeler, 2015). The OpenCV library provides the critical algorithms, functions, and tools required for basic image processing operations. The method was tested on a Mac OS Sierra operating system. The hardware configuration for the test includes an Intel® Core™ i5 CPU (Central Processing Unit) @ 2.30 GHz, 8 GB memory.

4.1.2 Test results

The images selected from two real construction sites in Toronto, Canada, were used for the test. The cameras were placed on the top of buildings next to the construction sites to capture the scene of the construction sites. Two video frames extracted from two videos were used for testing the accuracy of localization.

For initialization, 4 landmarks in each construction sites were selected to calculate the transformation matrix. Then their map coordinates and image coordinates were retrieved from Google map and the extracted video frames. The selected landmarks are shown in Figure 4.1-4.2. Based on the selected 8 point pairs, the transformation matrixes of two construction site views were calculated.

Figure 4.1 Localization scene 1 with selected landmarks

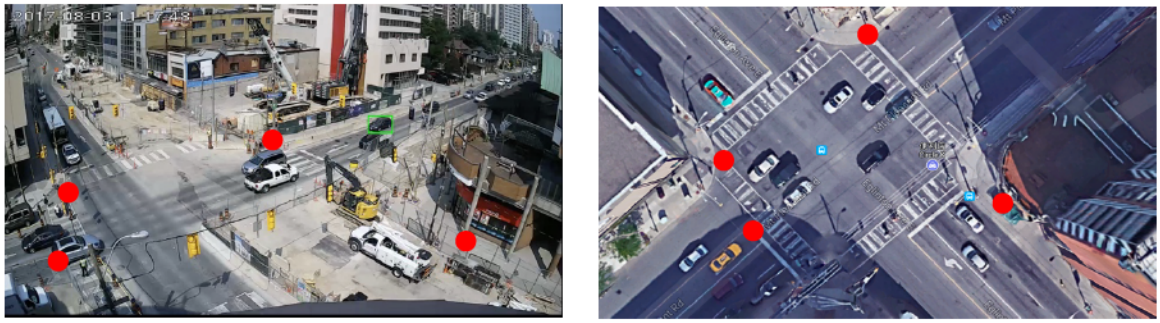

Figure 4.2 Localization scene 2 with selected landmarks

In order to test the accuracy of localization, ten landmarks on each construction site were selected. The original landmarks are marked on the maps by green points and their projection positions are marked by blue points. If the green point and the corresponding blue point are in the same working square, the green point is regarded to be matched correctly. The localization results are shown in Figure 4.3-4.4; the localization accuracy is

listed in Table 4.1.

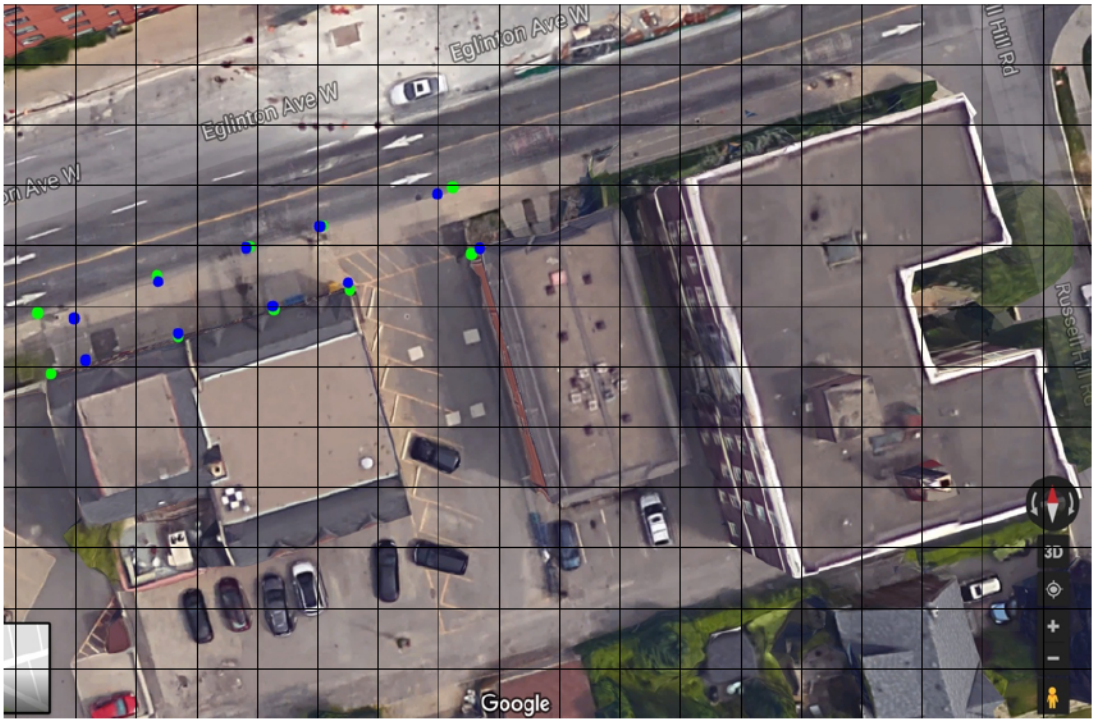

Figure 4.3 Localization test result 1

Figure 4.4 Localization test result 2

Scene number	Selected points	Correct localization results	Localization accuracy
1	10	9	90%
2	10	9	90%
Total	20	18	90%

Table 4.1 Localization results

4.1.3 Discussion

Besides the working square, the GPS coordinates of 10 selected landmarks were used to further evaluate the accuracy of the localization method. The accuracy was measured by the distance between the ground truth and the coordinates calculated by the proposed method. The GPS coordinate ground truth was retrieved from the Google Map. The distance was calculated by GPS coordinates (e.g., longitude and latitude) based on the equations:

$$C = \sin(LatA) * \sin(LatB) * \cos(LonA - Lonb) + \cos(\text{LatA}) * \cos(\text{LatB}) \quad (4.1)$$

$$Distance = R * Arccos(C) * \pi/180 \quad (4.2)$$

where (LonA, LatA) is the ground truth coordinate, (LonB, LatB) is the calculated coordinate, and R is the radius of the Earth which is 6371004m. The average localization error is 1.03m, the maximum localization error is 4.24m and the minimum localization error is 0.15m.

From Figure 14-15, it can be found that the accuracy of the proposed method is

relative to the position of the point in the image. The localization error of the points in the image center is small, while the error of the points on the image edges is large. The relation between the location of points and error is shown in Figure 4.5-4.6. As multiple cameras are used in the proposed framework, one worker may be captured by multiple cameras. It is recommended to select the image in which the worker is in the center of image to do the localization, in this way the most accurate localization result will be obtained.

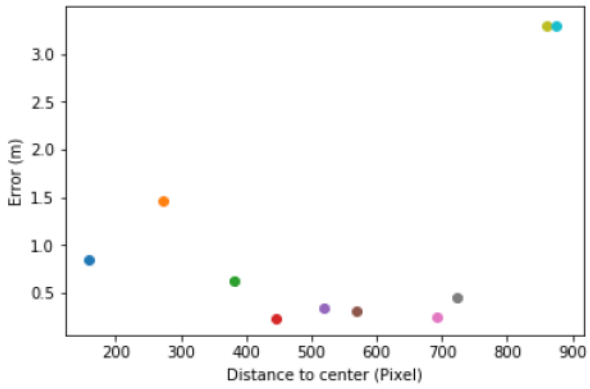

Figure 4.5 Relationship between localization errors and bounding box centers 1

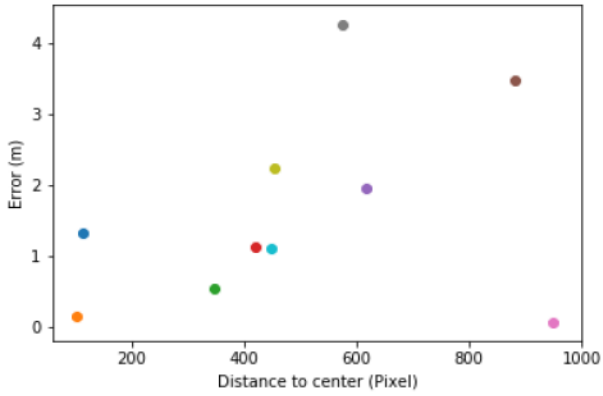

Figure 4.6 Relationship between localization errors and bounding box centers 2

Besides the localization of workers, the proposed can also be used for localization of construction equipment. By localization, the trajectories of the construction equipment or

workers can be obtained. Figure 4.7 shows the trajectories of construction equipment on a construction site. The real-time locations and trajectories generated by the proposed method can be used for other automatic construction management purposes like working efficiency analysis, collision alert, etc.

Figure 4.7 Trajectory of two construction equipment

4.2 Matching

4.2.1 Implementation

The proposed method has been implemented in the Python platform with the support of the OpenCV (Beyeler, 2015) and Munkres libraries (Pilgrim, 2017). Both libraries provide the critical algorithms, functions, and tools required for basic image processing operations. The method was tested on a Microsoft Windows 10 64-bit operating system. The hardware configuration for the test includes an Intel® Core™ i7-7700HQ CPU (Central Processing Unit) @ 2.80 GHz, a 16 GB memory, and a NVIDIA GeForce GTX 1070 GDDR5 @ 8.0 GB GPU (Graphics Processing Unit).

4.2.2 Test results

The images selected from a real construction site in Montreal, Canada, were used for the tests. A total of four high definition video cameras were placed on the site to record daily construction activities for a period of 6 months starting from August 2015. The placement of the cameras on the site is shown in Figure 4.8. Examples of the test images are shown in Figure 4.9.

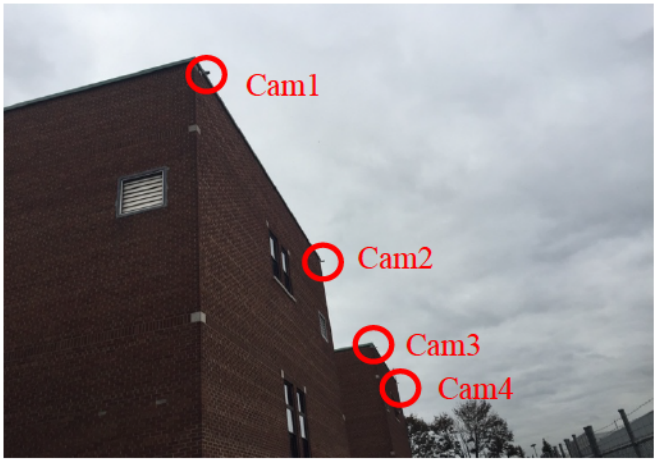

Figure 4.8 Matching camera placement

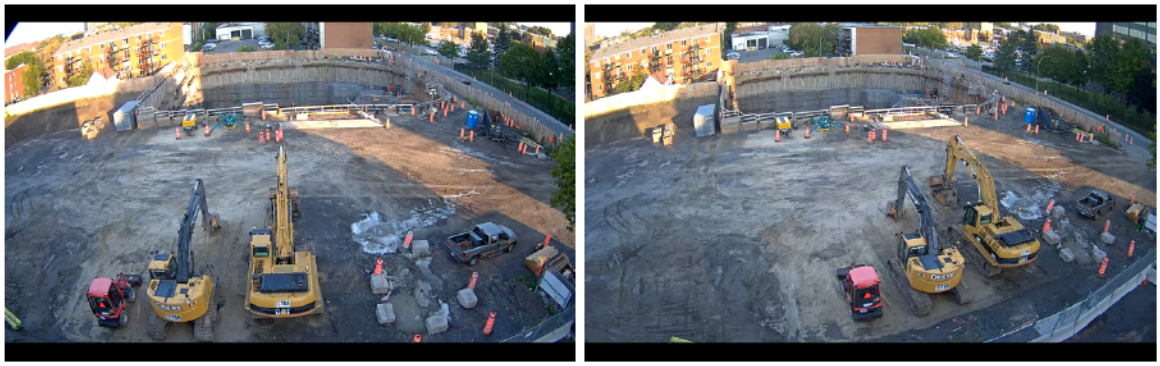

Figure 4.9 Matching test example

Different environmental conditions are considered for testing the performance of the proposed method. Specifically, the construction site images captured at day and night are used to evaluate the performance of the proposed method under different lighting conditions. The construction site images captured under sunshine and snow are used to evaluate the impact of weather conditions on the performance of the proposed method. Figure 4.10-4.12. shows the different test conditions.

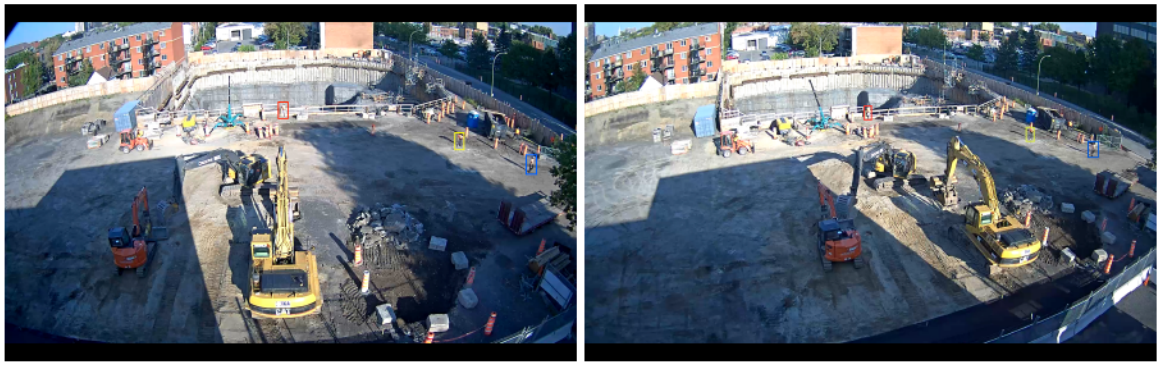

Figure 4.10 Workers matching at day

Figure 4.11 Workers matching at night

Figure 4.12 Workers matching under snow

The overall matching accuracy of the proposed method is 93.01%. Table 4.2 compared the performance of the proposed method when matching workers in daytime and nighttime. The accuracy for matching the workers in daytime and nighttime are 90% and 93.33%, respectively. Table 4.3 compares the performance of the proposed method on matching workers under different weather conditions. Their corresponding matching accuracies are 91.79% and 94.50% under the sunny and snowy conditions respectively.

	Correct Matched Pairs	**Total Pairs**	**Accuracy**
Daytime	54	60	90%
Nighttime	56	60	93.33%

Table 4.2 Matching results at daytime and nighttime

	Correct Matched Pairs	**Total Pairs**	**Accuracy**
Sunny	54	60	90%
Snowy	103	109	94.50%

Table 4.3 Matching results under different weathers

4.2.3 Discussion

The test results showed the effectiveness of the proposed method for matching workers on construction sites under different environmental conditions. In most cases, the matching accuracy of the proposed method could reach more than 90%. The matching accuracy on a sunny day is a little lower than the matching accuracy on a snowy day or at nighttime because on sunny days more equipment worked on the construction site and thus occlusions are more common. As the difference between the daytime and the nighttime condition is 3.33% and the difference between the sunny and the snowy condition is 4.5%, it can be concluded that the lighting and weather conditions have little effect on the matching accuracy of the proposed method.

It could be seen from the test results that the number of the workers in each camera view does not have to be equal. For example, there are 5 workers in the left camera view and 4 workers in the right camera views in Figure 4.13. The proposed method could successfully match the 4 out of 5 workers in the left camera view to the ones in the right camera view.

Figure 4.13 Matching unequal numbers of workers

On the other hand, it could be found that the matching accuracy of the proposed method was reduced when the workers are close to each other on the construction sites, which makes the matching more challenging. In addition, the strategy adopted in the proposed method is to find a total minimum matching cost between the pairs of workers in different camera views. When there is one pair of workers matched incorrectly, the error may be propagated and affect the correct matching of other pairs of workers, as shown in Figure 4.14.

Figure 4.14 Wrong matching results propagation

Furthermore, it is important to match visual feature points correctly, since the finding of the epipolar lines and the generation of the triangle meshes are both based on the

matched feature points. In this research, the matching of SIFT features under the different camera views relies on a threshold, which is a ratio describes the tolerance of the potential feature matching errors. A larger threshold indicates a higher tolerance of the matching errors and thus leads to more pairs of matched feature points, covering the larger overlapping areas. Figure 4.15 shows an example of the triangle meshes generated with different threshold values (0.5 and 0.9). It could be seen that the triangle mesh cover a smaller area when the threshold value is set to be 0.5, while the feature points are incorrectly matched with the threshold value equal to 0.9, although the triangle mesh covers a larger area.

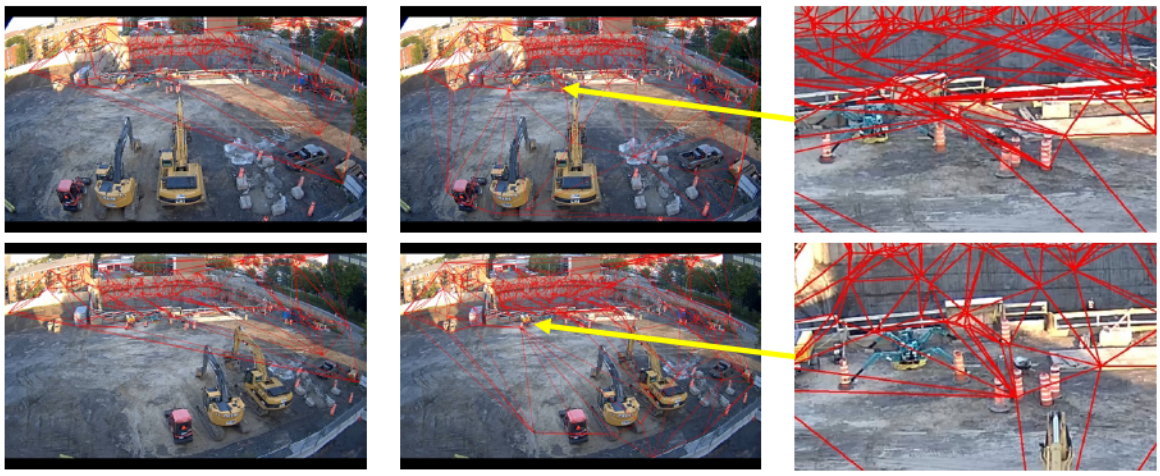

Figure 4.15 Triangle meshes generated by different thresholds

(Left: Threshold=0.5, middle: Threshold=0.9)

In order to determine the appropriate threshold values in this research study, the accuracy for matching the construction workers with different threshold values was evaluated. The values range from 0.5 to 0.9. Figure 4.16 shows the evaluation results. It

could be seen that the matching accuracy dropped when the threshold value was selected too large or too small. When the threshold is too large, it introduced many wrongly matched feature points to generate the triangle meshes, which reduced the accuracy for matching construction workers. When the threshold is too small, not enough feature points could be matched to generate the large triangle mesh. As a result, many construction objects of interest were outside of the mesh, and that affected the final matching accuracy. The maximum matching accuracy could be achieved when the threshold value is in the range of 0.6 to 0.8. In this research, the threshold value was selected to be 0.7.

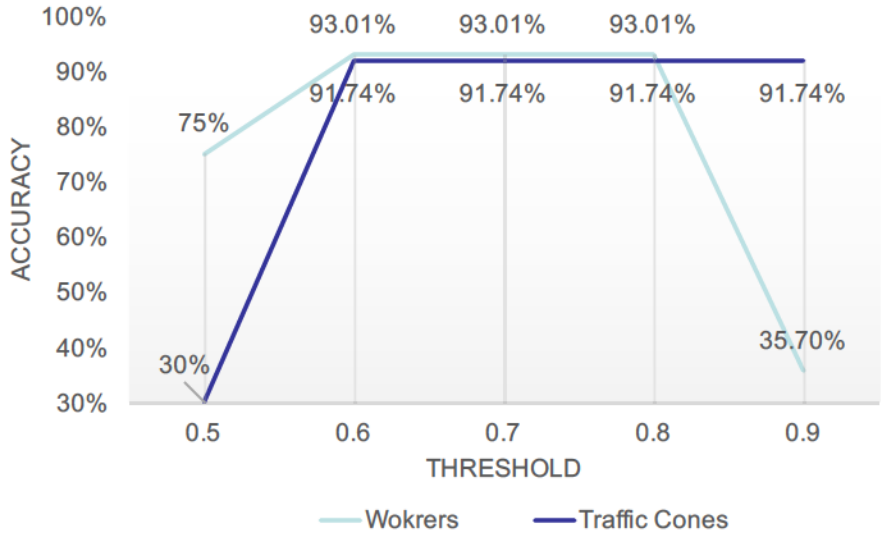

Figure 4.16 Matching accuracy with different thresholds

Another feature that influences the matching method is the view angle. The performance of the feature points matching between two opposite camera views was tested on another construction site, and the result is shown in Figure 4.17. The result demonstrates that even if the view angles of the camera views are very different, the proposed method

can still match the objects, as long as sufficient matched feature points are detected. However, if the view angle is too different that the overlap area of the camera views is very small, limited feature points on the construction site can be detected and thus the proposed method is ineffective. The proposed method is effective whenever the target objects are in the triangle meshes generated by matched feature points.

Figure 4.17 Matching with large camera view angles change

Moreover, the proposed method was compared with the recent research work conducted by Lee et al. (2016). It was found that the matching accuracy of the method proposed by Lee et al. (2016) with our tests set could only reach 53.57%. This is mainly because the method proposed by Lee et al. (2016) are not able to match construction objects when they are close to the same epipolar lines and/or partially occluded, as shown in Figure 29. These issues were well addressed in the proposed method.

Besides the workers, the proposed method can also be applied for matching other construction objects. Specifically, excavators, traffic cones are tested to evaluate the performance of the proposed method. Table 4.4 shows the matching accuracies of matching construction workers, excavators, and traffic cones. It could be seen that the matching

accuracies of other construction objects are also very high. With a high matching accuracy, the proposed method can contribute to automatic construction processes like object tracking, localization, etc.

	Correctly Matched Pairs	**Total Pairs**	**Accuracy**
Workers	213	229	93.01%
Excavators	40	40	100.00%
Traffic cones	100	109	91.74%
Total	353	378	93.39%

Table 4.4 Matching accuracy of different objects

4.3 Fall detection

4.3.1 Implementation

The proposed method has been implemented in the Python platform with the support of the OpenCV library (Beyeler, 2015). The library provides the critical algorithms, functions, and tools required for basic image processing operations. The method was tested on a Mac OS Sierra operating system. The hardware configuration for the test includes an Intel® Core™ i5 CPU @ 2.30 GHz, 8 GB memory.

4.3.2 Test results

Videos recorded in a lab in Concordia University by two GoPro4 were used for the tests. The cameras were placed on the top of the ceiling to simulate the view angle of cameras on the construction sites. The placement of the cameras is shown in Figure 4.18. Two videos captured by the two GoPro4 including several falls are used as the training set and test set of the artificial neural network.

Figure 4.18 Fall detection camera placement

To test the proposed method, totally 800 frames including 20 fall accidents are used. The features of each video frame are labeled with fall/not fall. 600 frames were used as the training set and 200 frames were used as the test set. For training, the features and labels of the training set were put into the neural network to find the weight and bias fit the model best. For testing, the features of the test set were put into the trained neural network and the fall accidents were counted and displayed. The result of the test is shown in Figure 4.19.

The green bounding box means that no fall accident was detected and the red bounding box means that a fall accident was detected.

Figure 4.19 Fall detection result example

In order to evaluate the performance of the fall detection method, the precision and the recall were employed. They were defined as follows:

$$\text{Precision}=TP/(TP+FP) \tag{4.3}$$

$$\text{Recall}=TP/(TP+FN) \tag{4.4}$$

where TP stands for true positive, which is the number of detected fall accidents, FP stands for false positive, which is the number of other actions detected as fall accidents, and FN stands for false negative, which is the number of not detected fall accidents.

The result of the proposed method is shown in Table 4.5. The proposed method has a precision of 83% and recall rate of 90%.

Actions	Number	Precision	Recall
Fall	10	83%	90%

Table 4.5 Fall detection results

4.3.3 Discussion

The test results demonstrate that the proposed method has a high recall rate and a low precision rate. The high recall rete means that most of the fall accidents can be detected correctly and the low precision rate means that some other actions will be detected as fall accidents as well. Higher precision or recall rate can be obtained by changing the threshold of the fall detection. In this research, if 6 in 10 successive frames were labeled as fall, a fall accident was detected by the proposed method. If a fall accident were defined when 5 in 10 frames were labeled as a fall, the recall would reduce but the precision would increase significantly. The result is shown in Table 4.6. The selection of the threshold should be decided based on the real application environment. If the focus is on the precision rate, the thresholds should be set higher, thus less non-fall will be detected as fall and hence false alarms are avoided. Oppositely, if the recall rate is concerned, the thresholds should be set smaller. Thus although some non-falls will be detected as falls, no fall accident will be omitted.

Thresholds	Precision	Recall
5 of 10 frames	62.5%	100%
6 of 10 frames	83%	90%

Table 4.6 Fall detection results with different thresholds

The parameters of the fall detection neural network influence the accuracy of fall detection. The most important parameter is the kinds of feature. According to Makantasis et al.'s work (2015), reduce one or more features will lead to the decrease of fall detection accuracy and add other features can not increase the fall detection accuracy apparently. Thus in this research, the six selected features including the height-width ratio of the bounding boxes, the change of height-width ratio between frames, the bounding ellipse angles, the change of bounding ellipse angle between frames, the actual height and the vertical motion velocity are used for fall detection.

Besides the parameters of fall detection neural network w and b, the hyperparameters of the neural network influence the performance of fall detection by influencing the training efficiency and the performance of the neural network. The hyperparameters include the number of hidden layers, learning rate, number of neurons and iteration epochs. To evaluate the influence of hyperparameters and to find the best selection of hyperparameters, different hyperparameters are tested in this research. Similarly, 600 frames were used as

the training set and 200 frames were used as the test set. The performance of the trained neural network is defined by the loss function, which is the difference between the output of the neural network and the ground truth:

$$Loss = -\frac{1}{m}\sum_{i=1}^{m}(y^i \log \hat{y}^i + (1 - y^i)\log (1 - \log \hat{y}^i)) \quad (4.5)$$

where y is the ground truth label of the dataset, $\hat{y}$ is the output of the neural network. The smaller the loss is, the more accurate the neural network is.

The number of hidden layers influences the performance of the neural network. With different number of layers, the neural network learns in different degrees. In this research, different numbers of layers from 1 to 4 were tested. The training results and test results are shown in Figure 4.20. It can be found that when the neural network only has 1 hidden layer, the final cost of the neural network is smallest. Also, as the training speed decreases when the number of hidden layers increases, 1 hidden layer neural network was selected for this research topic.

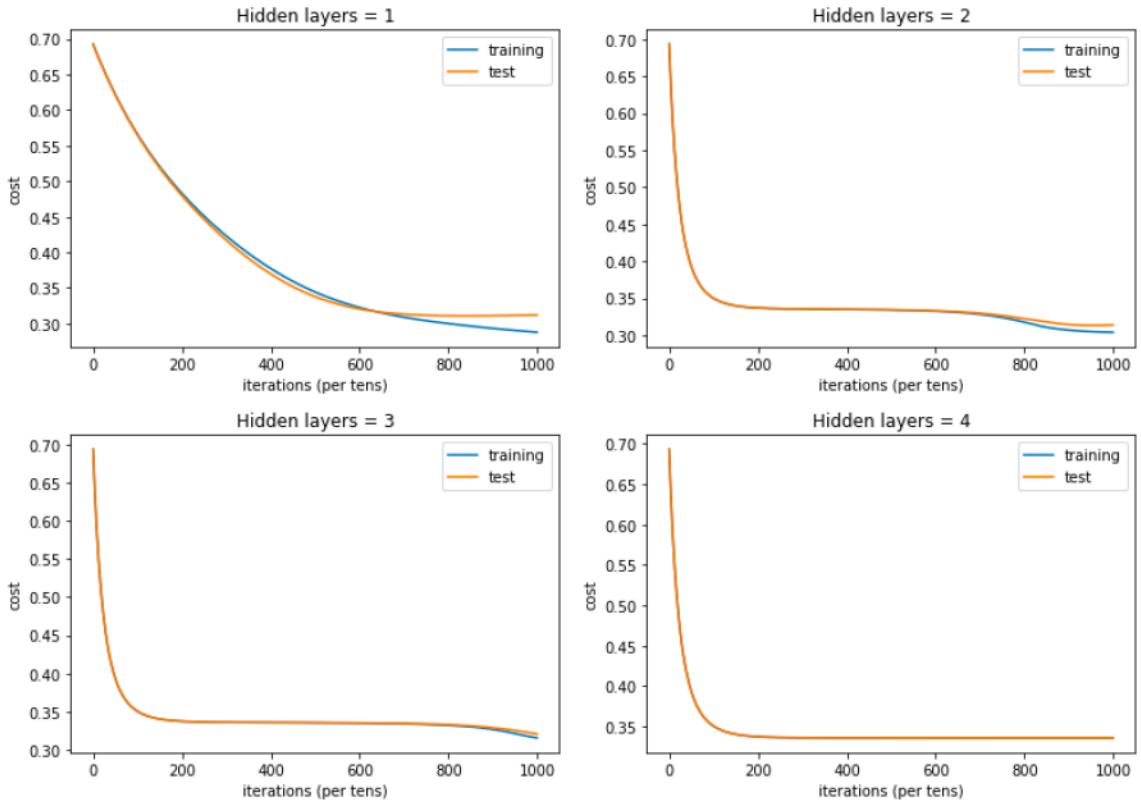

Figure 4.20 Training results with different number of hidden layers

The learning rate influences the speed of the gradient descent. If the learning rate is too large, the gradient descent will go over the minimum point. If the learning rate is too small, the gradient descent speed will be too slow that takes too much time for training. To find an appropriate value, the neural network is trained with different learning rates. The results are shown in Figure 4.21. From the figure, it can be found that if the learning rate is selected to be 0.001, the neural network will converge to the minimum at the fastest speed and won't be overfitting.

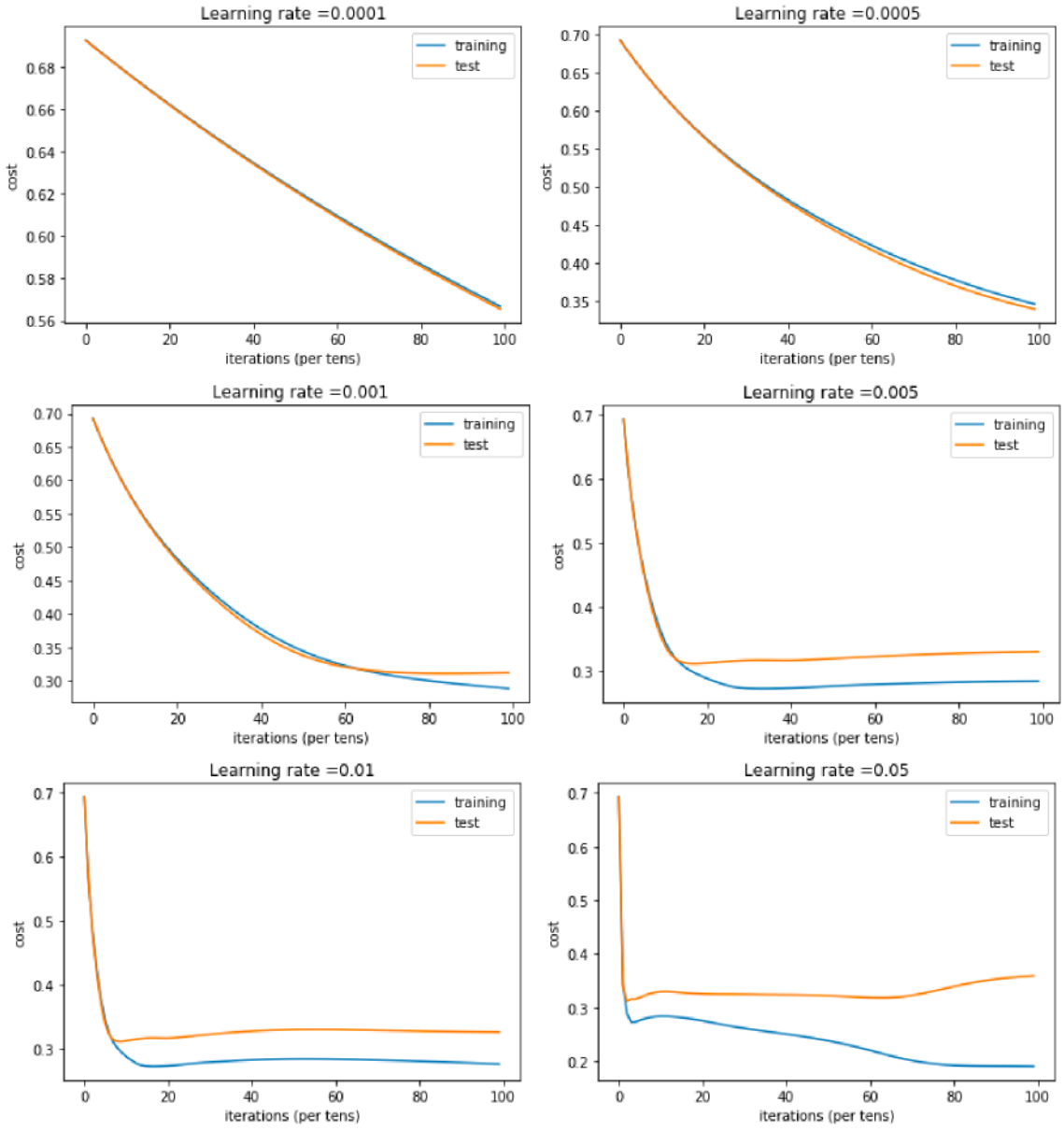

Figure 4.21 Training results with different learning rates

The iteration epochs influence the completeness of the learning process. If the number of epochs is too small, the gradient descent may not reach the minimum. If the number of epochs is too large, the gradient descent time will be too long to the final epoch and the neural network may be over-fitted. With learning rate set to be 0.001, the neural network is trained with different numbers of epochs. Figure 33 shows the training and test results

with diffident epochs. From Figure 4.22, it is easy to find that 7000 iteration epochs are the best choice when the learning rate is 0.001.

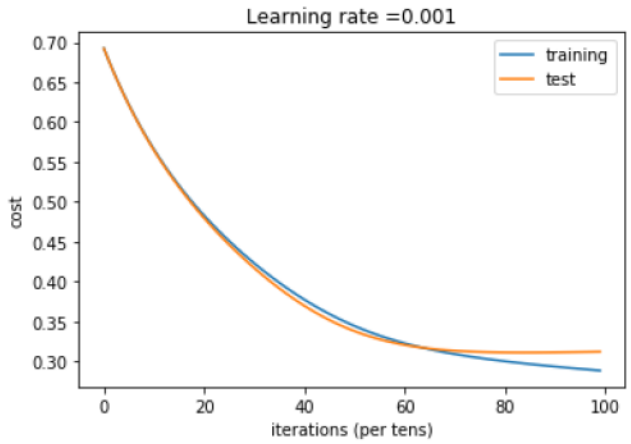

Figure 4.22 Training results with learning rate=0.001

The number of neurons influences the performance of the neural network. With different neurons, the neural network learns in different ways. In this research, different numbers of neurons from 10 to 500 were tested. The training results and test results are shown in Figure 4.23. It can be found that when the neural network has 50 neurons, the loss of the the neural network is the smallest and the training speed is fastest.

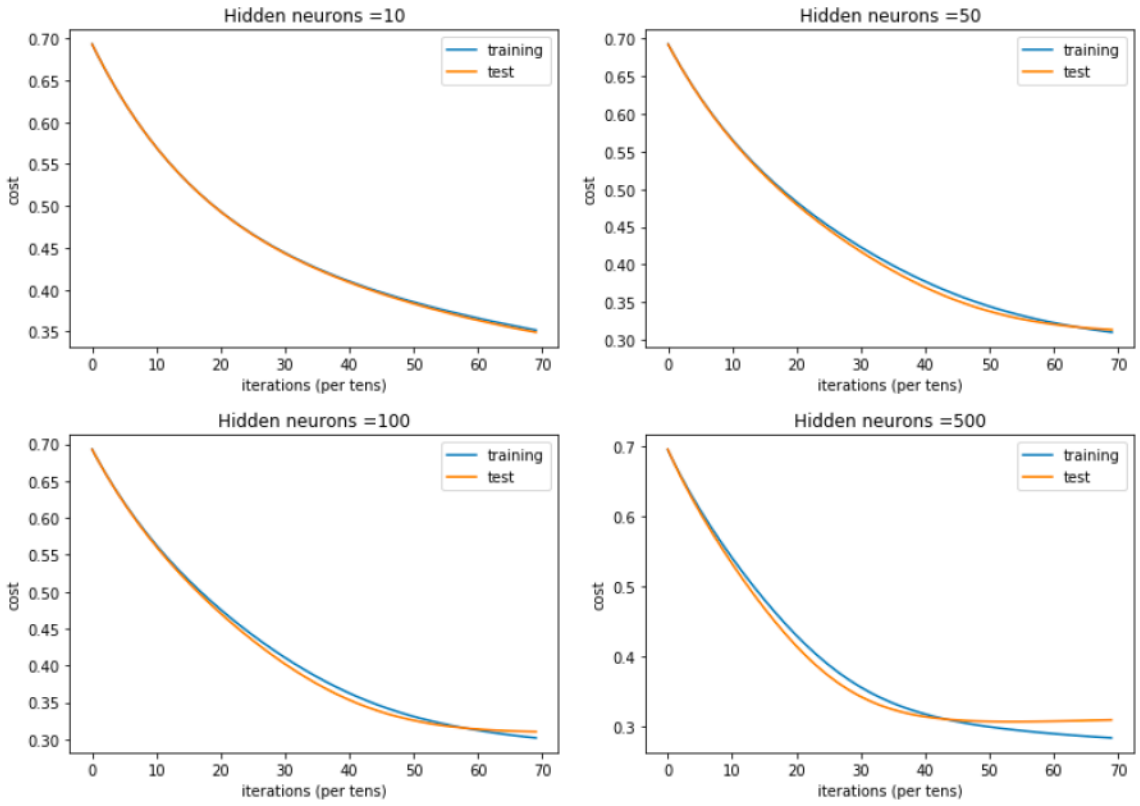

Figure 4.23 Training results with different hidden neurons

4.4 Contributions

This research contributed to the fundamental knowledge to retrieve construction workers' information on construction sites from videos. The information includes location, action, etc. This information can be used to support the construction safety management work on fall detection. Specifically, it can:

1. Provide the information about localization and fall detection without any wearable sensor or tag attached to the construction workers.

2. Reduce the time and cost of human monitoring by automatic localization and fall

detection with video cameras.

3. Report accidents timely to reduce the secondary harm caused by the delayed rescue actions.

4. Detect potential fall risks for unsafe areas analysis.

In addition to the fall detection, this research is expected to influence the research efforts in the area of automation in construction, and specifically facilitate:

1. Generating trajectories. With the proposed localization methods, the trajectories of workers and construction equipment can be obtained. This information can facilitate the automatic analysis of construction safety, productivity, etc.

2. Visual tracking. The proposed matching method may facilitate the visual tracking process. With multiple cameras and the proposed matching method, the difficulty of occlusion can be solved. Also, the proposed matching method makes tracking one object across multiple camera views possible.

3. Action classification. In this research, the artificial neural network was trained for fall detection. The neural network can also be trained for classifying other actions of workers with similar training process.

CHAPTER 5: CONCLUSIONS

In this chapter, the background and motivation of this research are reviewed first, then followed by the review of the proposed methodology. After the review, the discussions and conclusion of this research are presented. The recommendation and future works grow out of this research are discussed finally.

5.1 Review of background and motivation

According to statistics (BLS, 2015), the construction industry is one of the most dangerous industries and fall accident is one of the most frequent accidents on the construction sites. In order to reduce the losses incurred by fall accidents, it is important to detect them promptly and automatically. This way the rescue actions of the victims can be done immediately. Also, detecting the potential fall accidents can aid to safety management on finding dangerous areas on the construction sites. With potential fall detection results, corresponding measurements can be conducted to prevent fall accidents.

Currently, several kinds of methods have been employed for fall detection on the construction sites. These methods can be grouped into three categories including wearable sensor-based fall detection, ambiance based fall detection, and vision based fall detection. Those state-of-art methods have several limitations and issues when applying on the construction sites. The wearable sensors are not popular among workers because they don't

want their personal data being tracked. The ambiance based fall detection methods are not applicable because they are only capable of detecting fall accidents in a small area near the sensor. The vision based fall detection methods face the difficulties like occlusion, localization, processing speed, etc.

Although existing vision based fall detection methods have some limitations and issues, they still have significant potential to detect fall accident on construction sites. The motivation of this research is to solve the existing limitations and issues of vision based fall detection methods. The objective of this research is to propose a robust method which could detect and locate fall accident on the construction sites timely and automatically.

5.2 Review of methods

The proposed framework consists of three parts, localization, matching and fall detection. The objective of localization method is to find the working square of workers on the construction sites. When fall accidents happen, the injured workers can be found quickly based on the detected working squares. The objective of the matching method is to match the same workers in the different camera views. In order to overcome the occlusion, multiple cameras are used. With multiple cameras, one worker may be captured in several views, thus matching the same worker in different camera views can avoid the repetitive counting. The objective of fall detection method is detecting fall accidents based on videos

using an artificial neural network. The fall detection method can detect fall accident accurately and timely, and have the ability to detect potential falls.

The localization method locates workers by perspective transformation. Images coordinates of four corresponding point pairs are selected at first. The coordinate pairs are then used to calculate the perspective transformation matrix. With the transformation matrix, the map coordinates of the workers can be transformed from the workers' image coordinates in the video frames and then the workers' working squares can be determined.

The matching method is based on the spatial relationship of workers. Some matched points in different camera views are detected based on their feature first, then matched triangle meshes are generated by those points. The workers' triangle coordinates are calculated based on the generated triangle meshes. Multiple workers are then matched by combinatorial optimization based on the distance between their triangle coordinates in different camera views.

The fall detection method utilized an artificial neural network. Six features including height-weight ratio and its difference between frames, bounding ellipse angle and its difference between frames, as well as the actual height of workers and the vertical motion velocity, are used as the input of the neural network and the output is a binary classification result identifying fall or not fall. A set of videos including fall and other actions are used to train the neural network. After training, the neural network is able to detect fall accidents of workers on the construction sites from videos automatically.

5.3 Discussions and conclusions

The proposed framework was implemented on Python platform. The framework was separated into three parts for testing. The localization and the matching method were tested on the real construction sites and the fall detection method was implemented in the lab. The localization accuracy was 90%. The matching accuracy was 93.01%. The fall detection precision was 83% and the recall rate was 90%. The test results indicated that the proposed framework could detect fall accidents on construction sites accurately and timely.

In the localization method, workers were assumed to be on the same plane and then their working squares can be retrieved from videos. The localization result of workers whose working squares near the image centers were more accurate than the workers on the edges of images. Thus if a worker is captured in multiple camera views and the localization results are different, it is more accurate to consider the working square retrieved from the image, in which the worker is closest to the image center, as its coordinate.

For the matching method, the accuracy was influenced by the quality of triangle meshes, which relied on the feature matching method thresholds. The matching result will be more accurate if the triangle meshes are denser and matched more accurate. However, denser triangle meshes usually contain more incorrect matching point pairs. Thus a balance point of density and point matching accuracy should be found. In this research, it demonstrated that 0.7 is the best choice of the SIFT threshold value.

In the fall detection method, the parameters and the hyperparameters influenced the accuracy of the result. The parameter here is the number of frames n in 10 successive frames are detected as fall when the neural network indicates a fall accident. If n is large, the result's accuracy increase but recall rate decrease. If n is small, the result's recall rate increase but accuracy decrease. The hyperparameters (e.g., learning rate, epochs, number of neurons), are tuned in the research and the test results demonstrated that the best selections are 0.001, 7000, and 50 respectively.

5.4 Recommendations and future works

This research focuses on the vision based fall detection on construction sites. It consists of localization, matching, and fall detection. The experiment of this research provided valuable experiences and indicated the future works for better fall detection methods.

First, due to the safety constraint, the fall detection test was done in the lab simulating the environment of a construction site. For more convincible results, the fall detection neural network should also be trained and tested on data retrieved from real construction sites in the future.

Second, in this research, the fall detection is limited on a flat level surface, which is a 2D space. In the future, the fall detection may be extended to 3D space with other hardware

like depth cameras, stereo-camera, and/or with other fall detection algorithms using other features.

Third, the proposed fall detection neural network is a shallow neural network. With the development of GPU, it is now possible to work with the deep neural network for classification. The deep neural networks are usually more fast and accurate. Thus it may be a good research direction to build fall classifiers with deep neural networks.

Overall, this research study is just a tip of the iceberg in vision based fall detection on the construction sites. Much more information for automatic fall detection is still stored in the images and videos and waiting for us to retrieve.

REFERENCE

Alwan, M., Rajendran, P. J., Kell, S., Mack, D., Dalal, S., Wolfe, M., & Felder, R. (2006, April). A smart and passive floor-vibration based fall detector for elderly. In Information and Communication Technologies, 2006. ICTTA'06. 2nd (Vol. 1, pp. 1003-1007). IEEE.

Auvinet, E., Multon, F., Saint-Arnaud, A., Rousseau, J. and Meunier, J. (2011). Fall Detection With Multiple Cameras: An Occlusion-Resistant Method Based on 3-D Silhouette Vertical Distribution. IEEE Transactions on Information Technology in Biomedicine, 15(2), pp. 290-300.

Bahl, P., & Padmanabhan, V. N. (2000). RADAR: An in-building RF-based user location and tracking system. In INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE (Vol. 2, pp. 775-784). Ieee.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). Surf: Speeded up robust features. Computer Vision–ECCV 2006, 404-417.

Beyeler, M. (2015). "OpenCV with Python Blueprints." Packt Publishing, ISBN-13: 978-1785282690

Brilakis, I., Park, M. W., & Jog, G. (2011). Automated vision tracking of project related entities. Advanced Engineering Informatics, 25(4), 713-724.

Caldas, C. H., Haas, C. T., Torrent, D. G., Wood, C. R., & Porter, R. (2004). Field trials of GPS technology for locating fabricated pipe in laydown yards. Smart Chips Project Report, FIATECH, Austin, TX.

Carbonari, A., Giretti, A., & Naticchia, B. (2011). A proactive system for real-time safety management in construction sites. Automation in Construction, 20(6), 686-698.

Chae, S., and Yoshida, T. (2010). "Application of RFID technology to prevention of collision accident with heavy equipment." Automat Constr, 19(3), 368-374.

Chang, T. H., Gong, S., & Ong, E. J. (2000, September). Tracking Multiple People Under Occlusion Using Multiple Cameras. In BMVC (pp. 1-10).

Chang, T. H., & Gong, S. (2001). Tracking multiple people with a multi-camera system. In Multi-Object Tracking, 2001. Proceedings. 2001 IEEE Workshop on (pp. 19-26). IEEE.

Costin, A., Pradhananga, N., & Teizer, J. (2012). Leveraging passive RFID technology for construction resource field mobility and status monitoring in a high-rise renovation project. Automation in Construction, 24, 1-15.

Delahoz, Y. S., & Labrador, M. A. (2014). Survey on fall detection and fall prevention using wearable and external sensors. Sensors, 14(10), 19806-19842.

DoD, U. S. (2001). Global positioning system standard positioning service performance standard. Assistant secretary of defense for command, control, communications, and intelligence.

Dubey, R., Ni, B., Moulin, P. (2012). A depth camera based fall recognition system for the elderly. In: International Conference Image Analysis and Recognition, Springer, Berlin Heidelberg, pp. 106–113.

Ergen, E., Akinci, B., & Sacks, R. (2007). Tracking and locating components in a precast storage yard utilizing radio frequency identification technology and GPS. Automation in construction, 16(3), 354-367.

Fatal falls in the private construction industry, 2003–2013 : The Economics Daily. (n.d.). Retrieved September 24, 2017, from https://www.bls.gov/opub/ted/2015/fatal-falls-in-the-private-construction-industry-2003-2013.htm

Fatal occupational injuries by selected characteristics, 2003-2014. (n.d.). Retrieved September 24, 2017, from https://www.bls.gov/iif/oshwc/cfoi/all_worker.pdf

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381-395.

Foroughi, H., Aski, B.S., Pourreza, H. (2008). Intelligent video surveillance for monitoring fall detection of elderly in home environments. In: 11th international conference on computer and information technology, IEEE, pp. 219–224

Forsyth, D., & Ponce, J. (2011). *Computer vision: a modern approach*. Upper Saddle River, NJ; London: Prentice Hall.

Goodrum, P. M., McLaren, M. A., & Durfee, A. (2006). The application of active radio frequency identification technology for tool tracking on construction job sites. Automation in Construction, 15(3), 292-302.

Grau, D., Caldas, C. H., Haas, C. T., Goodrum, P. M., & Gong, J. (2009). Assessing the impact of materials tracking technologies on construction craft productivity. Automation in construction, 18(7), 903-911.

Han, S., & Lee, S. (2013). A vision based motion capture and recognition framework for behavior-based safety management. Automation in Construction, 35, 131-141.

Hartley, R., & Zisserman, A. (2003). Multiple view geometry in computer vision. Cambridge university press.

Hazelhoff, L., Han, J., With PH (2008). Video-Based Fall Detection in the Home Using Principal Component Analysis. In: International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, Berlin Heidelberg, pp. 298–309

Henriques, J. F., Caseiro, R., Martins, P., & Batista, J. (2015). High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3), 583-596.

Hildreth, J., Vorster, M., & Martinez, J. (2005). Reduction of short-interval GPS data for construction operations analysis. Journal of Construction Engineering and Management, 131(8), 920-927.

Hu, M. K. (1962). Visual pattern recognition by moment invariants. IRE transactions on information theory, 8(2), 179-187.

Ingram, S. J., Harmer, D., & Quinlan, M. (2004, April). Ultrawideband indoor positioning systems and their use in emergencies. In Position Location and Navigation Symposium, 2004. PLANS 2004 (pp. 706-715). IEEE.

Jang, W. S., & Skibniewski, M. J. (2009). Embedded system for construction asset tracking combining radio and ultrasound signals. Journal of Computing in Civil Engineering, 23(4), 221-229.

Joglekar, J., & Gedam, S. S. (2012). Area based image matching methods—A survey. Int. J. Emerg. Technol. Adv. Eng, 2(1), 130-136.

Jonker, R., & Volgenant, T. (1986). Improving the Hungarian assignment algorithm. Operations Research Letters, 5(4), 171-175.

KaewTraKulPong, P., & Bowden, R. (2002). An improved adaptive background mixture model for real-time tracking with shadow detection. Video-based surveillance systems, 1, 135-144.

Kartam, N. A., Flood, I., & Koushki, P. (2000). Construction safety in Kuwait: issues, procedures, problems, and recommendations. Safety Science, 36(3), 163-184.

Kaskutas, V., Dale, A., Lipscomb, H., Gaal, J., Fuchs, M. and Evanoff, B. (2010), "Changes in fall prevention training for apprentice carpenters based on a comprehensive needs assessment", J. Saf. Res. 41 (3) (2010) 221–227.

Kendzior, R. (2010). Falls aren't funny. Lanham, Md.: Government Institutes.

Khoury, H. M., & Kamat, V. R. (2009). Evaluation of position tracking technologies for user localization in indoor construction environments. Automation in Construction, 18(4), 444-457.

Kim, H. J., & Park, C. S. (2013). Smartphone based real-time location tracking system for automatic risk alert in building project. In Applied Mechanics and Materials (Vol. 256, pp. 2794-2797). Trans Tech Publications.

Konstantinou, E., & Brilakis, I. (2015). 3D Matching of Resource Vision Tracking Trajectories. In *Construction Research Congress 2016* (pp. 1742-1752).

Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., & Shafer, S. (2000). Multi-camera multi-person tracking for easyliving. In Visual Surveillance, 2000. Proceedings. Third IEEE International Workshop on (pp. 3-10). IEEE.

Lai, C. F., Chang, S. Y., Chao, H. C., & Huang, Y. M. (2011). Detection of cognitive injured body region using multiple triaxial accelerometers for elderly falling. IEEE Sensors Journal, 11(3), 763-770.

Lee, Y. J., Park, M. W., & Brilakis, I. (2016, January). Entity Matching across Stereo Cameras for Tracking Construction Workers. In ISARC. Proceedings of the

International Symposium on Automation and Robotics in Construction (Vol. 33, p. 1). Vilnius Gediminas Technical University, Department of Construction Economics & Property.

Lee, L., Romano, R., & Stein, G. (2000). Monitoring activities from multiple video streams: Establishing a common coordinate frame. IEEE Transactions on pattern analysis and machine intelligence, 22(8), 758-767.

Lee, D. T., & Schachter, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. International Journal of Computer & Information Sciences, 9(3), 219-242.

Li, H., Chan, G., Wong, J. K. W., & Skitmore, M. (2016). Real-time locating systems applications in construction. Automation in Construction, 63, 37-47.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In Computer vision, 1999. The proceedings of the seventh IEEE international conference on (Vol. 2, pp. 1150-1157). Ieee.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.

Lu, M., Chen, W., Shen, X.S., Lam, H.C., and Liu, J.Y. (2007) "Positioning and tracking construction vehicles in highly dense urban areas and building construction sites." Automation in Construction, 16(5), 647-656.

Luo, S., & Hu, Q. (2004, December). A dynamic motion pattern analysis method to fall detection. In Biomedical Circuits and Systems, 2004 IEEE International Workshop on (pp. 1-5). IEEE.

Makantasis, K., Protopapadakis, E., Doulamis, A., Grammatikopoulos, L., Stentoumis, C. (2012) Monocular Camera Fall Detection System Exploiting 3d Measures: A Semi-supervised Learning Method. In: European Conference on Computer Vision Springer, Berlin Heidelberg, pp. 81–90

Makantasis, K., Protopapadakis, E., Doulamis, A., Doulamis, N., & Matsatsinis, N. (2015). 3D measures exploitation for a monocular semi-supervised fall detection system., 75(22), Multimedia Tools and Applications, pp. 1-33.

Mastorakis, G. and Makris, D. (2012). Fall detection system using Kinect's infrared sensor. Journal of Real-Time Image Processing, 9(4), pp.635-646.

Mathie, M. J., Coster, A. C., Lovell, N. H., & Celler, B. G. (2004). Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. Physiological measurement, 25(2), R1.

Montaser, A., & Moselhi, O. (2014). RFID indoor location identification for construction projects. Automation in Construction, 39, 167-179.

Mubashir, M., Shao, L., & Seed, L. (2013). A survey on fall detection: Principles and methods. Neurocomputing,100, 144-152. doi:10.1016/j.neucom.2011.09.037

Ni, L. M., Liu, Y., Lau, Y. C., & Patil, A. P. (2004). LANDMARC: indoor location sensing using active RFID. Wireless networks, 10(6), 701-710.

Pang, Y., Li, W., Yuan, Y., & Pan, J. (2012). Fully affine invariant SURF for image matching. Neurocomputing, 85, 6-10.

Papadimitriou, D. V., & Dennis, T. J. (1996). Epipolar line estimation and rectification for stereo image pairs. IEEE transactions on image processing, 5(4), 672-676.

Park, M. W. (2012). Automated 3D vision based tracking of construction entities (Doctoral dissertation, Georgia Institute of Technology).

Park, M. W., & Brilakis, I. (2012). Construction worker detection in video frames for initializing vision trackers. Automation in Construction, 28, 15-25.

Park, M. W., Koch, C., & Brilakis, I. (2011). Three-dimensional tracking of construction resources using an on-site camera system. Journal of Computing in Civil Engineering, 26(4), 541-549.

Park, M. W., Makhmalbaf, A., & Brilakis, I. (2011). Comparative study of vision tracking methods for tracking of construction site resources. Automation in Construction, 20(7), 905-915.

Patel, S., Park, H., Bonato, P., Chan, L., & Rodgers, M. (2012). A review of wearable sensors and systems with application in rehabilitation. Journal of neuroengineering and rehabilitation, 9(1), 21.

Pilgrim, R. (2017) "Munkres' Assignment Algorithm", CSC 445 Readings <http://csclab.murraystate.edu/~bob.pilgrim/445/munkres.html> (August 3, 2017)

Pradhananga, N., & Teizer, J. (2013). Automatic spatio-temporal analysis of construction site equipment operations using GPS data. Automation in Construction, 29, 107-122.

Pratt, W. K. (2013). Introduction to digital image processing. CRC Press.

Priyantha, N. B., Chakraborty, A., & Balakrishnan, H. (2000, August). The cricket location-support system. In Proceedings of the 6th annual international conference on Mobile computing and networking (pp. 32-43). ACM.

Priyantha, N. B. (2005). The cricket indoor location system(Doctoral dissertation, Massachusetts Institute of Technology).

Rashidi, P., & Mihailidis, A. (2013). A survey on ambient-assisted living tools for older adults. IEEE journal of biomedical and health informatics, 17(3), 579-590.

Ray, S. J., & Teizer, J. (2012). Real-time construction worker posture analysis for ergonomics training. Advanced Engineering Informatics, 26(2), 439-455.

Saeki, M., & Hori, M. (2006). Development of an Accurate Positioning System Using Low-Cost L1 GPS Receivers. Computer-Aided Civil and Infrastructure Engineering, 21(4), 258-267.

Saidi, K. S., Teizer, J., Franaszek, M., & Lytle, A. M. (2011). Static and dynamic performance evaluation of a commercially-available ultra wideband tracking system. Automation in Construction, 20(5), 519-530.

Sathyanarayana, S., Satzoda, R., Sathyanarayana, S. and Thambipillai, S. (2015). Vision based patient monitoring: a comprehensive review of algorithms and technologies. Journal of Ambient Intelligence and Humanized Computing.

Sawyer, T. (2008) "South Korean research in electronic tagging is forging ahead." Engineering News-Record, The MacGraw-Hill Companies, April 2008.

Schwickert, L., Becker, C., Lindemann, U., Maréchal, C., Bourke, A., Chiari, L., ... & Bandinelli, S. (2013). Fall detection with body-worn sensors. Zeitschrift für Gerontologie und Geriatrie, 46(8), 706-719.

Shahi, A., Aryan, A., West, J. S., Haas, C. T., & Haas, R. C. (2012). Deterioration of UWB positioning during construction. Automation in Construction, 24, 72-80.

Simeonov, P., Hsiao, H., Powers, J., Ammons, D., Kau, T. and Amendola, A. (2010). "Postural stability effects of random vibration at the feet of construction workers in simulated elevation", Appl. Ergon. 42 (5) (2011) 672–681.

Skibniewski, M. J., & Jang, W. S. (2007). Localization technique for automated tracking of construction materials utilizing combined RF and ultrasound sensor interfaces. In Computing in Civil Engineering (2007) (pp. 657-664).

Soltani, M. M. (2013). Neighborhood Localization Method for Locating Construction Resources Based on RFID and BIM(Doctoral dissertation, Concordia University).

Statistics. (n.d.). Retrieved September 24, 2017, from http://awcbc.org/?page_id=14

Study Shows Interference with GPS Poses $96 Billion Threat to US Economy. (n.d.). Retrieved September 24, 2017, from http://www.forconstructionpros.com/business/press-release/10366178/study-shows-interference-with-gps-poses-96-billion-threat-to-us-economy

Wang, L. C. (2008). "Enhancing construction quality inspection and management using RFID technology." Automat Constr, 17(4), 467-479.

What Do Workers Really Think About GPS Monitoring? (n.d.). Retrieved September 24, 2017, from https://www.tsheets.com/gps-survey

Whitley, D. (1994). A genetic algorithm tutorial. Statistics and computing, 4(2), 65-85.

Woo, S., Jeong, S., Mok, E., Xia, L., Choi, C., Pyeon, M., & Heo, J. (2011). Application of WiFi-based indoor positioning system for labor tracking at construction sites: A case study in Guangzhou MTR. Automation in Construction, 20(1), 3-13.

Wu, G. E., & Xue, S. (2008). Portable preimpact fall detector with inertial sensors. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 16(2), 178-183.

Wu, W., Yang, H., Chew, D. A., Yang, S. H., Gibb, A. G., & Li, Q. (2010). Towards an autonomous real-time tracking system of near-miss accidents on construction sites. Automation in Construction, 19(2), 134-141.

Wu, B., Zhang, Y., & Zhu, Q. (2011). A triangulation-based hierarchical image matching method for wide-baseline images. Photogrammetric Engineering & Remote Sensing, 77(7), 695-708.

Yang, J., Vela, P., Teizer, J., & Shi, Z. (2012). Vision based tower crane tracking for understanding construction activity. Journal of Computing in Civil Engineering, 28(1), 103-112.

Yu, X. (2008, July). Methods and principles of fall detection for elderly and patient. In e-health Networking, Applications and Services, 2008. HealthCom 2008. 10th International Conference on (pp. 42-47). IEEE.

Zhang, Z., Deriche, R., Faugeras, O., & Luong, Q. T. (1995). A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artificial intelligence, 78(1-2), 87-119.

Zhang, T., Wang, J., Xu, L., & Liu, P. (2006). Using wearable sensor and NMF algorithm to realize ambulatory fall detection. Advances in Natural Computation, 488-491.

Zhu, Q., Wu, B., & Tian, Y. (2007). Propagation strategies for stereo image matching based on the dynamic triangle constraint. ISPRS Journal of Photogrammetry and Remote Sensing, 62(4), 295-308.