Kamal Dhakal

Development of Construction Safety System with Computer Vision

Metropolia University of Applied Sciences

Bachelor of Engineering

Degree Programme in Electronics

Bachelors Thesis

24 September 2018

Author Title Number of Pages Date	Kamal Dhakal Development of Construction Safety System with Computer Vision 49 pages 24 September 2018
Degree	Bachelor of Engineering
Degree Program	Electronics
Instructors	Jani Virtanen, Owner, MDS Finland Oy Matti Fischer, Principal Lecturer

This Bachelors Thesis explores the emerging field of automation technology to develop an advanced driver assistant system for the safety of construction workers in a construction site. The main goal of this study is to use different computer vision and machine learning algorithms to develop an autonomous system which alarms the driver of an excavator when construction workers are in near proximity of the vehicle.

The study evaluates different types of state of the art technologies for the development of vision-based detection and tracking system for the occupational safety of construction workers. The study includes the development of hardware system, and a software module for the operation of developed hardware unit. A stereo rig with two individual camera modules with high power LED rings around them is designed as a hardware unit. The software module is based on the algorithms of computer vision and machine learning for the effective operation of the stereo system to detect and track construction workers with safety vest on their body.

The developed system was tested in different indoor and outdoor environmental conditions. The real-time tracking and detection of persons with reflective vest on their body was successfully conducted. The camera module was placed in test environment and the persons with safety vest on their body were exposed to the field of view of camera module. The system performed extremely well in the indoor conditions and outdoor conditions with minimum sunlight facing the camera. The system showed satisfactory results in bright morning and evening conditions. The results were comparatively poor when a camera module was exposed directly towards the sun during the test conducted in mid-day of the summer season. The evaluation results were based on the real-time tests which were conducted in indoor hall and outdoor yard of a building in different times of the day. The system is yet to be tested in the real-world construction area with excavators and heavy machinery equipment.

The study concludes that the approach used in this study could be the effective solution to minimize accidents and risk hazards in the construction industry. The application of advanced machine learning technologies developed in the recent times will continue to push the development of this system further forward.

Keywords	Reflective Safety Vest, Construction Safety, Detection, Tracking, Near Infrared

Table of Contents

List of Abbreviations

ADAS	Advanced Driver Assistance System
BLOB	Binary Large Objects
BW	Bandwidth
BS OHSAS	British Standard Health and Safety Assessment Series
CMOS	Complementary Metal-Oxide Semiconductor
CNN	Convolutional Neural Network
dBs	Decibels
DBMS	Database management system. Software for maintaining, querying and updating data and metadata in a database.
EM	Electromagnetic
F-CNN	Faster Convolutional Neural Network
FOV	Field of View
GPS	Global Positioning System
GPIO	General Purpose Input Output
ISO	International Organization for Standardization
LADAR	Laser Detection and Ranging
LED	Light Emitting Diode
LIDAR	Light Detection and Ranging
MATLAB	Matrix Laboratory
NIR	Near Infrared
OHS	Occupational Health and Safety
ORB	Oriented FAST and Rotated BRIEF
ORM	Object-relational mapping. The set of rules for mapping objects in a programming language to record in a relational database, and vice versa.
OSHA	Occupational Health Safety Organization
PCB	Printed Circuit Board
RADAR	Radio Detection and Ranging
RFID	Radio Frequency Identification
RGB	Red, Blue, Green
ROI	Region of Interest
SIFT	Scale-Invariant Feature Transform
SNR	Signal-Noise Ratio
SSD	Single Shot Detector
SURF	Speeded Up Robust Features
SVM	Support Vector Machine

UWB Ultra-Wide Band

List of Figures/Table

1 Introduction

The proper management of health and safe working environment is the most important and crucial element to all the sectors of industries. The construction industry being particularly considered to be riskier with frequent and high accident rates, it demands the utmost priority in safety concerns. This industry makes up one of the major sectors of workforce and job market in the world and is considered as an important factor in driving the economy of the country.

Even though there has been huge development in state-of-art technologies in automation and robotics, it is disturbing to see the lack of sufficient and effective safety systems adjusted to the construction industry environment. There are various factors which have resulted this scenario. One of the factors is the lack of initiatives from the construction industry itself as there are financial issues related to it. Besides, the challenging and complex environment of the construction sites make it extremely difficult to directly apply the technology efficiently. However, many research works have been conducted in the last decade to introduce new technology for the implementation of efficient safety systems in the construction industry.

The main aim of this study is to test the various computer-vision based methods to develop an automated safety system which could be crucial in minimizing the accidents and casualties in the construction sector. Particularly, this study is focused on developing an Advanced Driver Assistance System that prevents the accidents that are being caused by the excavators in the construction sites. The study deals with the efficient detection and tracking of the construction workers on the specific parameters using the high definition industrial cameras.

The chapter 2 focuses on the construction sites vulnerability and the detailed description of excavators. The chapter has data and statistics on the various kinds of accidents that has occurred in the construction industry. The risks of excavators and the kinds of accidents related to the excavator are also thoroughly described in the chapter. Chapter 3 puts a light on the current state of art technologies used in the construction sector for the proper safety of the workers. Chapter 4 focuses on the literature of computer vision along with various methods and algorithms used in the field which has been pivotal in the development of autonomous systems. Chapter 5 gives the overall system principal and system approach that are applied in this study. Chapter 6 gives the overall description of hardware development used in the experiment. The chapter has all the detailed information for the propose of selection of the hardware and the details of each individual components. Chapter 7 gives the details of algorithms involved in the implementation of system procedure of the whole system and evaluates the tests results in different indoor and outdoor conditions. Chapter 8 discusses the results achieved in the study and possible future works which will be continued further. The conclusions of the study are placed in the chapter 9.

2 Literature Overview

2.1 Occupational Health

Occupational health according to National Institute of Environmental Health Sciences, USA, refers to the identification and control of the risks arising from physical, chemical, and other workplace hazards to establish and maintain a safe and healthy working environment [1]. These hazards may include chemical agents and solvents, heavy metals such as lead and mercury, physical agents such as loud noise or vibration, and physical hazards such as electricity or dangerous machinery. The vulnerabilities and loss associated with these hazards can be catastrophic and with risk factors at the workplace leading to cancers, accidents, musculoskeletal diseases, respiratory diseases, hearing loss, circulatory diseases, stress related disorders and communicable diseases and others [1]. Hence it is of paramount importance to create measures to ensure the safety against these hazards.

The occupational health is strictly linked to the dynamics of economic globalization. With the expansion of global market there is an expanding potential of risk in occupational health and safety. The increase in risk potential affects millions of workers and their families in the world. Globalization has promoted the introduction of market systems in many countries with the weakest capacities to create and enforce a regulatory system to protect workers and consumers [2]. As an outcome, large number of accidents are taking place in these countries at the workplaces. Various international organizations like World Health Organization (WHO), and International Labour Organization (ILO), have issued strict guidelines to ensure workplace safety and manage standard occupational health and safety programs. But they have met with little success and even more surprisingly, it has been ratified by only 37 of the 175 ILO member states. Only 23 countries have ratified the ILO Employment Injury Benefits Convention that lists occupational diseases for which compensation should be paid. OHS regulations cover only about 10% of the population in developing countries. These laws omit numerous major hazardous sectors like agricultural and domestic work, typically not considered “industries.” Only 5% to 10% of workers in developing countries and 20% to 50% of those in industrialized countries have access to adequate occupational health services. Although in a survey among International Commission on Occupational Health members from 47 industrialized and industrializing countries, 70% reported OHS being in place and 80% noted the existence of a national institute for OHS, the estimated coverage of workers with OHS services was only 18% [2]. According to Jorma Rantanen et al. [3] in their paper in National Center for Biotechnology Information Journal, a total of 31% of the respondent countries had more than 50% coverage of employees, but the majority had a lower percentage or did not provide data on their coverage. The emerging economies with a lot of workforce like India, China and Brazil had the largest share of employed population in numbers. However, the OHS coverage was found to be disappointing in those countries. The statistics showed that 10% of the population in each India and China had under coverage while Brazil had comparatively better statistics to show, being close to 26%. Small countries with comparatively less number of people in the global workforce and with a less impact on the global coverage, like Croatia, Finland etc. had a significantly high coverage ranging from 75 % to 97%. The countries that are considered more developed with big economies like, France, Italy and Japan, also showed encouraging results having as many as 75% or above workforce populace under coverage.

Finland gives occupational health a great importance and has regulated acts pertaining to it in the constitution to make sure about the existence of occupational safety and health divisions at the regional state administrative agencies, i.e. the OSH authorities in regional administration monitor compliance with OSH legislation. These OSH inspectors are entitled to access workplaces and conduct workplace inspections. By 2020, according to the Ministry of Social Affairs and Health Brochure, Finland [3], the incidence of occupational disease is expected to be down by 10%, frequency of occupational accidents by 25%, perceived physical strain by 20% while as perceived mental strain by 20%. The Ministry of Social Affairs and Health, working closely with other ministries, is responsible for OHS and has shown a lot of seriousness to meet the standards laid by international bodies and it has resulted some fascinating results. In last 50 years, the number of fatal accidents at work in Finland has fallen from over 400 to about 30, number of occupational accidents has now remained constant around 125 000 accidents per year for about ten years. The total number of suspected cases of occupational diseases has fallen to half of what it was eight years ago (2005–2013). There are some well-funded and renowned training and research institutes in Finland that are doing a remarkable work to collect updated information about the working life. They are working hard to improve the occupational safety and many other organizations have supported the research activities conducted on topics related to OHS. These include the Finnish Institute of Occupational Health, Finnish Work Environment Fund and Finnish Funding Agency for Technology and Innovation etc.

As mentioned by Statistics Finland, which is the only Finnish authority specifically established for statistics [5], the number of fatal accidents has been significantly decreasing since last two decades. The line graph of the data from last 25 years in figure 1 clearly proves that the massive improvements have been made in the Occupational Health and Safety Systems in Finland. The figure 2 shows the accidents involved in various industries and it has been clearly noticed in the figure that construction and transportation industry has extremely large number of accidents occurring in the workplace compared to other workplaces.

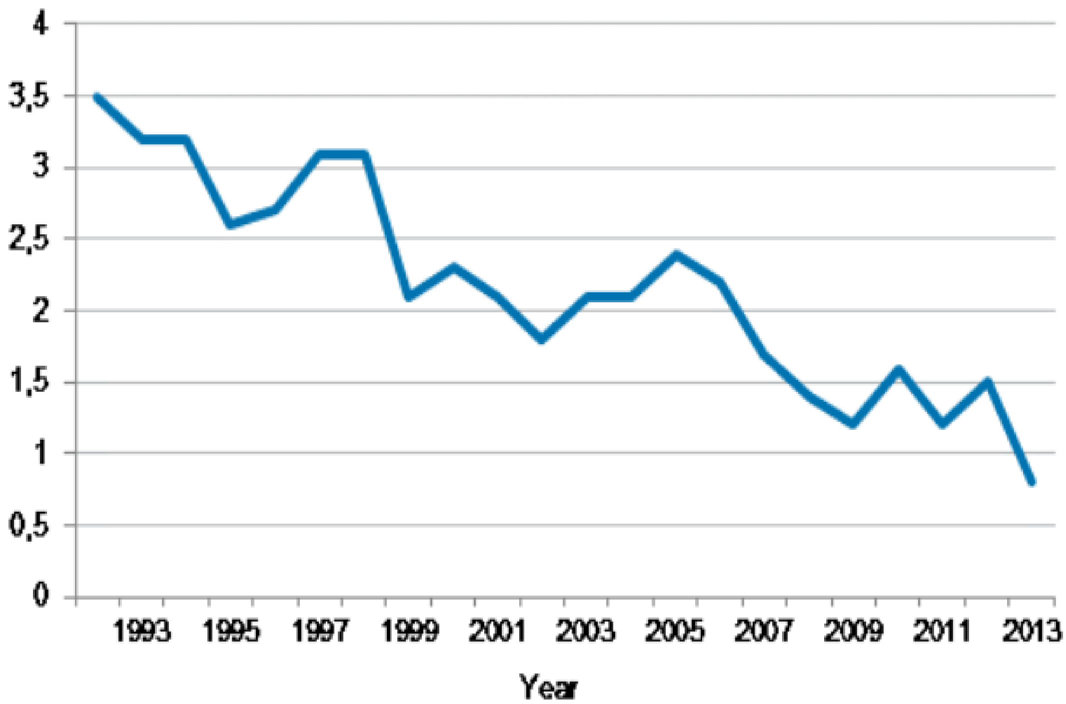


Figure 1. Fatal Accidents at work per 10000 wage and salary earners [4]

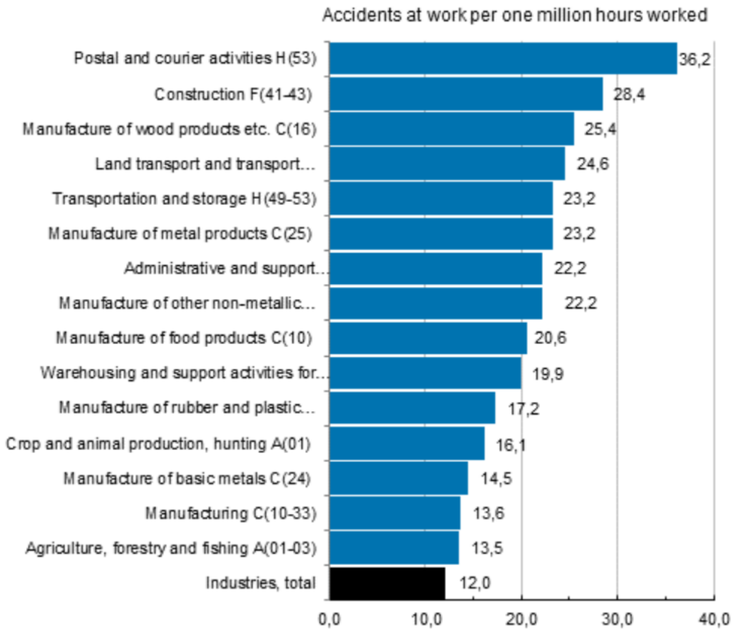


Figure 2. Data of accidents in different works in 2013 [4]

Occupational safety in Finland is defined by different frameworks and legislation. Every company must be aware of these frameworks to ensure safer workplace. Some of these frameworks are: BS OHSAS 18 000, ILO OSHM and Finnish legislation. Finnish legislation provides some acts that companies working with heavy equipment need to follow: Occupational Safety and Health Act (738/2002), Government Decree on the Safe Use and Inspection of Work Equipment (403/2008), Government Decision on the selection and use of personal protective equipment (1407/1993) , Government Decree on the Safety of Construction Work (205/2009) , Ministry of Labour decision on personnel rooms at construction sites (977/1994) , Council of State Ordinance concerning safety during blasting and excavation operations (644/2011) and Charger Act (423/2016). [6]

The table 1 shows the published accident cases involving excavator with help from Worker's Compensation Centre in Finland. It shows the record of the fatal accidents that were directly related to excavators.

Table 1. Total accident cases involving excavator in Finland recorded in Worker's Compensation Centre Finland [6]

Types of Accidents	Injury method	Number of cases	Fatality result
Resulted by movement of excavator (I)	Compressed.	19	Death or severe injured
Resulted by one or several tools of excavator (II)	Rotation parts.	11	Death or severe injured
Resulted by other problems	Compressed; Fallen down; etc.	16	Death or severe injured

2.2 Construction Site Safety: World and Finland

Construction sites are dynamic and critical sites where workers are involved in wide range of activities. The activities include moving objects, renovation and maintenance, exposure to heights, working in the periphery of heavy equipment thus exposing them to a threat of safety hazards. The most common hazards include falls, collisions, being caught between objects, electrocutions, and being struck by objects, hence causing injuries and deaths.

To reduce or eliminate the construction-related hazards, awareness should be created among workers and the supervisor. Similarly, new workers and employees should be specially trained and educated before the start of the job. Proper safety equipment should be assigned to the all the workers. A supervisor at every construction site should be educated and trained well to enforce all the safety standards and compliance with regulatory measures.

In 2014, the US had 4679 fatal occupational injuries, an incidence rate of 3.3 per 100000 full-time employed workers. In the same year, fatal work injuries in construction and extraction occupations increased by 5%. One in five deaths of workers in 2014 were related to construction work. Construction industry shares about 6% of total workers in USA but a whopping 17% of the fatalities (which is the largest number of fatalities reported for any industry sector) belongs to the construction industry. In the United Kingdom, the construction industry is responsible for 31% of fatalities at work and 10% of major workplace injuries. In South Africa there are 150 fatalities and approximately 400 injuries each year related to construction sites. In Brazil, the incidence rate for all occupational fatalities is 3.6 per 100,000. However, there is little to no information regarding construction fatalities could be found in Asia, South American continent, Africa, and the Antarctic region.

Finland is the frontrunner when it comes to safety on the construction sites. Finnish legislation forces construction sites to perform a specific safety measurement at the sites on a weekly basis. This is meant to be a pre-emptive action to prevent potential safety hazards and to make sure that site workers get notified of anomalies at the site. In 2008, number of fatalities (per annum per 100,000 workers) was 5.9 according to the data presented in the Brochure of Ministry of Social Affairs and Health [7, 9].

According to the European commission [8], harmonized product standard or ETA basic requirements for construction workers include: mechanical resistance and stability, safety in case of fire, hygiene, health and the environment, safety and accessibility in use, protection against noise, energy economy and heat retention, sustainable use of natural resources, etc.

The Ministry of Social Affairs and Health is responsible for the enforcement and development of occupational safety and health and for the preparation of related legislation. The ministry acts in close cooperation with other ministries, namely, Ministry of Labour, Ministry of Trade and Industry, Ministry of the Environment, Ministry of Transport and Communications and with various special authorities on matters associated with occupational health and safety. The Ministry of Labour prepares statutes in the field of labour law as well as deals with matters concerning employment, unemployment benefits and employees' pay security. Settling labour disputes is also the responsibility of the Ministry of Labour [9].

To develop the working life and workplace welfare in Finland, national programmes have been implemented and are being implemented. The programmes are coordinated with concerning ministries, labour market parties and other players acting in close cooperation. The most important programmes are:

a) Finnish national workplace development programme (1966- present)

b) National occupational accident prevention programme 2001 – 2005

c) National action programme on prolonging work careers, coping with work life and rehabilitation (Veto programme) 2003 – 2007 [9]

2.3 Theoretical Background on Excavators

Excavators are a popular heavy-duty equipment designed and used especially for excavating from earth and moving large objects like rocks. They are heavily used in construction of roads, buildings, demolition of constructed structures, moving of material like logs, boring, excavation of sand from river beds and much more. Figure 3 shows the standard excavator and its main labelled parts are described in the points below:

a) A heavy-duty engine, most commonly running on diesel, drives the hydraulic pumps that provide oil at a high pressure to the slew motor, rams, track motors, and several accessories summing up to controlling the boom and powering the undercarriage.

b) An undercarriage is the backbone of the excavator that supports the weight of the machine, allows traction on unstable surfaces and to move. It has wheels or tracks for mobility.

c) A boom is the large arm of the excavator that is hydraulically articulated. It’s one end is fixed to the excavator and the other end does the apparent work and it can be fit with different tools such as buckets, claws, compaction wheels, jackhammers and other industry specific tools to cater to specific needs.

d) A 360-degree rotating cab housed on undercarriage and wheels where the operator controls the machine. Nowadays, these cabs are comfortable, let a very less noise into the cab, and are ergonomically and aesthetically great.

e) A dipper which is the arm above the bucket

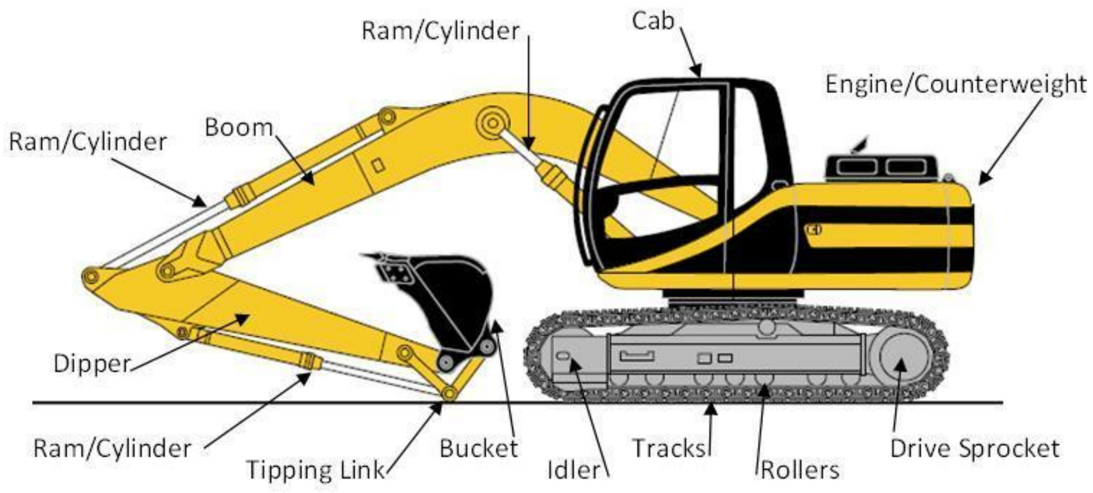


Figure 3. A standard labelled excavator [6]

Since the vision levels of excavator cabs can be limited, there can be a lot of possible risks involved with a working excavator, both to the people working near it and the operator. Some risks involved with the excavator are: (1) Striking a pedestrian especially while reversing (2) Slewing which means tapping a person between excavator and a fixed structure (3) Inadvertent falling of an attachment or an accessory that might hit a person. Operating an excavator must be carried out after a proper training and supervision and it is strictly required that all the workers must participate in the safety training. Moreover, the working area surrounding the excavators must properly be marked with safety symbols to prevent the possibility of any kinds of accidents.

3 Related Works and Current Practices Construction Safety

With the advancement of automation technology in recent years, state-of-art technologies are employed for the efficient monitoring of the workers working in the construction sites. Some of the safety systems that have been implemented in the construction sector are described in this chapter.

3.1 Global Positioning System

The GPS system was first developed by United States Department of Defense for the military navigation purposes, but it was made available for civilian use purposes in 1983

[10]. Currently, this technology has been utilized by wireless and telecommunication companies for the navigation and mapping. This has also been utilized in construction sites for monitoring purposes. This technology is based on a set of satellites that revolve around the earth and it is operated 24 hours a day. The GPS sensors are fitted in certain construction equipment or clothing in case of humans and the central module is established to communicate with the all the sensors. This has been effective in outdoor environment for tracking of construction materials against theft and robbery. It has also been effective to track the humans when they are lost in some catastrophes as the location can be tracked via satellites. However, the real time near collisions are difficult to prevent with this technology as the average error of this GPS is around 8 meters according to the official governmental agency to provide information of GPS [11]. In the fast-moving conditions as in excavators, this level of accuracy is not enough to prevent the collisions.

3.2 Radio Frequency Identification

RFID-based systems are also popular in construction projects. The system consists of RFID tags, a reader, and a data management system. The RFID tags are placed in persons clothing particularly, helmet, belt or vest that are detected by the tag reader in the equipment. This has been effective in near end collisions in cranes and excavators. This technology is extremely promising as there are minimum chances of false alarms [12]. However, it is extremely expensive technology and many construction companies are reluctant to use it citing the economic factors.

3.3 Ultra-Wide Band (UWB)

UWB is another type of technology that is currently in use of the monitoring of the worker in various indoor and outdoor application. It is based on radio technology that can be applied to short range communications. Usually the wireless and telecommunication systems transmit radio signals at around 30 – 5000 Hz of bandwidth. On the contrast, UWB transmits signals at very high bandwidth from 500 MHz to several GHz at very low power spectral densities. The interference of the signals is minimal because of extremely wide bandwidth which makes this technology attractive. The wide range of higher operational frequency and the low power density of signals limit the system to operate in shorter distance. The short length of the signal pulse helps in maintaining the accuracy during the measurement of distance. UWB uses very short signals of around 1 ns which allows the tracking of distance with great accuracy. The use of this technology was applied on construction for the work zone safety and material tracking by Teizer et al. in their paper [13].

Figure 4 shows the range of frequency used by other wireless communication system and the UWB network.

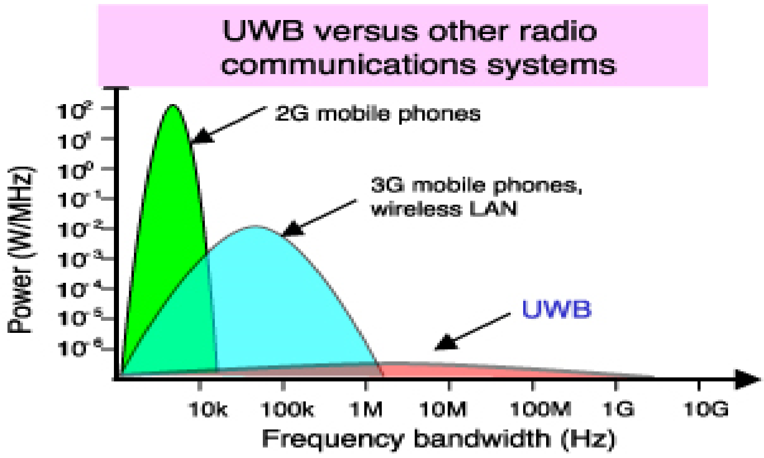


Figure 4. Operational BW of UWB vs Other Radio Communication Systems [14]

3.4 Optical Sensor Based Tracking

The Laser based technologies are also currently in practice in construction industry in challenging working environment. Laser Scanners along with LADAR based systems are highly effective in 3D imaging along with the accurate distance measurement of the of the object. The depth map generated by the laser-based systems are highly accurate and are extremely useful for short range safety operations. However, the high cost of LADAR technology has been the massive drawback in its effective implementation in all kinds of construction projects. Moreover, it has also been reported of reflectance of surface as mentioned by the Teizer et al. in their published journal [15]. RADAR based safety systems have been also implemented in various vehicles and automobiles used in construction areas. However, the challenges of massive false alarms have prevented this technology for the effective implementation.

3.5 Vision Based Detection and Tracking

The development of advanced algorithms on Computer Vision and Machine learning has encouraged many research works in the field of vision-based object detection and tracking. The advent of effective and robust image sensors and the micro computer technology have been encouraging factors in this process. Several object detection algorithms have been introduced to work in real time. In vision-based object detection and analysis, the mechanism uses vision as the main sensor for the machine to work in real time [16]. Prominent application of vision-based detection and tracking can be seen in self driving cars, robotics, large scale industrial inspection etc. The survey of vision-based vehicle-based detection was carried out by J. Joseph Antony et al. in their paper where they have analyzed Histogram of Oriented Gradients (HOG) of the pixels based extraction and Haar-like-features based detection algorithms [17]. In their paper, they have mentioned that the vision-based detection in vehicles have reduced the number of traffic accidents. They have also highlighted the need of ADAS in automobile which are efficient and reliable.

Vision based object detection and tracking use features such as shape, color, motion etc. of the object for the detection purpose. Haar like features was presented by Viola and Jones in their paper in which they utilized shape features for object detection using the distribution of intensity gradients [18]. Similarly, Histogram of Oriented Gradients was introduced by Dalal and Triggs in CPVR 2005 paper [19]. They utilized the distribution intensity gradients inside the rectangular region of the image modeling with histograms. They used HOG features trained with linear SVM for the efficient human detection. Similarly, Zhu et al. in their paper improved the human detection by Adaboost feature selection [20]. The paper mentions the use of cascade of HOG for the fast human detection with greater accuracy.

Similarly, color has also been the useful feature for the efficient object recognition. Swain and Ballard in their paper in International Journal of Computer Vision, 1991 have used the effective use of color histogram as a description of the object for the effective recognition [21]. The use of the RGB color histograms for the safety vest detection for the construction safety was implemented by Man-Woo Park in this paper on automated 3D vest tracking [22]. In the paper, author has implemented the stereo system for estimating the 3D position of the object. He computed the color histograms of the green safety vest to detect and track the workers wearing green safety vest in their body. The author has claimed to have obtained the encouraging test results. The vision-based approaches of object detection have been utilized by the autonomous cars for the pedestrian protection system [23]. The camera modules most commonly used by the automotive cars are Thermal Infrared(TIR) and the Near Infrared(NIR). The Near Infrared camera is used in this study for the hardware setup of the experiment. Thermal cameras use the temperature of an object to form an image. The description of the NIR image sensors is explained in the hardware setup in chapter 6. The use of NIR cameras was implemented by Rafael Mosberger et al. for the detection of construction workers with the safety vest on their body. The experimental works for the detection and tracking of construction workers with vision-based approach are presented their journals [24, 25, 26, 27]. This study is closely related to their works as we have implemented similar approach to the problem statement. However, the algorithms used in this study are much simpler and hardware setup is comparatively less expensive.

The development of vision-based detection and tracking has seen the new high in recent times with development of robust machine learning algorithms. The deep learning based object detection models are being practiced compared to the conventional vision technique of object detection and tracking. CNN based object model called YOLO was introduced by Joseph Redmond et al. in their paper on IEEE conference on Computer Vision and Pattern Recognition [28]. Many experiments have been conducted using this algorithm for real time object detection [29, 30] and the authors have claimed this algorithm to be robust in speed. However, significant amount of errors in the localization limits this algorithm in the application of safety applications. Similarly, the other popular frameworks of object detection are Faster-CNN, Single shot detectors. The deep learning models use the base network which are usually pre-trained to perform classification on the image datasets. The common base networks used for image classification using deep learning are:

- VGGNet
- Resnet
- Mobilenet
- DenseNet

The detailed description of Deep Learning based models and the frameworks is not the primary goal of this research as these are part of the future works which will be more discussed in the chapter 8. The fundamental principle of deep learning and the models which are applied in real world conditions are presented in the references [31, 32].

4 Computer Vision

Computer Vision is the sub-field of Artificial Intelligence that works on computers to see, identify and analyze the image in the similar way as the human vision system. Humans see the object through the eyes and analyze objects in the field of view with the help of neurons of brains evaluating the properties in a rapid procedure. The capacity of analyzing and interpreting things that we see is a result of the continuous learning of things by the brain since our birth. Accordingly, our biological visual concept is applied to computers artificially to give them capability to see and learn to analyze the image similar as humans. This process involves the image processing steps and the various machine learning algorithms processing in a synchronized manner.

The understanding of image by computers is not simple as biological human understanding and it has to go through many complex procedures of optical-vision, interpretation and analysis. This involves the image abstraction (image clicks or frames in a video), processing steps, detecting features, extracting features and analyzing the image based upon those extracted features with a suitable machine leaning algorithm. These processing steps could involve individual images, or the sequence of images represented as frames in the video. Human beings have the natural tendency of seeing the images as the real representation of texture, color, shape, sizes along with the emotions involved in it but that is not the case with computers. The computers see the images as a huge matrix represented in numbers of pixel values and the pixel value in each co-ordinate represents the type along with the shape and size of the image. More precisely, image is represented in terms of array of pixels with different intensity value and the position of these pixels are given by representing them in coordinate system. Image size 256 x 256 represents array of 256 rows and 256 columns with total number of 256 x 256 = 65536 pixels.

A computer with algorithms of image processing and machine learning analyzes these matrices of numerical values to define the features and interprets as a certain object based upon its properties of the features. The most challenging part of a computer analysis compared to humans is that it is entirely dependent on the numbers in the matrix and those numbers do not define the emotions that human beings feel about the object. For example, the computer does not understand a dummy and a real person and a toy dog or a real dog provided they appear similar in the physical appearance.

4.1 Basic Understanding of Image

Images are the fundamental blocks of building any computer vision application. In simple words, the image is an exact representation of any person, object or thing. The images are observed when the light hits and object and the rays are reflected to eyes. There cannot be existence of any images in the dark. The detailed process of biological phenomenon of vision and the nervous system involving the total optical phenomenon is out of the scope of this report and the images that will be discussed in this report will be digital images obtained from the camera. The images are representations of set of pixels and pixels are then considered as the intensity of light that appears in a given place at the image. There are different types of images based upon pixel representation of grayscale or color channels. Grayscale images represent each pixel within intensity range of 0 - 255. The number 255 is the highest value and it represents white pixel whereas 0 is the lowest value representing black pixel. Binary images are represented in two pixels with intensity of each pixel being either 0 or 255. Color is represented in the image in 3 channels which are represented in Red, Blue and Green (RGB). Each pixel in color image consists of the 3 individual values of RGB channels in the form of tuple. For example, the white pixel in a color image is represent as (255, 255, 255) and black is represented as (0, 0, 0) and three numbers inside the parentheses are the representation of (red, blue, green) values. The figure 5 represents the grayscale format of the image in which each pixel is distributed in between 0 and 255. The figure 6 represents the binary format of the same image in which pixels are represented in either black (0) or white (255). In the binary image shown in figure 6, the lighter pixels are set as 0 and the heavier pixels are set as 255.

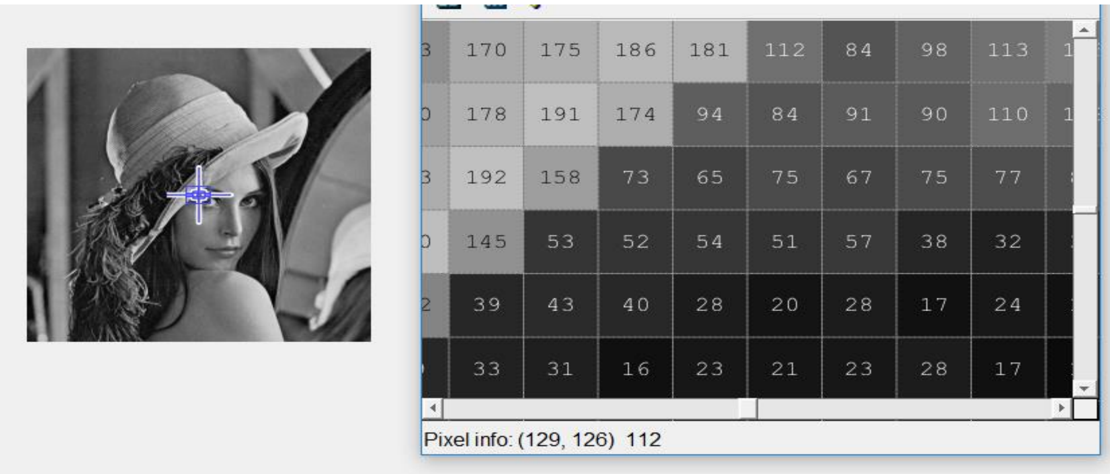


Figure 5. Standard Grayscale representation of Image (Pixels generated by MATLAB for standard test image)

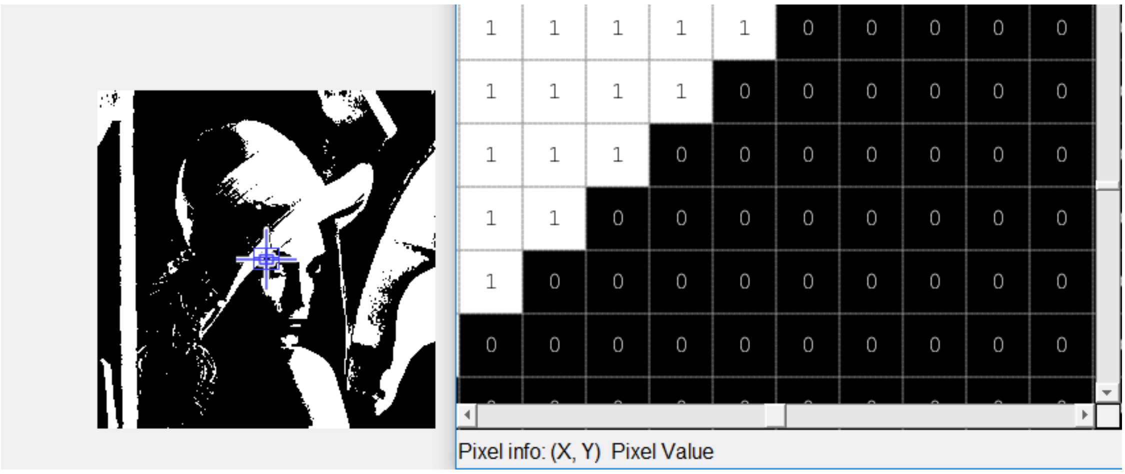


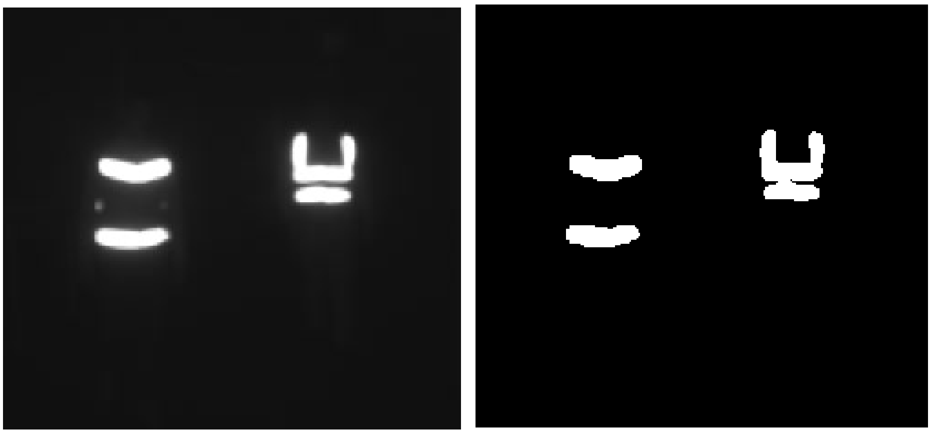
Figure 6. Standard Binary representation of Image (Pixels generated by MATLAB for standard test image)

4.2 Image Segmentation

Image Segmentation is a process of partitioning a digital image into multiple segments that share similar attributes (like color, texture, brightness) so that it is simplified and can be interpreted into something more meaningful [36]. It is one of the first processes of image processing in the field of image analysis and computer vision. The segmentation of input image from the camera in this experiment is done with the process of local adaptive thresholding which results into the formation of distinct blobs of the reflective clothing with bright pixels. Local adaptive thresholding adjusts the thresholding value automatically according to the changing lighting conditions.

Thresholding is just a simple way of converting a RGB or grayscale image into a binary image. Binary image is simply an image with a black and white representation. In technical terms, it is an image with only two pixels values i.e. 0 and 255. Generally, when image processing is carried out to perform segmentation with thresholding, color image is converted to grayscale and then the grayscale is converted into binary with different methods of thresholding techniques. During the thresholding process, the grayscale image which is represented in pixel value 0 to 255 is provided a threshold value and all the pixels below the value is settled at either 0 or 255 and vice versa. The different thresholding types include simple thresholding and adaptive thresholding. Simple thresholding means that we manually provide the threshold value whereas the adaptive thresholding adjusts the threshold automatically by analyzing the neighborhood pixels. The adaptive thresholding can be carried out in global and local parameters. The global parameters represent the whole image representation whereas the local parameters include the parameters in the region which we are interested in the image. In this experiment we have experimented different methods of adaptive thresholding procedure to quantify the portions of image and figure out the most suitable one for the segmentation.

Figure 7(a) is the raw taken from the camera module used in the experiment. The image was taken at indoor conditions during the bright day. The raw image taken from the camera was segmented into two background and foreground region with local adaptive thresholding. The thresholded image is shown in 7(b).



(a) (b)

Figure 7. (a) Raw flash image from camera. (b) Image (a) after thresholding

4.3 Contour Extraction

Contours are defined as the boundary of an object in an image. Contours are one of the most important techniques to perform processing of computer vision applications of object detection and tracking. Contours are usually detected with the edge detection algorithms. The edges in the images are defined as the regions in an image where there is the sharp change in the value of intensity of the pixels.

The contour properties are often overlooked in the processing of images in the recent times as the modern machine learning algorithms have developed significantly to compute complex problems. The advanced machine learning algorithms perform all the processes of segmentation with the hidden layers of neurons in a network, but the wise use of the contour properties often makes the processing much simpler thus resulting into faster and efficient computation of an algorithm. The most common contour properties are 1. Centroid 2. Area 3. Parameter 4. Bounding Boxes 5. Rotated bounding boxes 5. Minimum enclosing Circle 7. Fitting an eclipse [33].

In this experiment, the use of contour properties has been implemented to find our object which is a reflective material of the safety vest in an image. We have experimented with several contour properties and implemented the best suitable properties computing the centroid of the area of the resulting bounding boxes drawn over the blobs originating from the reflective objects.

The centroid is also termed as “Center of Mass” which is center (x, y) co-ordinate of the center of the certain area and is calculated based on mean position of all the co-ordinates inside the contour of the object. The bounding boxes are rectangle boxes which are drawn enclosing the entire contoured region of an image. However, it is not rotational invariant box so that the rotated contour properties are not equal to the original contour.

4.4 Localization

Localization is the process of finding the position of object in the image. Recently with the development of deep learning-based algorithms localization is performed with the classification of image with the pre-trained model of huge dataset. The hidden layers inside the CNN perform the operation of localization of image by training the dataset with the bounding box in the object. However, the use of reflective vest in this experiment has resulted the peculiar and distinct pattern of white blobs obtained from the safety vest. Thus, simple contour-based properties are applied for the localization of the object in an image. The contours of the vest are not compact together as there are two stripes of reflective material apart from each other which results into the formation of two distinct blobs in a safety vest. Here, we are interested in combining those two patches into a single unit as in the figure 8(b) using a series of morphological erosions and dilations are performed. Then, a bounding box is drawn over a newly formed single patch as seen in figure 9.

Erosion is a morphological operation in image processing in which it erodes the foreground object and decreases the size of the contour. In a more precise meaning, it discards the pixels which are near the boundary of an object. The application of erosion helps us to remove small noises around the contour boundary of an object to get a smoother distinct object pattern. It also helps to filter out the distractions of small false blobs in the segmented image. In this experiment, we have applied the process of erosion to remove the noise and small reflective points from the external light and small reflective distraction blobs so that reflective pattern of the object will be more distinguished and traceable.

Dilation is one of the operations in image processing which helps in joining the broken parts of the image by growing the pixels in the foreground patch. Normally the process of dilation is carried out in binary images. In this experiment, the raw image after the process of segmentation has resulted into binary image with the reflective patch represented in white and the non-reflected patch represented in the black pixels. Dilation has made the processing easier for localization by joining the reflective patch together to form as a single unit and represent whole vest as a one object in most of the cases. The resulted image after the process of dilation can be seen in figure 8(b).

The figure 8 below in the left is the raw frame captured from individual camera. The camera module which was used to capture this frame is described in the hardware setup in chapter 6. The person is in backward position from the camera view and the flash of LED has resulted the distinct pattern of the reflective vest. The resulted image after the local adaptive thresholding is represented by the right image in the figure 8.

(a) (b)

Figure 8. a. Raw image taken with flash / b. Resulting image after thresholding and dilation

Figure 9. Bounding box around the safety vest

4.5 Image Feature

The image feature is a region of an image which is unique and easily recognizable. The image features are represented in global and local formats. The global image features are those features which deal with the entire image whereas the local features are those features which deal only with the interesting regions of an image. The computation of local features is mainly performed in two steps namely 1) Keypoint Detection and 2) Feature Extraction.

4.6 Keypoint Detection

Keypoint detection is the process of detecting important regions of an image. The important regions are those regions in an image which attract more attention with unique and distinct features in them. The region of an image with these distinct properties gives the unique identity to the image. For example, the image of a plain surface in the universe would not help us to estimate the location of a place in an image but an image with water and green trees would help us recognize that the image belongs to a place in the earth. The presence of water in the image in the example is a distinct feature and can be referenced as a keypoint in the image. Likewise, for computers the interesting regions in an image are the references which help computers to better understand the information in an image. However, digital images are interpreted in terms of pixels and pixels are distributed in different orientation in different images. The most common keypoints that can be represented in digital images include edges, blobs, corners or the region in an image with a uniform intensity. The keypoints that are referred in this experiment are these regions of an image and most of the work is based on these features and their interpretation with the help of computer vision and machine learning algorithms. There are numerous keypoint detection algorithms which can be applied during the processing of images and selection of the algorithms is entirely dependent on the type of images and the features that are considered important for the application.

4.7 Feature Description

Feature description is a methodology to compute the locally quantified interesting regions of an image. These interesting regions are also known as keypoints. The feature description is accomplished by the use of feature descriptor algorithms. The kepoints that are detected are fed into a feature descriptor and multiple feature vectors are generated to describe the keypoint in an image. The number of keypoint detected in an image equal to the number of feature vectors generated from the feature descriptor.

4.8 Feature Extraction

The process of quantifying the contents of an image is called feature extraction. The features that are considered in this research are local features. Each of the keypoints that are detected are quantified in the form of feature vectors by feature descriptor as described in section 4.7. The feature vectors which are generated by feature descriptors are extracted by feature extractor algorithms. These features describe the respective region of an image surrounding the keypoint. The features that are extracted are usually stored in a disk which are used as a training data on a classification process. The feature extraction can be operated as a parallel processing system using multi-core processing units if the dataset is too large. The described feature vectors after extracting into a disc are then indexed to perform the classification.

5 System Approach

The core principal used in this experiment is a vision-based approach tracking of construction workers working in the construction site wearing safety vests with high reflective material on their body. The reflective material in the safety vest illuminates when exposed to NIR range of light wave. The illuminative property of a reflective material used in safety vest makes up the base foundation for the implementation of the idea used in the study. The stereo camera is designed consisting of two individual monochrome cameras operating in the NIR range of light wave in the Electromagnetic Spectrum. The camera system is equipped with a wide angle lens and an optical band pass filter with the LED ring surrounding the lens of each cameras. The LEDs are synchronized with the camera system using the GPIO ports to obtain the flash and non-flash images alternately.

Figure 10 shows the overall framework of the system. The frames obtained from live streaming of the two cameras in the stereo setup form the input to the system. The algorithm used in this experiment consists of camera geometry calculation, detection of reflective pattern of safety vest and tracking of the detected objects in the FOV of the camera. The frames obtained from the camera then are preprocessed to remove the distortion in the images. The rectified frames are entered into the main algorithm chain for detection and tracking purposes.

The detection method implemented in this study is mainly based on the segmentation process. The segmented image consists of the foreground white blobs and the background is represented in black pixels. The high intensity blobs obtained in the segmented image are of two types, namely, 1. Blobs obtained from the LEDs used in the experiment 2. Blobs obtained from the external light source. These two types of blobs obtained from the process of segmentation are separated by feature tracking algorithm. The matched features in the respective two frames after feature tracking are removed thus eliminating the unwanted blobs originated from the external light sources. The bounding boxes are drawn over the blobs which are obtained after the filtering of the unwanted blobs. The centroid of the of the bounding boxes is computed to determine the construction workers with safety vest.

The 3D pose of the workers was estimated utilizing flash images (I_f^1, I_f^2) of both the cameras using the triangulation to compute depth map of patterns originated from reflective safety vest. Two-dimensional tracking of the workers was performed using centroid based tracking algorithm.

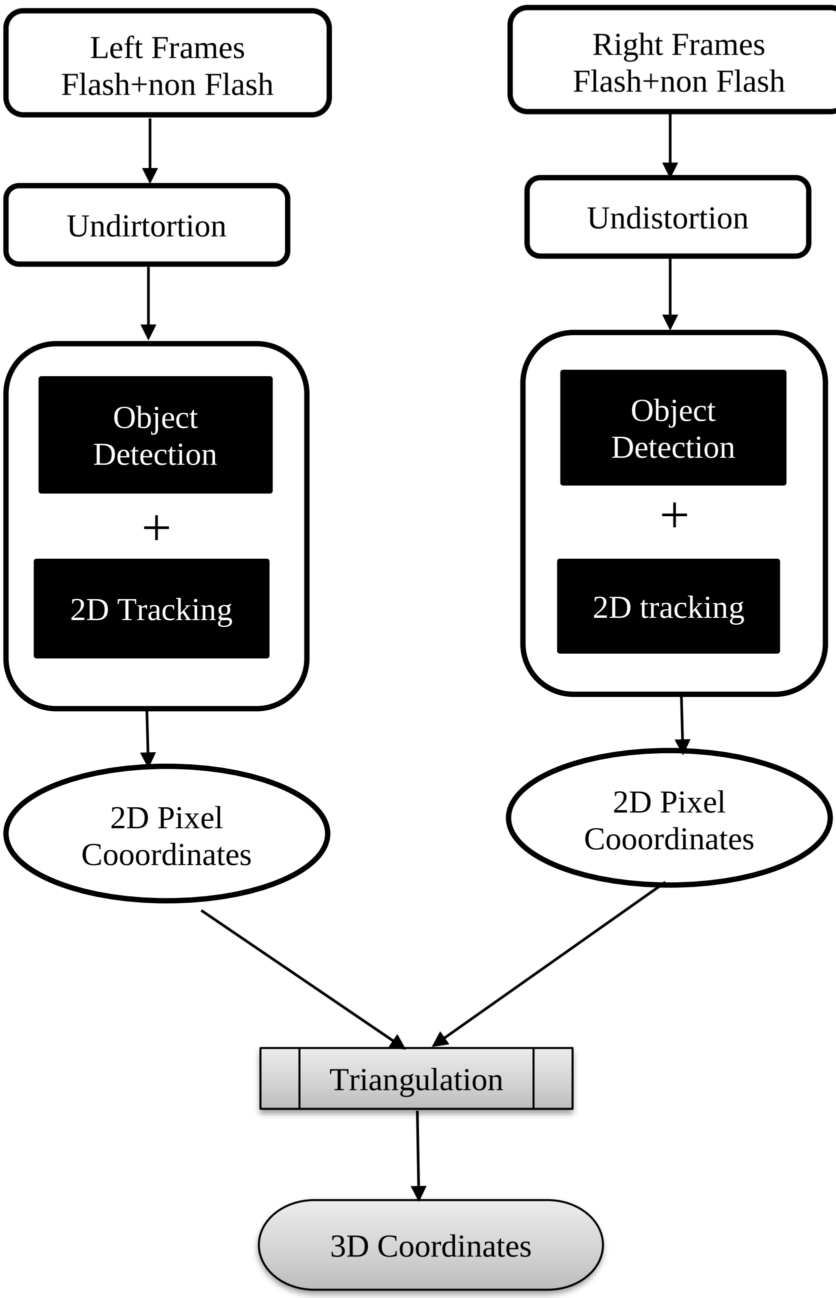


Figure 10. Overall framework of detection and tracking using stereo Camera

6 Hardware Setup

The hardware is a most essential part of the development of any computer vision application. In computer vision application used in real-time, the principal hardware component is a camera unit with the type of image sensors used in them. The sensors collect the data from the field of view are present them in the form of images. The main hardware used in this study is NIR monochrome camera equipped with CMOS image sensor. The camera is then manually assembled with a suitable band pass filter and a wide-eyed lens to form a single unit camera system. The high-power LEDs of central wavelength 850 nm are placed around the camera system. The selection of 850 nm as the centroid wavelength is based upon the quantum efficiency of the monochrome sensor that is used in this system. However, the use of 940 nm could have been more effective in the bright sunlight exposure considering the effect of sunlight is minimum in the 940 nm as shown in the spectral index graph of the sun in NIR range of light in figure 11. The two camera modules are then combined to form a stereo system. The high power LEDS are assembled as a ring and they are controlled by the GPIO pins which are placed in the PCB board of the camera unit. However, the LEDs used in this research consume high power which was beyond the capacity of the GPIO power output and this was solved with the use of external power using the NPN transistor and resistors. The different parts of camera unit are described in the sections below.

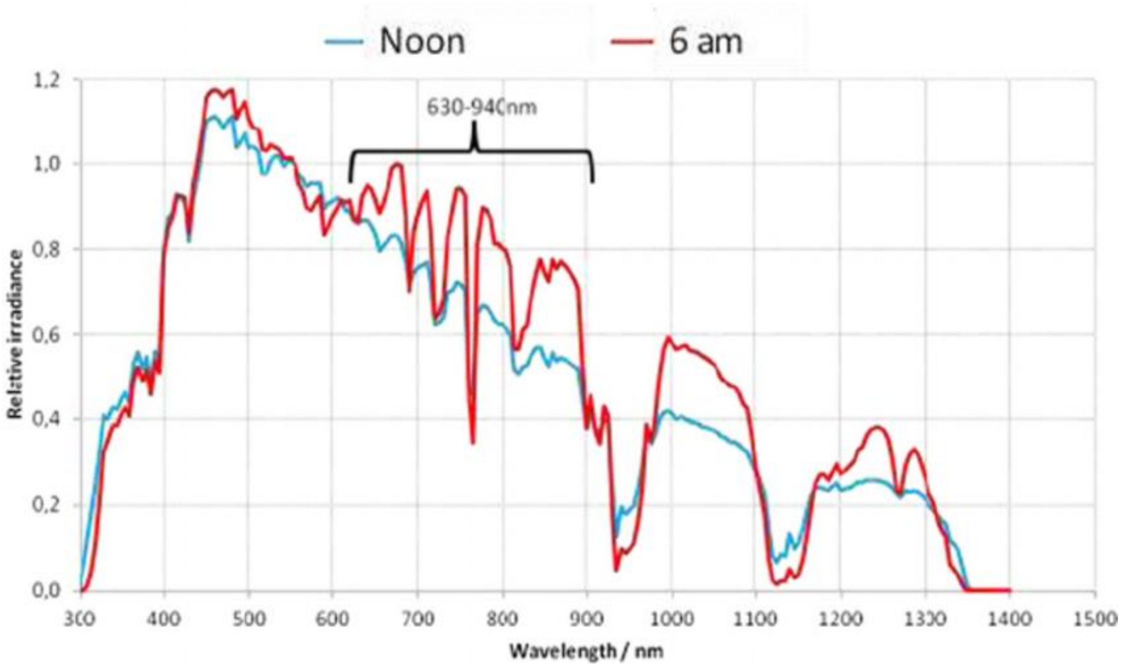


Figure 11. Graph showing the effect of solar radiation in various wavelength (100-2000 nm) [37]

6.1 NIR Monochrome Camera

NIR cameras operate in the NIR region of light of the electromagnetic spectrum. It is equipped with CMOS digital image sensor with a S-mount lens holder with USB 3.0 interface. The NIR range in electromagnetic region is shown in the figure 12 below. The whole explanation of the NIR range in the electromagnetic spectrum is out of the scope of this study but however the references [38, 39] give a better understanding of how the light waves are categorized according to their respective wavelengths. The wavelength of Near Infrared rays ranges from 800 nm to 2500 nm in the electromagnetic spectrum of light. The camera used in this experiment consists of the sensor which has the capability to perform effectively in the NIR range. The designed module is bound to perform in the frequency range of 850 nm, so the band pass filter of center frequency 850 nm is added on the top of camera sensor. The bandwidth of the optical bandpass filter that is used in this experiment is 20 nm. The construction sites have large range of area for surveillance and for this purpose the wide-angle lens of 110 degrees is mounted on the camera.

The NIR range is chosen for the experiment NIR light produces high illumination in the retroreflective material that is placed in the standard safety vest. The use of the safety vest is mandatory in the construction work. This illumination property of the reflective material makes the algorithm efficiently detect the presence of people equipped with safety vest on their body. The use of thermal camera was also taken into consideration for this experiment, but the extreme environmental conditions of the Finland made us consider that the NIR monochrome camera would be an ideal option. The construction works are also carried out in the night conditions and the dark winter conditions in Finland makes the selection of the NIR camera more effective.

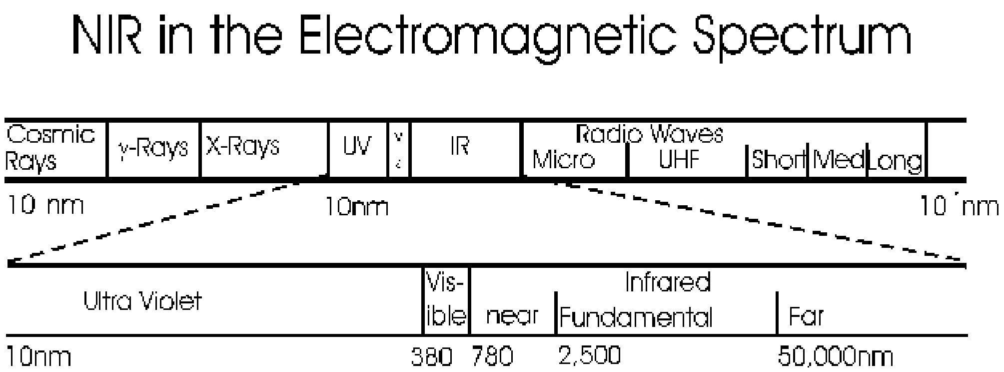


Figure 12. Near Infrared range in the Electromagnetic spectrum [38]

6.2 CMOS Monochrome sensor

CMOS sensor is an electronic chip which converts photons to electrons for digital image processing. CMOS digital image sensor used in this study has active pixel array of 1280H x 960V. The images are captured with rolling-shutter readout. It has good control over exposure and windowing and it is equally efficient with both video and single frame modes. It can be programmed with serial interface as USB is used in this camera module and it takes clear and sharp images and videos which made a logical choice of our application. The sensor is a progressive-scan sensor which generates a stream of pixel data at a constant frame rate.

This experiment is based on the NIR range of light and the camera module with CMOS sensor was selected for the effective performance in the NIR range of light. The graph of the quantum efficiency of monochrome sensor with respect to wavelength is shown in the figure 13. The graph shows that the efficiency of the sensor used in the experiment is 30% in 850 nm and it was turned out to be sufficient enough for the experiment as high resolution images were not necessary for the experimental purposes. The efficiency is measured in quantum unit of light and the wavelength is presented in nanometers as shown in the figure 13 below.

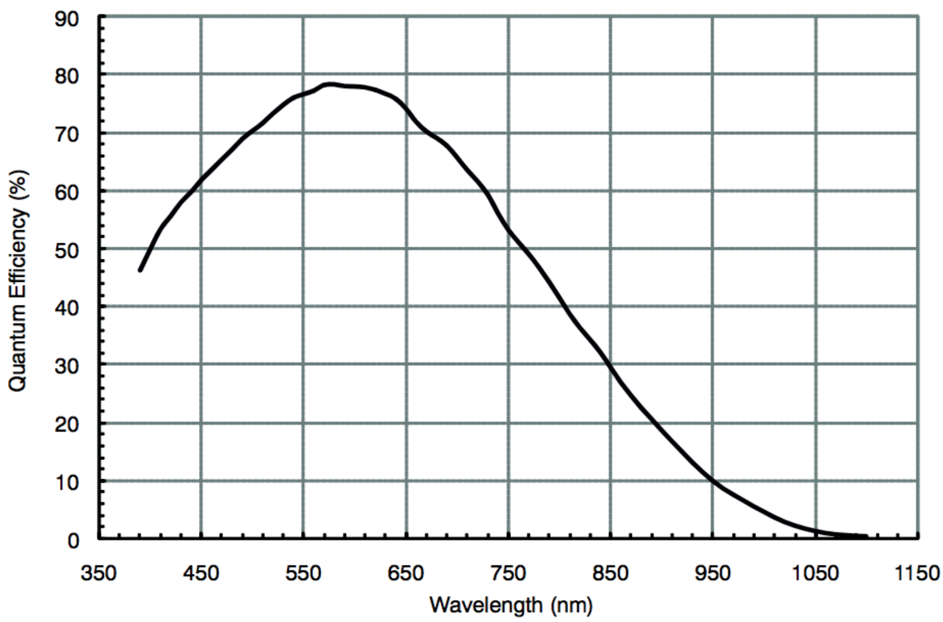


Figure 13. Graph representing quantum efficiency vs wavelength (Source: Camera Developer's Manual)

6.3 Optical Band Pass Filter

The optical band pass filters are placed in between lens and the camera to concentrate our working frequency on a specific narrow bandwidth. As it has already been discussed that the range of NIR range is quite large. The camera module is designed with the LED ring of central wavelength 850 nm, it is not necessary to operate working procedure in the overall NIR region of the EM spectrum. Specifically, it is better to operate in between 20 nm wavelength bandwidth from 850 nm central wavelength. For this purpose, an optical band pass filter was used in in the camera setup. The filter cuts off the lower and higher wavelengths of light outside the bandwidth of a filter. The bandpass filter was adjusted in between camera sensor and the optical lens. The central frequency of the optical filter used is 852 nm and the bandwidth at 10% cutoff frequency is 15nm.

The graph of the optical filter that is used in the study showing transmission of the NIR waves at various wavelengths of the EM spectrum is shown in figure 14.

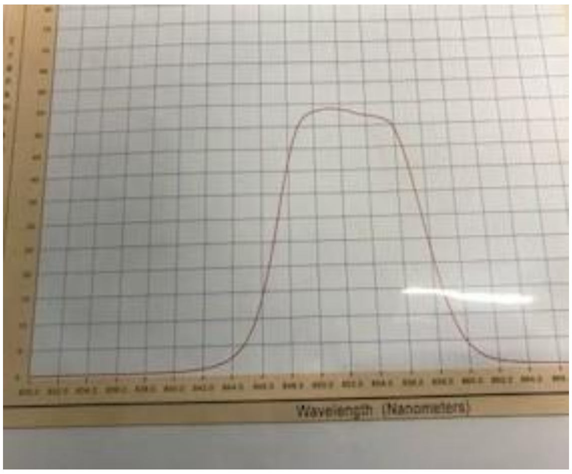


Figure 14. Graph showing the properties of the optical bandpass filter used in the experiment (Source: Filter Manual)

6.4 Camera Parameters and Stereo Geometry

Camera calibration is the one of the most important prerequisite process which needs to be carried out before analyzing any images or video taken from the system. The calibration of camera is necessary to remove the distortion in images and to analyze the 3D geometry of the object in the real-world space. The raw images taken directly form the camera have distortions in them and these distortions are categorized as radial and tangential distortion. Camera has several important parameters in the form of extrinsic and intrinsic parameters and these parameters must be studied well to remove the distortions in the images. Camera calibration in the context of 3D machine vision is defined as the process of determining the properties of the camera systems. The overall optics and geometry of the camera are presented in the references [40,41]. The definitions of important camera properties which are important to this experiment are described below.

6.4.1 Intrinsic Parameters

The characteristics which are related to the hardware of the camera are known as intrinsic parameters. These parameters are found inside the camera and are specific to the hardware of a camera. The intrinsic properties include the focal length, optical center, lens distortion and the pixel coefficients.

These intrinsic parameters are included inside the camera matrix. The equation 1 shows the camera matrix. In the equation 1, (fx, fy) are the co-ordinates of focal length of the camera and (cx, cy) are the co-ordinates of optical center of the camera. It is extremely important to know the intrinsic parameters of the camera to remove the lens distortion caused by the wide-angle lens.

$$\text{Camera Matrix} = \begin{bmatrix} fx & 0 & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{bmatrix} \quad \ldots\ldots\ldots\ldots \quad (1)$$

6.4.2 Extrinsic Parameters

The characteristics related to the orientation of camera in the space are termed as extrinsic parameters. These parameters correspond to the rotation and the translation vector and which translates the coordinates of 3D point to a coordinate system. It is extremely important to know these parameters for stereo calibration and the 3D reconstruction. Knowing these parameters also helps to measure the size of the object.

6.4.3 Lens Distortion

The distortions are caused in the camera because of several internal and external factors. The distortion that is related to this study is a radial distortion caused by wide angle lens. The radial distortion is caused when the light rays bend in a greater amount through the edges of the lens than the optical center [40].

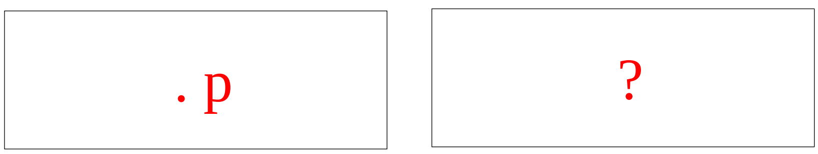
The radial distortion is represented in the equations 2 and 3 below. The variables $(x_{distorted}, y_{distorted})$ are the distorted image points of original (x, y) points in a real image position. The coefficients k_1, k_2, k_3 are the radial distortion coeffients. The radial distortion is corrected using camera matrix as represented in equation 1. Camera matrix is obtained by the calibration of the camera.

$$x_{distorted} = x(1 + k_1r^2 + k_2r^4 + k_3r^6) \quad \ldots\ldots\ldots\ldots (2)$$

$$y_{distorted} = y(1 + k_1r^2 + k_2r^4 + k_3r^6) \quad \ldots\ldots\ldots\ldots (3)$$

6.4.4 Epipolar Geometry

The basic geometry of stereo imaging system is referred to as epipolar geometry. The most important part of stereo imaging is to know the location of cameras and corresponding point of an object in two individual frames of the cameras in a stereo rig. The rectangle boxes below represent simultaneous frames of two individual cameras in a stereo rig. The figure 15(a) is the field of view seen from the camera 1 and figure 15(b) is the corresponding frame seen from the camera 2 of the stereo system. As the two cameras are placed in different co-ordinates of the 3D real world space, the corresponding view from the cameras will also be different. Thus, it is extremely important for both cameras to know the corresponding points to obtain 3D imaging of the scene.



(a) (b)

Figure 15. Rectangle boxes as a representation of two frames of two cameras

7 Procedure and Experiments

This chapter describes the overall processing of the system which includes the detection and tracking of the workers in the construction site. The processing is based upon the streaming of the camera module in real-time for detection and tracking of the construction workers equipped with the reflective safety vest. Most of the work in this experiment is carried out in the indoor environment because of the hardware constraints. The camera module is placed in the tripod stand as shown in the figure 20. The figure shows the stereo unit developed from the two individual cameras.

7.1 Camera Calibration and 3D Estimation

This study is based upon the stereo vision built up from the two separate CMOS monochrome cameras. The details of hardware setup have already been described in chapter 6. Both cameras in a stereo rig were calibrated individually and then the processes of stereo matching were conducted. The calibration was an essential process in this application as there was radial distortion in raw images that were obtained from the cameras. For individual calibration process of the cameras, the chessboard pattern was used to determine the correct points in the chessboard square pattern [40]. The intrinsic and extrinsic parameters of the camera were determined along with the calibration process. The obtained parameters were used to remove distortion from the images/frames.

After the successful calibration of individual cameras, the stereo system was adjusted so that the epipolar geometry was well preserved. For this propose, the feature points are identified and matched with two cameras in the stereo setup. The feature tracking algorithms such as SIFT developed by Lowe [43], SURF developed by Bayer et al. [44], and ORB based trackers are commonly practiced for the robust accuracy. In this study we have analyzed different feature tracker algorithms to test the accuracy and precision of the stereo geometry. Finally, the depth map was created from two separate cameras to ensure the hardware was constructed with correct geometrical accuracy.

The overall preprocessing steps of the stereo formation can be summed up in the following steps.

- Camera Calibration
- Epipolar Geometry
- Triangulation
- Depth Map from Stereo Images

7.2 Image Acquisition

In this study, each frame is analyzed in a streaming video from two cameras and the two frames taken simultaneously as a flash and non-flash images are processed as a single unit. The two flash and non-flash images from two cameras in the stereo setup form four individual images which are entered into the processing system as a single unit. The timing in between the flash and non-flash image was kept small so that there won't be much difference in the image planes in between two continuous frames in a streaming video. Two images taken as flash and non- flash image pairs are processed as a pair in the real time operation of the system. The distortion in the frames were rectified with a camera calibration process explained in chapter 7.1. The input frames were reduced in dimensions and fetched to the processing chain for the further processes. This process helped in reducing the computational effort and the processing speed of the overall system was increased.

The figures 16 and 17 are the two continuous frames of the streaming camera. Figure 16 is the frame captured with flashed LEDs and the figure 17 is the frame captured when the LEDs were in low state making it a non-flash image. These frames were taken in quick succession with the flash and non-flash state of high-power LEDs used in the system. The figures 15 and 16 show the distinct properties in the images of two continuous frames.

It can be clearly observed that distortion is minimum in both the figures 16 and 17. The reflective pattern of vest is distinct in figure 16 when the image was taken in flash condition of LEDs. However, there can be seen no pattern of the safety vest in the figure 17 as the image was taken in non-flash conditions. The outside conditions when the frames were captured was bright and sunny. The real color view of the indoor environment was brighter compared to the background intensity seen in the figures. The dark background conditions in the images were the results of band-pass filter which filtered out all the frequencies which do not fall under its bandwidth.

Figure 16. Frame captured with flash in indoor condition

Figure 17. Frame captured with non-flash in indoor conditions

As explained above, the difference in reflective pattern of the safety vest is seen in two frames. The reflective patterns of the reflective vest did not appear after the LEDs were turned to low state (non-flash). The mirror reflection can be observed similar to the one in figure 15 as these were the result of reflection from the sunlight. Both of the figures seem to be identical except the pattern of reflective vest.

7.3 Segmentation

The resulted undistorted frames obtained from the earlier process are entered into image processing. The segmentation process is applied to the resulted frames to obtain binary images. For the process of segmentation, local adaptive thresholding was applied to obtain the white blobs which represented higher intensity pixels in the images. The two images obtained from the flash and non-flash systems of the LED ring were processed together for the segmentation to obtain two binary images with distinct patterns of blobs. The segmentation of frames/images was first experimented by simple thresholding. The simple thresholding technique performed well in particular environment as thresholding value was adjusted accordingly for the controlled conditions. However, the thresholding value had to be adjusted each time as we changed the testing environment because of the occurrence of various lighting conditions in different environment due to sunlight and other external lights. To solve this problem, the adaptive thresholding methods were applied so that the same algorithm would be effective for different lighting conditions throughout the day and night in whole time of the year. First, the Otsu’s method of thresholding was applied, and the better results were obtained compared to simple thresholding algorithm. Then, the local adaptive thresholding was implemented to obtain the much better results in different light conditions compared to the previous results. The figure 18 shows the images obtained from the color camera and the NIR camera and the corresponding processed form of images ontained from NIR camera.

The figure 18(a, b) are the color images taken from external mobile camera. 18(c) and 18(d) are the images taken from NIR camera in non-flash and flash condition respectively. The figure 19 is the resulted frame obtained after filtering out the same intensity pixels from the flash and non-flash images. The same intensity blobs in the consecutive flash and non-flash frames are removed leaving only the reflective safety vest pattern in the output frame.

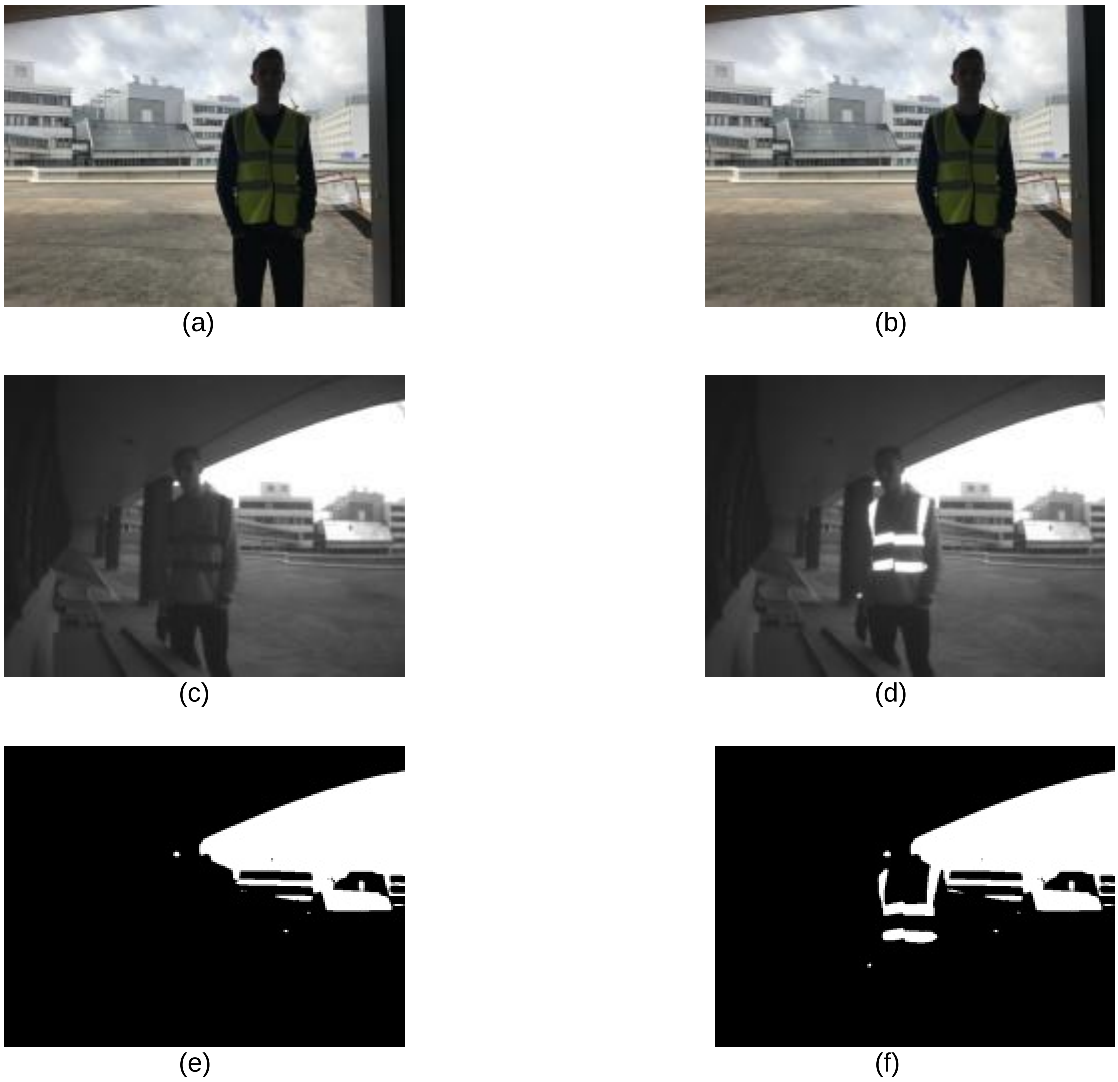


Figure 18. (a, b).Images taken from color camera (c ,d).Raw images taken from NIR camera in non-flash and flash conditions. (e, f) Respective thresholded images of c and d

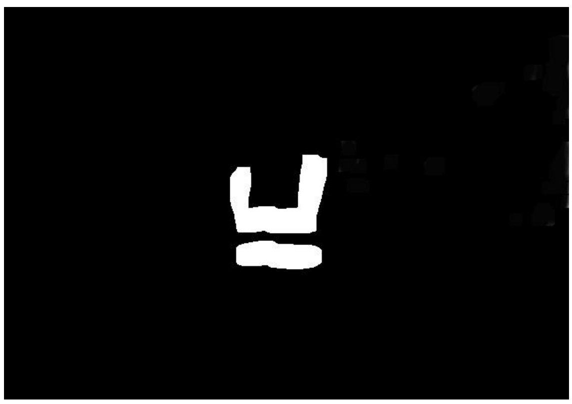


Figure 19. Fully segmented image after processing of flash and non-flash images shown

7.4 Detection and Computation of Bounding Boxes

The binary frames were thus obtained from the segmentation procedure as explained in the section 7.3. The binary frames obtained consisted of distinct pattern of reflective vests as blobs. The resulted blobs in the images had possibility of origin from different sources. Those blobs were either obtained from the illuminative surfaces or other external bright intensity surfaces originated from the external light sources. The feature matching was conducted between the two frames and the features which closely matched in two consecutive frames with flash and non -flash of the LEDs were removed by the bitwise operation. The blobs features which were matched in both the frames were removed thus only leaving the blobs originating from the illuminating reflective vest. Then the bounding boxes were computed using the function in software environment using the properties of available contours.

The individual bounding boxes were thus created over all the blobs which were received after the removal process of unwanted blobs. The centroids of each contours were computed which were of tremendous importance in the tracking algorithm. The computation of the centroid would help in the computation of the keypoints which is an essential process in the classification system of obtained blobs. The images of the bounding boxes drawn over the detected persons with safety vest are presented in tests and verifications results.

7.5 Tracking

Tracking is simply defined as locating the specific object in the subsequent frame of the video to determine its relative movement with respect to other objects. The identity of object in the tracking process is preserved as long as it in the FOV of the camera. Usually the object is first determined by the detection process. Then a unique id is assigned to the object in a bounding box and then the identity of the object is preserved in the successive frames of the video. The tracking algorithms can be operated either by operating object detection in every frame of a video or by running object detection algorithm in the first frame and then preserving its identity throughout the operation. The tracking algorithm can preserve the identity of an object even when the object is lost for small amount of time due to partial or full occlusion. The tracking algorithm is implemented in the study for the specific purpose of partial and full occlusion that is expected to happen in the construction area. The construction area offers high possibility of partial or full occlusion of the construction workers. The tracking algorithm preserves the identity if the workers are lost and it can be a great safety aspect when the workers are occluded by the arm of the excavator itself. Here in this study, the implemented contour-based approach of blob detection is applied, and it has acted as the sole basis of the tracking system in the subsequent frames.

The overall tracking algorithm of the system that we have implemented in the system is noted in the points below.

1. Accept the bounding boxes of the safety vest detected from the detection algorithm.

2. Compute the Euclidean distance between the existing bounding boxes and new bounding boxes along the centroid.

3. Update (x, y) co-ordinates of the object represented by bounding boxes.

4. Object update by registering new objects and deregistering lost objects.

7.6 Tests and Verifications

The tests and verifications of the study were carried out in different environment conditions with varying light conditions. The tests were carried out in the 4 different types of environment conditions. Test experiment 1 was conducted in indoor environment with different kinds of lighting conditions in the room/hall. The test experiment 2 was conducted in outdoor condition in the morning time with bright conditions. The experiment 3 was carried out in the evening with deem sunlight. Finally, the experiment 4 was conducted in extreme bright condition with strong sunlight during the day with the camera directly facing the sun.

The setup of the first experiment is shown in the figure 20. The stereo system is placed in a tripod stand at the height of around 1.6 meters. The breadboard is attached to the tripod stand for the electrical and power operations involved in the study. Persons were equipped with a standard reflective safety vest on their body and the camera module was streamed in real-time for evaluation purposes. The persons showed quick movements in various body postures. The system performed extremely well in those test conditions as there were few unwanted reflective materials in the site. The system was computationally fast, and it showed good accuracy of detection. There minimal false positives were detected due to the reflections from the mirror.

Figure 20. Camera Module Setup for the tests and verifications

The test results of different body positions in indoor conditions are shown in figure 21. The images 21(a, b, c) are the color images taken from the external color camera. The color images are represented only for the better understanding of the test scenario. The various body positions and the movements were experimented as seen the images below. Figure 21(c) shows the case of partial occlusion and images 21(a, b) show different postures of body positions. The figures 21(d, e, f, g) are the frames obtained from camera module used in this experiment as shown in figure 20. The figures 21(d, e, f) show the robust detection in forward and backward bending postures. However, two detections are observed for the single person in figure 21(d) which was a limitation of this test as it's not logical for one person to be represented by multiple detection.

It can be seen in figure 21(g) that the right arm of a person has disrupted the regular pattern of the safety vest. The image in a given scene is a side view from the camera angle and it is supposed to be one of the major challenges of detection. Our system showed excellent result in this situation as seen in the figure 21(g) with a perfect bounding box around the reflective vest. The figure 21(h) shows the partial occlusion in forward position of the person from the camera angle. The full view of the person is blocked by the obstacle, but the reflective pattern is obvious and distinct. As a result, the system showed efficient detection in this situation.

The second test experiment was conducted in outdoor condition during the morning time. The test site was an open outdoor yard and the conditions were bright with normal morning sun in the sky. The camera module was placed in the yard in the same tripod stand as in indoor conditions. The verification results obtained from the test are presented in figure 22 below. The figure 22(a, b, c, d) are the color images where the persons are equipped with safety vest in their body with different body positions. These figures show the original condition of the test environment. The figure 22(e, f, g) show the results of the detection and tracking.

The reflective patterns of the safety vest are represented with the rectangular box with the centroid drawn over the center of a rectangular bounding box for the tracking purposes. The system showed satisfactory results in those conditions as there were few false detections. There can be seen false detection of the background object in the figure 22(e). It was bright spot in the sky originated from the sunlight. It is expected to be removed by the classification algorithm during the future works. Moreover, the multiple detection of the same reflective vest persisted in this test sequence as well. The multiple detection created confusions in identity of workers during the tracking purposes. The occlusion of the person in figures 22(e, g) is efficiently detected by the system which is a huge positive thing for the development of the system.

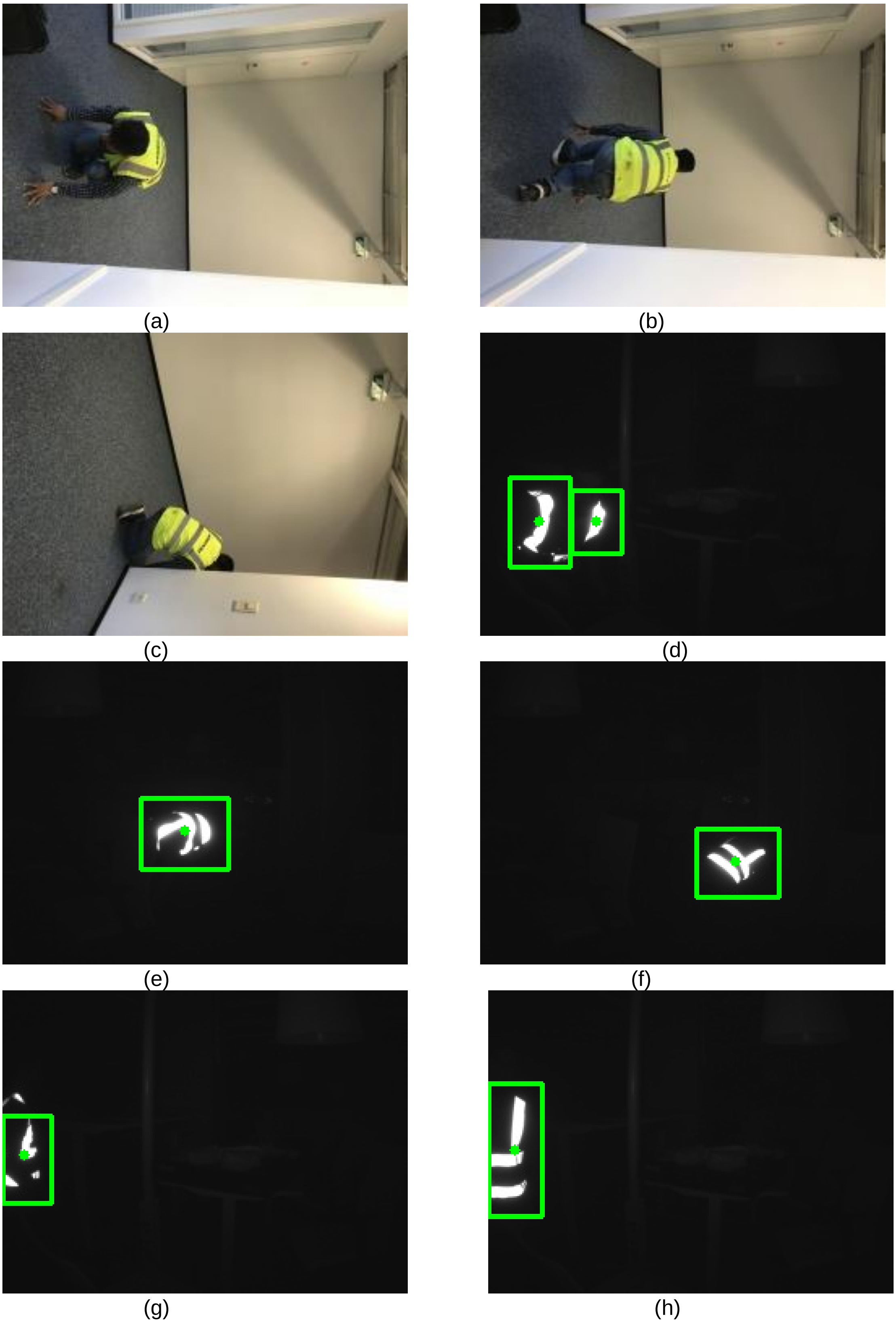


Figure 21. a, b,c. Color images taken from external camera. b,c,d,e,f.Real-time evaluation results in indoor condition

(a)

(b)

(c)

(d)

(e)

(f)

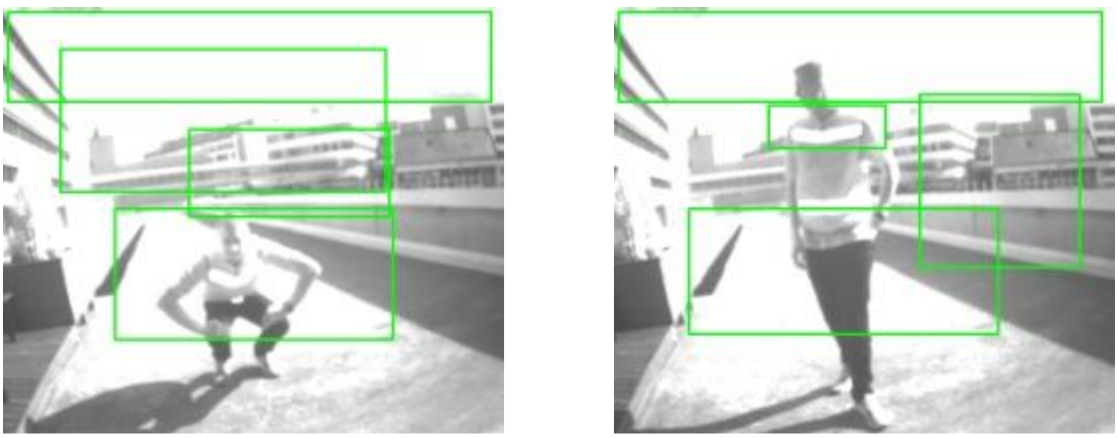
(g)

Figure 22. a,b,c,d. Color images taken form external camera. e,f,g.Real-time verification results in outdoor conditions during morning time

The third test was carried out in a bright sunny condition during a hot summer day with strong sunlight facing the camera. The similar approach of experimental set up was applied for the testing as in previous two conditions. The results of the test experiment 3 in bright sunny conditions are presented in figure 23. The multiple false results can be seen in figure 23(a, b). The bright sunlight has nullified the effect of LEDs and flash and non-flash images resembled similar. These false detections can be overcomed by the use of simultaneous color camera in the system or by the use of depth sensor. The use of better range of NIR region is also one of the possibility to minimize the false detections.

The fourth test was conducted in the evening conditions during a moderate evening summer time of the year. The sunlight was not strong but sufficient enough to create clear bright condition of the evening as seen in the figure 24(a). The similar approach as in tests 1 and 2 and 3 was applied for the hardware setup and body positions.

The results of the test experiment 4 are presented in figure 24. The color image representation of the test can be seen in figure 24(a). The other images in the figure 24(b, c, d) are the verification results obtained from the test. The figure 24(b, c, d) show that the visible rays of light are filtered by the band pass filter creating ideal condition for the efficient detection. The test results showed great accuracy in this test conditions. The noise in the frames were almost removed perfectly and the wanted blobs were also filtered out. The system efficiently identified the persons wearing safety vest in their body as seen in the images of figure 24 (b, c, d). The test results in those conditions showed encouraging results for detection and tracking.



(a) (b)

Figure 23. Real-time verification results in outdoor conditions during bright sunny day

Figure 24. a. Color image taken from external camera. b,c,d.Real-time verification results in outdoor conditions during evening time

8 Discussions and Future Works

The images in the figures 21, 22, 24 show that this model works well for the various body positions. The reflective material is detectable when small portion of the torso of the construction worker is exposed to the camera. The system in tested in indoor environment and outdoor environment. The results in the indoor environment were robust despite the prevalence of varying light conditions. The outdoor test results were of good accuracy when there was no bright sunlight. However, the results were poor when the bright external sunlight passing through the camera sensor was quite strong during the mid-day conditions. The problem of sunlight illumination could be solved by the use of band pass filter of 940 nm as suggested by the authors in the journal [26].

The encouraging factors of the use of NIR LEDs is that the system is extremely effective in the dark conditions. The model is capable of functioning during the construction works in the evening or night hours. The effective detection and tracking in the dark hours can be extremely useful in the Finland as the conditions are dark throughout the day and night during the winter. However, the problems of illuminating snowflakes and the rainy conditions are expected challenges to be encountered during the further works.

Although the main goal of this study was to develop a system which could be an upgrade to the previous works, various factors such as time and the hardware constraints limited the improvement of overall accuracy of the results. However, further research works are expected to be carried out for the development of the Advanced Driver Assistant System which could be a milestone in the occupational safety systems in the construction field.

The immediate future work which is extremely essential to upgrade this system is the classification of the various reflective objects which were detected during the real-time test environment. This could be achieved by the traditional algorithms such as SVM and Random Forest or the advanced deep learning-based classification models that that been pertained to localize and detect the objects. The preparation of the dataset is underway, and it is expected to be an immediate future work. The authors in the paper [26] have achieved high classification accuracy using the Random Forest algorithm. Deep learning-based algorithms of classification have turned out to be more precise in accuracy in real-time object detection and tracking entities in recent times.

Recently, the semantic segmentation-based tracking algorithms are employed in autonomous driving cars [45]. This could also be used along with the system as the classification runs in a semi supervised manner. The use of semi supervised algorithm in the real-world scenario could be more effective than fully-supervised ones as new objects are expected to appear randomly in real world situations. Nonetheless, the model developed in study has promising initial outcomes and is expected to be upgraded in the future.

9 Conclusion

The goal of this study was to develop the Advanced Driver Assistance System for the excavators which would minimize the occupational hazards that have been frequently occurring in the construction industry. The study consisted of the development of hardware and software modules for the efficient detection and tracking of the construction workers wearing safety vest on their body. The hardware system consisted of the stereo camera module built up of two individual monochrome camera modules with the optical filter, lens and the high powerful LEDs. The system was designed to work in the Near Infrared range of light and for this purpose high power LEDs were placed in the camera modules. The software was developed with various algorithms of computer vision and machine learning detection and tracking purposes. The open source platform of computer vision library(OpenCV) and MATLAB were used for the software development.

The reflective pattern of safety vest was used for the detection purposes. The use of optical band pass filter was used for the purpose of eliminating the unwanted wavelengths of light. The use of filter boosted the formation of distinct patterns of the reflective safety vest. The on and off conditions of LEDs was used to eliminate the unwanted bright patterns originated from other external false objects. The detection and tracking were solely dependent on the patterns of reflective vest based upon the flash condition of high power LED ring.

The tests and verifications were carried out in various indoor and outdoor conditions. The persons were equipped with safety vests in their body and real time streaming of the camera system was conducted for the detection and tracking purposes. The results showed high accuracy of detection in the indoor conditions and in the environment with no external light conditions. The results obtained in the morning and evening conditions were quite satisfactory with detection but there were few false detections of other reflective materials. The false detections are expected to be overcome with the classification of those detected regions with the implementation of advanced algorithms of machine learning. The test results obtained in the bright sunny condition directly facing the camera were quite poor. The use of better optical instruments and sensors are expected to minimize the false detections in bright sunny conditions.

Overall, the approach of this study and the results obtained from the various test experiments look promising for the future works. The use of better hardware equipment and advanced algorithms for software development is necessary for the development of robust and reliable autonomous system for the occupational safety in construction works.

References

1 “Occupational Health”. National Institute of Environmental Health Sciences. [Online] Availabe at: https://www.niehs.nih.gov/health/topics/population/occupational/index.cfm [Accessed on: 11 08 2018]

2 Lucchini, R.G & London, L., 2014. “Global Occupational Health: Current Challenges and Need for Urgent Action.” Annals of Global Health, vol 80, issue 4, pp:251-256.

3 Rantanen, J., Lehtinen .S & Iavicoli, S. “Occupational health services in selected International Commission on Occupational Health (ICOH) member countries.” Scand J Work Environ Health (2013);39: 212e6.

4 Peltola, H. , Luoma, H. “Road accidents in Finland and Sweden; “A comparison of associated factors.” [Online] Available: https://www.researchgate.net/publication/292552641_Road_accidents_in_Finland_and_Sweden_A_comparison_of_associated_factors [Accessed on: 11 8 2018].

5 Official Statistics of Finland (OSF):” Occupational accident statistics [e-publication]. ISSN=1797-9544. 2013, Wage and salary earners’ accidents at work. Helsinki:” Statistics Finland. [Online] Available at: http://www.stat.fi/til/ttap/2013/ttap_2013_2015-11-27_kat_001_en.html [Accessed on: 11 8 2018]

6 Quang-Nguyen, V., 2017. “Safety System in Construction Industry: A protection system preventing fatal injuries by excavators in construction sites.” Helsinki Metropolia University of Applied Sciences [Online]. Available at: http://www.theseus.fi/handle/10024/139396 [Accessed on: 18 3 2018]

7 Legislation on building products. Ministry of Environment. [online] Available at: http://www.ym.fi/enUS/Land_use_and_building/Legislation_and_instructions/Legislation_on_building_products [Accessed on: 11 08 2018]

8 EOTA Guidance on the content of a European Technical Assessment (ETA). [Online]. Available at: https://www.eota.eu/ckfinder/userfiles/files/ec-vacancy/7-eota-guidance-on-eta-content-cpr-draftv05072013.pdf [Accessed on: 29 08 2018)

9 Occupational Safety and Health in Finland. Ministry of Social affairs and Health, 2007. [Online] Available at: https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/70253/Es200616eng.pdf?sequence=1&isAllowed=y [Accessed on: 29 8 2018]

10 Wikipedia, “Global Positioning System” [Online]. Available at: https://en.wikipedia.org/wiki/Global_Positioning_System [Accessed On: 6 08 2018]

11 "GPS Accuracy" [Online]. Available at: https://www.gps.gov/systems/gps/performance/accuracy/ [Accessed on: 4 8 2018]

12 Valero E., Adán A., Cerrada, C. "Evolution of RFID Applications in Construction: A Literature Review". Fontecha J, ed. *Sensors (Basel, Switzerland)*. 2015;15(7):15988-16008. doi:10.3390/s150715988.

13 Teizer, J., Lao D., Sofer, M. (2007)." RAPID AUTOMATED MONITORING OF CONSTRUCTION SITE ACTIVITIES USING ULTRA-WIDEBAND." 24th International Symposium on Automation & Robotics in Construction (ISARC 2007), KOCHI, KERALA, INDIA. [Online] Available at: https://www.iaarc.org/publictions/fulltext/isarc2007-1.1_3_091.pdf [Accessed on. 5 9 2018]

14 Advanced Ultraband Radar: Signals, Targets and Applications. [Online]. Available at: http://spendergast.blogspot.com/2017/01/advanced-ultrawideband-radar-signals.html [Accessed on: 6 8 2018]

15 Teizer, J. (2008). "3D range imaging camera sensing for active safety in construction", ITcon Vol. 13, Special issue Sensors in Construction and Infrastructure Management, pg. 103-117 [Online] Available at:https://www.itcon.org/2008/8 [Accessed on: 5 9 2018]

16 G. J Awcock and R.Thomas," Applied image processing", McGrew Hill, 1966

17 Anthony, J. J., Suchetha, M. "Vision based Vehicle Detection: A Literature Review." [Online]. Available at: https://pdfs.semanticscholar.org/54d3/3ebacd199a9aca578f6402ddc85bf265a14.pdf [Accessed on: 6 7 2018]

18 Viola, P. and Jones, M., "Rapid Object Detection using Boosted Cascade of Simple Features" [Online] Available at: https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf [Accessed on: 6 7 2018)

19 Dalal, N. and Triggs, B., "Histograms of oriented gradients for human detection"; [Online]. Available at: https://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf [Accessed on: 6 7 2018]

20 Zhu, Q., Avidan, S. , Yeh, C. M. and Cheng, T. K. ; "Fast Human Detection using a cascade of Histogram of Oriented Gradients" [Online]. Available at: https://lear.inrialpes.fr/people/triggs/student/vj/zhu-cvpr06.pdf [Accessed on: 6 7 2018]

21 Swain, M.J. and Ballard, D.H. (1991); "Colour indexing. International Journal of Computer Vision"

22 Park M., "Automated 3D vision based Tracking of construction entirties." [Online]. Available at: https://pdfs.semanticscholar.org/ [Accessed on: 19 7 2018]

23 Shashua, A., Gdalyahu, Y and Hayun, G, 2004.” Pedestrian detection for driving assistance systems: single–frame classification and system level performance”. In: IV, Parma, Italy.

24 Andreasson, H. & Bouguerra, A., Stoyanov, T., Magnusson, M. & Lilienthal, A. (2011). Vision-based People Detection Utilizing Reflective Vests for Autonomous Transportation Applications. [Online]. Available at: https://www.researchgate.net/publication/264081850_Vision-based_People_Detection_Utilizing_Reflective_Vests_for_Autonomous_Transportation_Applications [Accessed on: 18 3 2018]

25 Mosberger R., Andreasson H., Lilienthal, AJ. “A Customized Vision System for Tracking Humans Wearing Reflective Safety Clothing from Industrial Vehicles and Machinery.” Sensors (Basel, Switzerland)17980.doi:10.3390/s141017952. [Online]. Available at: https://www.researchgate.net/publication/264081850_Vision-based_People_Detection_Utilizing_Reflective_Vests_for_Autonomous_Transportation_Applications [Accessed On: 18 3 2018]

26 Mosberger, R., & Andreasson, H. (2013). An Inexpensive Monocular Vision System for Tracking Humans in Industrial Environments. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) (pp. 5850–5857). https://doi.org/10.1109/ICRA.2013.6631419

27 Mosberger, R. (2016). Vision-based Human Detection from Mobile Machinery in Industrial Environments (PhD dissertation). Örebro university, Örebro. [Online] Available at: http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-48324 [Accessed on: 18 6 2018]

28 Redmon J., Divvala S., Girshik R., Farhadi A. “You Only Look Once: Unified, Realtime Object Detection.”

29 Liu, G. “Real-Time Object Detection Based on Deep Learning” Texas A & M University [Online]. Available at: http://tamuccir.tdl.org/tamuccir/bitstream/handle/1969.6/5637/Liu%2C%20Guangrui%20thesis.pdf?sequence=1&isAllowed=y> [Accessed on: 25 8 2018]

30 Simonyan, K. and Zisserman, A. ,2015. ”Very Deep Convolutional Networks for Large-Scale Image Recognition.” ICLR, pp 1–14.

31 Konig, D., 2017. ”Deep Learning for person Detection in Multi-Spectral Videos.”University of Ulm, Germany

32 Rosebrick, A., “Practical Python and OpenCV”

33 Deshpande, A., “Object Recognition Using Large Datasets”. [Onlin Available at: https://pdfs.semanticscholar.org/86da/acfca95204c1bff000da7fa5fcfb430205e5.pdf [Accessed on: 6 8 2018]

34 Standard Test Images. Online: URL: https://www.ece.rice.edu/~wakin/images/ [Accessed on: 6 6 2018]

35 Web Page. http://pages.cs.wisc.edu/~dyer/cs766/slides/segmentation/segment-4up.pdf [Accessed on: 5 6 2018]

36 Infrared and Skin: Friend or Foe - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/Early-morning-6-am-relative-irradiance-of-the-sun-is-higher-in-the-visible-and-NIR_fig8_287807359 [Accessed on: 17 9 2018]

37 Electromagnetic Waves. [Online]. Available at: https://crisp.nus.edu.sg/~research/tutorial/em.htm [Accessed on: 6 8 2018]

38 Web Page URL: http://dew.globalsystemsscience.org/key-messages/near-infrared-and-the-electromagnetic-spectrum [Accessed on: 6 8 2018]

39 Wang, Q., Wang, Z. and Smith, T. "Radial distortion correction in a vision system," Appl. Opt. 55, 8876-8883 (2016)

40 Bradski G. and Kaehler A, "Learning OenCV"

41 Web Page: https://docs.opencv.org/3.1.0/dc/dbb/tutorial_py_calibration.html [Accessed on:6 8 2018]

42 Lowe, David G. (1999). "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision 2. pp. 1150–1157.

43 Bay, H., Tuytelaars T., Gool, V. L., "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008

44 Sematic Segmentation [Online] Available at: https://medium.com/intro-to-artificial-intelligence/semantic-segmentation-udaitys-self-driving-car-engineer-nanodegree-c01eb6eaf9d [Accessed on: 25 8 2018]

45 Siam, M., Elkerdawy, S., Jagestand, M., Yogamani S., "Deep Semantic Segmentation for Automated Driving: Taxonomy, Roadmap and Challenges." [Online] Available at: https://arxiv.org/pdf/1707.02432.pdf [Accessed on: 25 8 2018]

46 Konig, D., 2017. "Deep Learning for person Detection in Multi-Spectral Videos."University of Ulm, Germany.

47 Viola, P., Jones, M., and Snow, D., 2003."Detecting pedestrians using patterns of motion and appearance". ICCV.

48 Lim, J. J., Lawrence, Z.C., and Dollár, P., 2013. " Sketch Tokens: A Learned Mid-level Representation for Contour and Object Detection." IEEE Conference on Computer Vision & Pattern Recognition, 9(4):pp3158—3165.

49 Gavrila, D.M. and Davis, L.S., 1996 "3D model-based tracking of humans in action: a multiview approach" IEEE International Conference on Computer Vision and Pattern Recognition, pp 73-80

50 Bertozzi, M., Binelli, E., Broggi, A., and Del Rose, M., 2005, "Stereo Vision-based approaches for Pedestrian Detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 3.

51 Ebihara, W., Ote, I., Koya, H., Murakami, T. & Nakata, S., 2016."Image Sensing techniques for analysis and interpretation of surveillance video." Hitachi Review Vol 65, No.5 page 53-57

52 Nilson, R. and Nordgren, M., 2003."Vision Based Control of the Overhead crane".LTH, Sweden.