A Study of Fall Detection: Review and Implementation

Muhammad Mubashir

A thesis submitted for the degree of

Master of Philosophy

October 2011

A Study of Fall Detection: Review and Implementation

Muhammad Mubashir

Communications Research Group
Dept. of Electronic and Electrical Engineering
University of Sheffield, UK

A thesis submitted for the degree of

Master of Philosophy

October 2011

Abstract

This thesis presents a research study on Fall detection with a comprehensive survey on available related literature and an evaluation experiment. Fall detection is a major challenge in the public health care domain, especially for the elderly, and reliable surveillance is a necessity to mitigate the effects of falls. The technology and products related to fall detection have always been in high demand within the security and the health-care industries. An effective fall detection system is required to provide urgent support and to significantly reduce the medical care costs associated with falls. In this thesis, we initially give a comprehensive survey of different systems for fall detection and their underlying algorithms. Fall detection approaches are divided into three main categories: wearable device based, ambience device based and vision based. These approaches are summarised and compared with each other and a conclusion is derived with some discussions on possible future work. Then we present an evaluation of fall detection using optical flow. Optical flow is one of the widely used approaches in computer vision to estimate motion. The literature of optical flow is briefly reviewed and some of the methods are implemented with discussion on experimental results. The best output yielding algorithm with respect to accuracy is used to setup an evaluation of fall detection. The evaluation compares our experimental results with the results obtained using other techniques. At the end we draw a conclusion in general on our research study and in particular on our contributions: Fall detection survey and Fall detection Evaluation. We also point out the futuristic direction of our research study with suggestions on possible areas with further development.

Acknowledgements

I would like to take this opportunity to dedicate this thesis to my late father who has always been the source of inspiration. I am truly indebted and thankful to my supervisor Dr Ling Shao whose encouragement, supervision and support from the preliminary to the concluding level enabled me to develop an understanding of my research study. I owe sincere and earnest thankfulness to Dr Luke Seed for the support and guidance he showed me throughout the research study. I would like to thank Mum, Anne, Zahid, Irfana, Saima, Rizwan and the rest of my family members for keeping me motivated at all times. Lastly, I offer my regards and blessings to my fellow colleague Mr Xiantong Zhen who was always willing to help and give his best suggestions and to all of those who supported me in any respect during the completion of the project.

Contents

Publication

Our research paper as a comprehensive survey of different systems for fall detection and their underlying algorithms is accepted for publication in Elsevier's journal called Neurocomputing.

M. Mubashir, L. Shao and L. Seed, *A Survey on Fall Detection: Principles and Approaches*, Neurocomputing, Elsevier, 2011

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

The quality of an individual's life is significantly affected by the levels of functional ability. Plenty of research has been done in this area to develop systems and algorithms for enhancing the functional ability of the elderly and patients. The maturity of cameras, sensors and computer technologies make such systems feasible. Such systems can not only increase the independent living ability of the elderly, by raising the confidence levels in a supportive care environment within the public sector, but also save on manual labour in terms of the presence of nurses or support staff at all times.

The elderly population is expected to grow dramatically over the next decades. The number of people requiring will grow according, while the number of people able to provide this care will decrease. Without receiving sufficient, elderly are at risk of losing their independence. Thus an intelligent monitoring system allowing elderly people to live safely and independently at home is more than needed.

A simultaneous increase in, fall related injuries and a decrease in qualified staff hires, is observed by Hospitals and nursing homes. "*In countries like England, falls account for 32.3% of reported patient safety incidents in hospitals* [142]". The implementation of preventative measures is step forward towards the solution of the problem through the minimisation of incidents leading to injury without necessitating a larger staff. Stopping falls from ever occurring is very difficult, but prevention and reduction of injuries would definitely lessen the dilemma.

In this research study, we studied algorithms and system design for automatic monitoring and detection of human actions and events of interests using video sequences. Recognising human behaviour/activity has become a popular research area recently. The existing techniques can only recognise simple actions with fixed lighting condition and no complex background clutter. To study a reliable system, many computer vision algorithms, e.g. Human detection and tracking, background subtraction, Motion detection etc have to be utilised and refined.

1.2 Motivation

Falls being 40% more likely to occur in a hospital are of the most common injuries in hospitals than in other industries and locations [143]. By comparison elderly population has likely possibility of falling approximately 50% more than the general population [144]. Doctors and care takers administration have been held responsible and liable in the lawsuits and therefore prone to overspend in areas like personnel [145].

This research study attempts to suggest an optimal solution for common problem like fall that has had multiple attempts at solutions. Some, like a physical alarm, would be intrusive on the surrounding patients and may even increase the likelihood of falling [146]. Incorporating ideas from various approaches and algorithms within and widening the research area can yield the system for appropriate surveillance in places like hospitals and nursing homes.

Many approaches currently and in the past require user's direct input for the system to be able to function such as use of a belt-size alarm with a button on it that only sounds when pushed. The system prevents fall to an extent in some types of injuries but obviously failed to show robustness in the case of falls due to unconsciousness. Some techniques are based on generalisation of a fall action but their accuracy remains a big question mark in the case of

detecting most fall scenarios. As there has been no empirical evidence of a decrease in injuries resulting from falls so research is being sought out.

1.3 Contributions

There are two main contributions in this thesis.

1.3.1 Survey on Fall Detection

One of the main contributions of this research study is a comprehensive survey on currently implemented fall detection algorithms. The survey classified different approaches into different categories.

1.3.2 Fall Detection using Optical Flow and Evaluation

The other main contribution of this research study is using the idea of optical flow in the research area of fall detection surveillance. The implementation of fall detection using optical flow is discussed with respect to the results achieved in terms of both accuracy and efficiency. The evaluation of fall detection compares our achieved results with other techniques.

1.3.3 System Spec Used

Processor	2.00GHz Core 2 Duo
Graphics Card	256 MB NVIDIA Quadro FX 570
Memory	2048 MB of DDR2 Ram
Coding Software	MATLAB R2009a

Operating System	Windows XP
Datasets	KTH, Fall Action, [148],[154]

Once video sequences are processed into MATLAB, a detection program using an evolved approach from optical flow estimation will analyze through motion estimation and classification of the data to detect whether or not it is a fall.

1.4 Thesis Outline

This thesis is organised as follows:

In Chapter 2 we presented a survey on fall detection approaches and techniques. Firstly, different types of fall are introduced, followed by the classification of fall detection methods. Secondly three different categories of fall detection approaches are reviewed. Finally, a conclusion is drawn and future direction of research in the area of fall detection is discussed.

In Chapter 3 we reviewed the some of the related literature of Background Subtraction and focus our attention on an easy to implement approach with efficiency and accuracy. The reviewed approaches are taken from well recognised conferences and journals. Firstly Background Subtraction is introduced in terms of its applications both in the industry and in the research community. Considerable range of methods and algorithms are reviewed followed by a brief discussion. Secondly different methods are implemented and their implementation is discussed with respect to the outputs achieved.

In chapter 4 we reviewed some of the related literature work of Optical Flow Estimation. Optical flow is introduced and its evolution is discussed through several techniques. We have reviewed wide range of optical flow algorithms from the very basics to the state of the art approaches. The reviewed approaches are discussed with the suggested

way forward to the future. Some approaches are implemented and their implementation is discussed in terms of the quality of the results. At the end of chapter 4, we have performed an evaluation of fall action detection using our implemented approach and compared it against two other techniques.

Finally Chapter 5 draws the conclusion of our study and points out possible future work.

Chapter 2

A Survey on Fall Detection: Principles and Approaches

In this chapter, we give a comprehensive survey of different systems for fall detection and their underlying algorithms. Fall detection approaches are divided into three main categories: wearable device based, ambience device based and vision based. These approaches are summarised and compared with each other and a conclusion is derived with some discussions on possible future work.

2.1 Introduction

Falls are a major cause of fatal injury especially for the elderly and create a serious obstruction for independent living. Statistics [62] show that falls are the primary reason of injury related death for seniors aged 79 or more and the second leading cause of injury related (unintentional) death for all ages. The demand for surveillance systems, especially for fall detection, has increased within the healthcare industry with the rapid growth of the population of the elderly in the world. It has become very important to develop intelligent surveillance systems, especially vision-based systems, which can automatically monitor and detect falls. it has been proved that the medical consequences of a fall are highly contingent upon the response and rescue time. Thus, a highly-accurate automatic fall detection system is likely to be a significant part of the living environment for the elderly to expedite and improve the medical care provided whilst allowing people to retain autonomy for longer.

The quality of an individual's life is significantly affected by the levels of functional ability. Plenty of research has been done in this area to develop systems and algorithms for enhancing the functional ability of the elderly and patients. The maturity of cameras, sensors and computer technologies make such systems feasible. Such systems can not only increase the independent living ability of the elderly, by raising the confidence levels in a supportive care environment within the public sector, but also save on manual labour in terms of the presence of nurses or support staff at all times.

The rest of the paper is organised as follows. In Section 2, different types of fall are introduced, followed by the classification of fall detection methods. We review three different categories of fall detection approaches in Sections 3-5. Finally, we conclude and discuss future directions of research in Section 6.

2.2 Classification of Falls and Fall Detection Techniques

In this section, different kinds of falls are first identified. Specifying different types of falls help towards an understanding of the existing approaches. It also guides and contributes towards the design of new algorithms.

Different scenarios should be considered when identifying different kinds of falls: falls from walking or standing, falls from standing on supports, e.g. ladders etc., falls from sleeping or lying in the bed and falls from sitting on a chair. There are some common characteristics among these falls as well as significant different characteristics. It is also interesting to note that some characteristics of fall also exist in normal actions, e.g., a crouch also demonstrates a rapid downward motion. Noury *et al.* [61] and Yu [60] reviewed principles and methods used in existing fall detection approaches. These are the only review papers on fall detection and their scope is limited, which prompts us to write a comprehensive survey of recent fall detection techniques.

Figure 2.1: *Fall from Sitting on a chair (Frames from Simulated fall sequence)*

Existing fall detection approaches can be explained and categorised into three different classes to build a hierarchy of fall detection methods. Different methods under these categories are discussed further in the following sections. Fall detection methods can be divided roughly into three categories: wearable device based, ambience sensor based and camera (vision) based. Figure 1 depicts the classification of fall detection techniques.

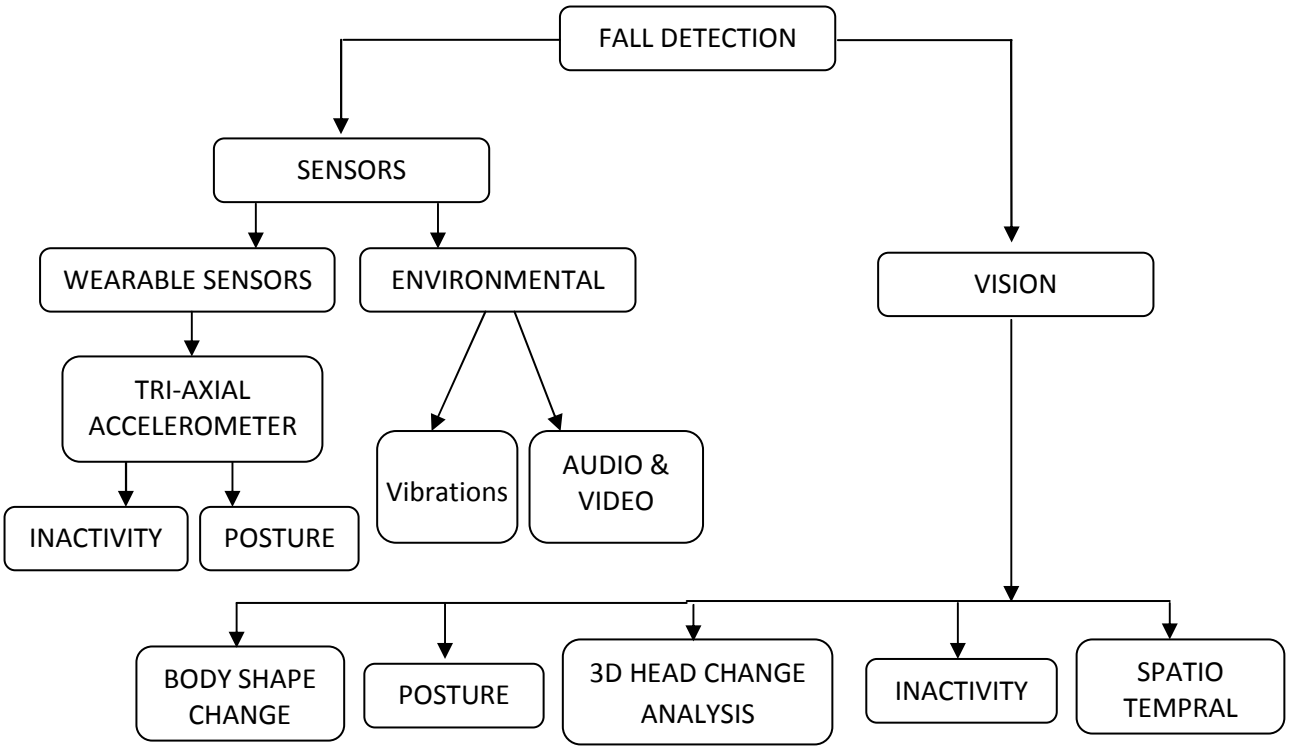

Figure 2.2: *Classification of fall detection methods*

Wearable devices can be further divided into posture based and motion based devices. Ambience devices can be further classified into presence and posture based sensors. And the camera (vision) based systems can be further categorised into three classes as shape change,

inactivity and 3D head motion. Most of the existing approaches share the same general framework. Data acquisition varies from one sensor to multiple sensors and from one fixed camera to multiple cameras and moving cameras. Figures 2 & 3 illustrate the general framework for a fall detection system based on Wearable & Ambient and Vision based approaches respectively.

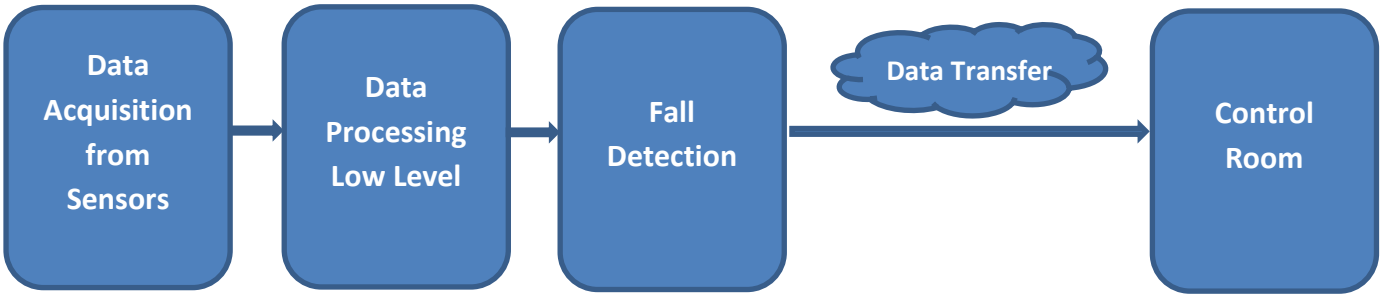

Figure 2.3: *Framework for existing wearable sensor and ambience based approaches*

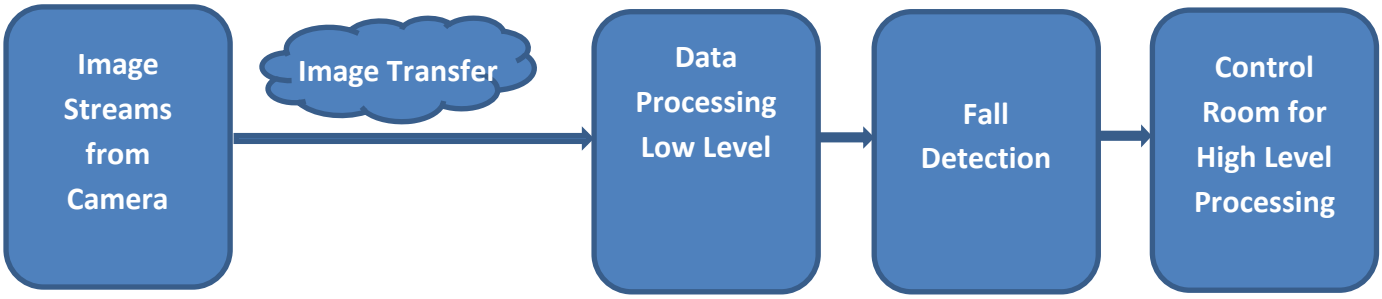

Figure 2.4: *Framework for existing vision based approaches*

2.3 Sensors Based Approaches

2.3.1 Wearable Device Based Approaches

Wearable device based approaches rely on garments with embedded sensors to detect the motion and location of the body of the subject. In the following we summarise the different methods.

2.3.1.1 Accelerometery

Technological developments have yielded devices that can measure activity using accelerometers. Accelerometry is composed of measure of acceleration of the body or parts of the body. It is one of the most extensively-used methods implemented for measuring physical activities to monitor activity patterns. Merryn et al. [42] used an integrated approach of waist-mounted accelerometry. A fall is detected when the negative acceleration is suddenly increased due to the change in orientation from upright to lying position. A barometric pressure sensor was introduced by Bianchi et al. in [40], as a surrogate measure for altitude to improve upon existing accelerometer-based fall event detection techniques. The acceleration and air pressure data are recorded using a wearable device attached to the subject's waist and analysed offline. A heuristically trained decision tree classifier is used to label suspected falls. Estudillo-Valderrama et al. [35] analysed results related to a fall detection system through data acquisition from multiple biomedical sensors, then processed the data with a personal server. The hardware and software design issues are clearly discussed when processing of bio-signals is involved during analysis. A wearable airbag was incorporated by Tamura et al. in [36] for fall detection by triggering airbag inflation when acceleration and angular velocity thresholds are exceeded. The system design consists of an accelerometer and a gyro sensor. Such a fall detection system can be very useful, especially at construction sites etc., for reducing fall related injuries.

Chen et al. [8] created a wireless, low-power sensor network by utilising small, non-invasive, low power motes (sensor nodes). The on-board device performs the sampling of acceleration sequentially, thus reduces the burden on the network. The dot product of acceleration vectors, from the orientation information, produces the angle of change during the fall event. The acceleration vectors are calculated using the average across a one-second

window. The accelerometric data were analysed by Narayanan et al. in [19] and a platform based on the "PreventaFall" ambulatory monitor (PFAM) and MiiLink data portal (MiiLink) can be used to monitor the accelerometric data. In [21], Wang et al. applied reference velocities and developed a system that uses an accelerometer placed on the head. The reference velocity is calculated using the backward integration of accelerations. By using the reference velocity and a predefined threshold, falls are distinguished from normal daily activities.

2.3.1.2 Fusion of Accelerometry & Posture Sensors

Physiological responses such as varying heart rate or blood pressure may result from physical activity and changes in body position. That makes the assessment of motion and posture a key factor in an ambulatory monitoring environment. Acceleration vectors were represented in a 3D space in [6] when Luo et al. implemented a group of sensors on a belt that filter noisy components with a Gaussian filter and generate a three dimensional body motion model that can be related to various body postures and the accelerometer's outputs. A two-axis accelerometer with a posture sensor was used in [27] for fall detection. The authors developed a wrist-worn prototype that integrates a health monitoring device with tele-reporting functionality for emergency telemedicine that contains a fall detector. The measured bio-signals have limited fidelity because the wrist area has limited body contact. This shortcoming could be overcome with further development in the posture sensor.

Ghasemzadeh et al. analysed machine learning and statistical techniques in [37] to create a physiological monitoring system that collects acceleration and muscle activity signals and performs analysis on those signals during standing balance. The objective of this system is to assess the behaviour of the electromyogram (EMG) signals to interpret the activity of postural control systems in terms of balance control.

2.3.1.3 Inactivity with Accelerometry

Accelerometry provides detailed information of measurement of physical activity and inactivity. This information can be used to measure more comprehensive relationships among movement frequency, intensity and duration. An array of relatively cheap infrared detectors was used by Sixsmith et al. in [24] for the design of a wearable system called Smart Inactivity Monitor using Array-Based Detectors (SIMBAD). The target motion was analysed to detect characteristic dynamics of falls. Inactivity periods were also monitored and compared within the viewing field with a map of acceptable periods of inactivity in different locations. Ghasemzadeh et al. in [38] implemented a similar approach for inertial sensor nodes that constructs motion transcripts from biomedical signals and identifies movements by taking collaboration between the nodes into consideration. The system relies on motion transcripts that are built using mobile wearable inertial sensors.

Srinivasan et al. [17] and Lee et al. [19] both used motion sensors along with wireless accelerometer sensor modules to monitor general presence or absence of motion. A smart fall sensor was designed by Noury et al. for fall detection in [3]. The software application transmits the data remotely through the network as well as exploiting data locally. The data are further analysed to determine the current state such as lying after a fall, sleeping, walking, etc.

2.3.1.4 Tri-axial Accelerometry

Tri-axial accelerometers are designed for simultaneous detection of acceleration in three axial directions. Lai et al. in [41] combined several tri-axial acceleration sensor devices for joint sensing, when an accidental fall occurs. The model transmits the information fed by the sensors which are distributed over various body parts. The system can determine the possible occurrence of a fall when the acceleration significantly exceeds the usual acceleration range. The impact acceleration and normal acceleration can be compared to determine the level of injury. Inertial sensors and the data logging unit are combined by Wu et al. in [30] to develop

a portable pre-impact fall detection system. The inertial sensor unit consists of accelerometers and tri-axial angular rate sensors. The inertial frame vertical velocity is the key variable that detects the fall prior impact and is applied using a threshold detection algorithm. Adaptive thresholding has been quite successful for the reduction of false positives.

Embedded intelligence was employed in [7] for the design of a system that performs the vast majority of signal processing on-board the wearable unit. The tri-axial accelerometer output is acquired from the portable unit containing an embedded microcontroller and the tracking information regarding the user's motion is transmitted to a local receiver unit. The analysis of acceleration thresholds in [16] was carried out where Kangas et al. used acceleration thresholds to detect falls using tri-axial accelerometric measurements taken at the waist, wrist, and head. The threshold values for different parameters are adjusted to optimise the detection of falls.

The trunk angle change was observed in [45] when Boissy et al. applied motion sensors on subjects to derive impact magnitudes and trunk angle changes. Motion sensors were placed on the front and side of the trunk along with three dimensional accelerometers. The deceleration as hitting the ground and trunk angle change in relation to hitting the ground represent two separate events. The fall detection algorithm is able to identify these two events as they are common to most falls. Wolf et al. [46] followed a popular low cost approach of a tri-axial accelerometer with a wireless transceiver. The algorithm is very similar to other accelerometric approaches discussed in this survey as data acquired from accelerometers are transmitted through a wireless transceiver for further sophisticated analysis. The algorithm applies acceleration thresholds to detect falls.

Zhang et al. [49] applied a similar idea of wearable tri-axial accelerometers for fall detection but with the introduction of non-negative matrix factorization (NMF). The method uses the vertical axis of the human body and acceleration sequences as input vectors. Vector

decomposition is performed through NMF. Finally a fall occurrence is determined via the k-nearest neighbour algorithm. Interestingly, as opposed to [49], Zhang et al. in [51] used a cell phone with a tri-axial accelerometer embedded in it. Data pre-processing is performed using 1-class support vector machine (SVM) and the wireless channel for Internet connection. Classification is achieved through the k-nearest neighbour (*k*-NN) algorithm and kernel fisher discriminant (KFD) analysis.

2.3.1.5 Posture Based

Multichannel accelerometry can be used to distinguish between posture and basic motion patterns. Body orientation as posture is measured to detect falls. Kaluza et al. [52] presented a posture-based fall detection algorithm. Falls along with abnormal behaviours are detected based on the ideology of reconstruction of an subject's posture. Small inexpensive wireless tags are placed on body parts, such as hips, ankles, knees, wrists, shoulders and elbows, identifying them as significant places. The locations of the tags are detected by the motion capture system. The posture is reconstructed in a 3D plane after locating the tags. Acceleration thresholds along with velocity profiles are applied in the fall detection algorithm.

Kangas et al. carried out study with the aim of developing a new fall detector prototype in [56] based on fall associated impact and end posture. A waist-worn tri-axial accelerometer, transceiver and microcontroller unit is used for data acquisition, transmission and processing. Sensitivity and specificity are also defined with respect to different fall detection algorithms. Sensitivity and specificity are achieved based on fall associated impact and end postures. Some backward falls cannot be detected by impact monitoring. This may partly be caused by the study set-up with intentional falls.

2.3.1.6 Discussion on Wearable Devices

Wearable devices have their advantages as well as disadvantages. The biggest advantage remains the cost efficiency of wearable devices. Installation and setup of the design is also not very complicated. The devices are relatively easier to operate. The disadvantages include intrusion and fixed relative relations with the subject, which could cause the device to be easily disconnected. Such disadvantages make wearable devices an unfavourable choice for the elderly.

2.3.2 Environmental Sensors Based Approaches

Environmental sensors based devices attempt to fuse audio and visual data and event sensing through vibrational data.

2.3.2.1 Audio & Video

Image sensing and vision-based reasoning were presented in [2] by Tabar et al. for verification and further analysis of sensor-transmitted events. A bridge like operation via a wireless badge node is created between the user and the network. The badge node detects falls through event sensing functions. Along with fall detection it also creates a voice communication medium between the user and the Monitoring Control when the system detects a problem and alerts the control. The monitoring control continuously tracks the approximate location of the user using signal strength measurements via the network nodes. A fusion of image sensing and network nodes is created for further analysis of the field-of-view and the user's status during fall detection.

Zhuang et al. [54] proposed a different approach to the method in [48] using the audio signal from a single far-field microphone. A Gaussian mixture model (GMM) super vector is created to model each fall as a noise segment. The pairwise difference between audio segments is measured using the Euclidean distance. The kernel between GMM super vectors constitutes the support vector machine employed for the classification of various types of

noise and audio segments into falls. Accelerometric data with video streams are used in the algorithm in [57]. Wearable sensors transmit the motion data wirelessly. Classification is achieved from acquired data using support vector machines (SVM) to detect fall events. Finally video streams are transmitted from a context-aware server. The image sequences are coded according to both the patient and the network status.

2.3.2.2 Event Sensing Using Vibrational Data

The detection of events and changes using vibrational date can be useful in many ways such as monitoring, tracking and localisation etc. A completely passive and unobtrusive system was introduced by Alwan et al. in [10] that developed the working principle and the design of a floor vibration-based fall detector. Detection of human falls is estimated by monitoring the floor vibration patterns. The principle is based on the vibration signature of the floor. The floor's vibration signature generated by the human fall is different from normal activities, such as walking. A special piezoelectric sensor is used which is coupled to the floor surface. A battery powered pre-processing circuit alongside is employed to analyse the vibration patterns. A binary fall signal can be generated in the case of a fall event.

A slip-fall detection system, using the sliding linear investigative platform, was proposed in [23] by Robinson et al. Classification of acceleration thresholds has been used to identify true slip-falls. Data such as tri-axial head accelerations and the centre of pressure in terms of psychophysical response are measured. Slip-fall vibrations are distinguished easily due to noticeable small vibration translations. The movement parameters require precise control and its advantages are discussed in terms of usefulness. The concept of floor vibrations with sound sensing is unique in its own way in [34]. Pattern recognition is applied to differentiate between falls and other events. Shock response spectrum is one of the key special features used in classification. The system is unique in the detection of falls in critical cases, such as

an subject being unconscious or in a stressful condition. The algorithm can be further developed with the calibration of the floor

Rimminen et al. in [39] proposed to use a floor sensor based on near-field imaging. The shape, size, and magnitude of the patterns are collected for classification. A set of features is computed from the cluster of observations. The postural estimation is implemented using Bayesian filtering instead of the features being classified directly. The system has problems with test subjects falling onto their knees as this produces a pattern very similar to a standing person. Toreyet al. [48] fused the multitude of sound, vibration and passive infrared (PIR) sensors inside an intelligent environment equipped with the above fusion elements. Wavelet based feature extraction is performed on data received from raw sensor outputs. Regular and unusual activities, such as falls, are used for training the Hidden Markov Models (HMM). The process of fusion is applied to all outputs from sensors to detect falls.

Nyan et al. in [59] distinguished backward and sideway falls from normal activities using gyroscopes (angular rate sensors). The gyroscopes are securely placed on different positions, such as underarm and waist. The angular rate is measured for normal activities and falls in lateral and sagittal body planes. A high speed camera is used to capture video image sequences of motion for body configuration analysis in the event of a fall. High speed cameras have the frame rate of 250 frames per second. The fusion of high speed camera images and gyroscope data is synchronised. Gyroscopes rely on the idea of acceleration thresholds to differentiate fall events from normal activities.

2.3.2.3 Discussion on Environmental Sensor Devices

Most ambient device based approaches use pressure sensors for subject detection and tracking. The pressure sensor is based on the principle of sensing high pressure of the subject due to the subject's weight for detection and tracking. It is a cost effective and less intrusive for the implementation of surveillance systems. However, it has a big disadvantage of sensing

pressure of everything in and around the subject and generating false alarms in the case of fall detection, which leads to a low detection accuracy.

2.4 Vision Based Approaches

Cameras are increasingly included, these days, in in-home assistive/care systems as they convey multiple advantages over other sensor based systems. Cameras can be used to detect multiple events simultaneously with less intrusion.

2.4.1 Spatiotemporal

Shape modelling using spatiotemporal features provides crucial information of human activities which is used to detect different events. Image analysis requires efficient and accurate shape modelling methods. Foroughi et al. [4] developed a method for detecting falls using a combination of the Eigen space approach and integrated time motion images (ITMI). ITMI can be described as a spatiotemporal database that contains motion information and time stamps of motion occurrence with an emphasis on the final action. Feature reduction is performed using the Eigen space technique. Feature vectors obtained from the feature reduction process are then fed to the Motion Recognition and Classification Neural Network classifier that can deal with motion data robustly. In [33], a mobile human airbag release system was designed for fall protection for the elderly. The system consists of 3D MEMS accelerometers, gyroscopes, a Micro Controller Unit and a blue-tooth module. The subject's motion information is recorded by the accelerometers. A high speed camera is used for the analysis of falls. Gyro thresholding is applied to detect a lateral fall. The classification of falls is performed by using a support vector machine (SVM) classifier. The real time fall detection system contains an embedded digital signal processor.

An asynchronous temporal contrast vision sensor was developed for fall detection in [22]. The method extracts changing pixels from the background and reports temporal contrast

(compared to a threshold), which is equivalent to the change in image reflectance in the presence of constant lighting and finally an instantaneous motion vector computation reports fall events. The device requires a power socket nearby that makes the deployment of the system very simple. The motion detector protects the patient's privacy because no data are communicated until an emergency is detected.

2.4.2 Inactivity/Change of Shape

In this section, we describe algorithms based on shape change analysis as well as inactivity detection. From tracking data, McKenna et al. in [1] automatically obtained spatial context models by using the combination of Bayesian Gaussian mixture estimation and minimum description length model for the selection of Gaussian mixture components through semantic regions (zones) of interest. Ceiling-mounted visual sensors are used to reduce the influence of occlusion. Human-readable summaries of activity are produced and unusual inactivity is detected through the resulting contextual model. The contextual model can differentiate unusual activities, such as falls, from normal activities. Foroughi in [5] applied an approximated ellipse around the human body for shape change. Projection histograms after segmentation are evaluated and any temporal changes of the head position are noted. Segmentation of moving subjects is obtained initially, and the next step involves extracting features by carrying out shape change analysis in the video sequence through an approximated ellipse around the human body. Further analysis of projection histograms (both horizontal and vertical) and temporal changes of the head position are carried out to extract feature vectors with optimised information. Extracted feature vectors are then fed to a MLP Neural Network similar to the earlier approach of Foroughi in [4] for classification of motions and fall events. Miaou et al. [9] captured images using an omni-camera called MapCam for fall detection. The personal information of each individual, such as weight, height and electronic health history, is also considered in the image processing task. Object

segmentation is performed by using methods such as background subtraction. Noise reduction is applied during and after segmentation for accuracy. A bounding-box like approach is used by creating a rectangle enclosing the subject. The ratio of height to width of the subject is calculated at each frame. The ratios are then analysed by considering the last six consecutive image frames and result, in total, in five ratio changes between two adjacent frames. The occurrence of fall becomes likely if the first three ratios are all greater than 1 and the last three ratios are all less than 1. The system's decision of fall detection is based on the last two ratio changes with respect to a threshold. Each individual, due to different body figures, has different ratio changes between normal and fall states, and the Body Mass Index (BMI) value is used to adjust the threshold. Therefore, the system is flexible enough to adjust the detection sensitivity on individual basis.

Tao et al. [11] developed a detection system based on Miaou et al.'s [10] approach of using background subtraction (another approach based on the shape change analysis algorithm) but with an addition of foreground extraction, extracting the aspect ratio (height over width) as one of the features for analysis, and an event-inference module which uses data parsing on image sequences. A simple two-state machine in combination with falling motion inference is implemented. The two states are "standing/walking" and "falling down". Rougier et al.'s [14] approach is based on a combination of motion history image (MHI) and human shape variation. The MHI is an image projected from multiple motion images. The recent information of motion in an image sequence is represented by the pixel intensity and the most recent motion is more emphasised than that happened in the past. Shape change analysis in combination with inactivity analysis is performed using the approximated ellipse.

Fall incident detection in a compressed-domain is discussed in [15]. Object segmentation within the compressed domain is applied for the extraction of moving subjects using the combination of global motion estimation and local motion clustering. The three extracted

features used are: short time period range of fall occurrence, significant and rapid centroid change of the falling human, and the vertical projection histogram of the falling human. Fleck et al. [32] proposed a very unique idea of processing the stream at the point of sight and transmitting the processed stream to the control leaving no further processing to be done except for the higher level abstraction. The system design consists of a distributed network that contains smart cameras. Geo-referenced tracking and activity recognition are performed simultaneously, embedding in each camera node. An FPGA module and a Power PC processor are used for low level computations. The efficiency of the automated video analysis algorithm plays an important role towards the performance of the system. The proposed system could be further developed with self-diagnostic tools. Further improvements such as comprehensive processing and better decision making could be used as one of the major research directions for future development.

Rougier et al in [64] proposed a classification method for fall detection by analyzing human shape deformation. Segmentation is performed to extract the silhouette and additionally edge points inside the silhouette are extracted using a canny edge detector for matching two consecutive human shapes using shape context. The mean matching cost and Procrustes analysis is applied for shape analysis. Both of these methods contribute in quantifying abnormal shape deformation. "*A fall is characterized by a peak on the smoothed full Procrustes distance curve or mean matching cost curve followed by a lack of significative movement of the person just after the fall.* [64]" GMM (Gaussian Mixture Model) Classifier is implemented to detect falls. Further computation of the sensitivity, specificity, accuracy and the error rate obtained from GMM classifier is performed for the analysis. Ensemble classifier is used later to combine the results of all cameras. The mean matching cost and the Procrustes analysis reduced the error rate to 4.6% and 3.8% respectively. Further

development can be made by the introduction of a remotely activated method which learns the inactivity zones automatically to improve the recognition results.

Wu et al. [43] uniquely identified velocity profile features between normal and abnormal activities, such as falls, for automatic detection. The fall activities contain forward and backward falls from standing and tripping, etc. Horizontal and vertical velocities are measured at different locations of the trunk. The trend of velocity increase shows an interesting pattern as it increases in one direction but does not in another. Two different characteristic patterns for falls are exhibited by the horizontal and vertical velocities. Differentiating falls from normal activities during the descending phase of falls heavily depends on these characteristics, i.e. the change in magnitude and timing when the change in magnitude occurs in both velocities.

Willimas et al [65] detected and localised falls through distributed network of overlapping smart cameras. The system design is composed of battery powered camera sensor nodes of same type on a single tier. Each node contains a camera sensor, an on board processor with RAM, wireless radio communication and flash memory. The system design works on a principle assumption of at least one leader node with calibrated camera to the world that has known homography between its image coordinates and the world coordinates of the 2D ground plane. Human detection is performed through background subtraction. Planar homography is estimated through the normalized Direct Linear Transform that gathers point correspondences between the two images of interest. Fall is detected through the feature of aspect ratio (width of the person divided by height) extracted from segmentation. Simple thresholding is used to classify the fall. Localization is achieved via automatically system estimated pair-wise camera homographies. Transformation of fall point into the destination node's image coordinates is performed before every hop, until it reaches the leader and transformation into world coordinates is achieved through intelligent weighting procedure

(the inverse of the cumulative mean squared transformation). The system works under an assumption of at most one moving person about the environment and fairly stable lighting conditions due to the processor and RAM constraints. The idea of simplistic algorithmic design on low power devices is appealing to an extent but it is prone to false positives although the evaluation results show very low false positive rate due to the data set used.

Vishwakarma in [44] followed an adaptive approach for the detection of moving objects by using background subtraction as well as bounding boxes. The described fall model is based on feature extraction analysis, detection and classification. Features extracted include horizontal and vertical gradients, aspect ratio and the centroid angle to the horizontal axis of the bounding box. Falls are confirmed when the angle reaches a value less than 45 degrees. The image stream from the thermal detector is monitored by the fall detector proposed in [47]. The analysis is focused on measuring vertical velocities of the subject using the coloured segmentation algorithm and identifying features in the pattern of velocities over time. These velocity estimates are then fed into a neural network-based fall detector that identifies the characteristic patterns of velocities present during a fall. Cucchiara et al. [53] instead applied a multi-camera system for image stream processing. The processing includes recognition of hazardous events and behaviours, such as falls, through tracking and detection. The cameras are partially overlapped and exchange visual data during the camera handover through a novel idea of warping "*people's silhouettes*". The video server (multi-client, multi-threaded transcoding) transmits sequences for further processing to confirm the validity of received data. The bandwidth usage is optimised through event-based transcoding and semantic methods. Anderson et al. [58] used a multi-camera system, similar to Cucchiara et al. [59], and applied silhouettes to form a 3D model of the human subject. The membership degree of the subject is measured using fuzzy logic to a pre-determined number of states at each image. The fall detection method consists of two levels. The first level deduces the

number of states for the object at each image. The second level deals with linguistic summaries of the object's states called "*Voxel Person*". Further derivations are performed regarding the activity.

2.4.3 Posture

The use of posture information contributes towards accurate fall detection. Different body positions are used to calculate postures. Specific types of postures are identified and localised in image sequences. Cucchiara et al. [25] carried out analysis of human behaviours by classifying the posture of the monitored person and consequently detecting falls. Projection histograms are calculated and compared with the stored posture maps (training). The tracking also deals with occlusions. Accuracy levels achieved are up to 95%. A different posture classification approach based on a neural fuzzy network was introduced in [32] by Juang et al. Standing, bending, sitting, and lying are the postures used for classification. After segmentation (background subtraction and extraction), projection histograms are used and discrete Fourier transform is applied. A neural fuzzy network is used for classification. The results could be improved with better segmentation, such as better elimination of shadows and filtering the illumination influence.

Thome et al. [12] developed a Hierarchical Hidden Markov Model (HHMM) with two layers for modelling motion. The first layer has two states, an upright standing pose and lying. Fall detection, in terms of sudden change, has dedicated motion features from the first layer. 3D angle relationships and their image plane projections have been carefully observed. After performing an initial image metric rectification, theoretical properties are derived from binding the error angle for a standing posture during the image formation process. This simply differentiates other poses as "non-standing" ones. Thus falls can be accurately detected from other actions, such as walking or sitting. Computer vision systems typically use

cameras only for recording and capturing video signals. Once transmission of the stream of images is completed, data processing is performed.

A support vector machine (SVM) approach is applied in [29] by Khandoker et al. for the classification of balance impairments, such as risks of fall for the elderly, based on the minimum foot clearance (MFC) principle which is used on the samples taken while walking on treadmill during training. Foot clearance data was collected using a 2-D Motion Analysis system. Unobstructed walking sequences were recorded for foot motion using a high speed camera. The SVM model builds on the effectiveness of multi-scale analysis of a gait variable which is based on a wavelet in comparison to histogram plot analysis during feature extraction. There is a clear indication of better performance of the SVM model based on multi-scale exponents (by wavelet analysis) in the results than the model based on MFC statistical features.

2.4.4 3D Head Position Analysis

Head position analysis is based on head tracking that determines the occurrence of large movement within the video sequence. Different state models are used to track the head based on the magnitude of the movement information. Rougier and Meunier in [50] obtained image streams from a monocular camera. This methodology of fall detection is based on 3D head trajectories and the idea that the subject's head remains visible in the image sequence and undergoes a large movement when a fall occurs. The 3D ellipsoid is used for bounding around the head. The 3D ellipse is a projection of ellipses in 2D image planes. A particle filter extracts the 3D head trajectory for tracking. The 3D head trajectory also contains features, such as 3D velocities, which are applied for fall detection.

Hazelhoff et al in [63] aimed at incidents involving falls in unobserved home situations by presenting the design and real time implementation of a fall detection system. The design involves segmentation of foreground subjects in the image streams obtained from two fixed,

uncalibrated, perpendicular cameras. The direction of the main axis of the body and the ratio of the variances in x and y direction are calculated through principal component analysis (PCA). A head tracking module is used for human detection as well as increasing the robustness of the system. Head position is estimated as blob using Gaussian skin-colour model and is tracked by searching for skin-coloured blobs nearby the head position. The classification is performed through a Gaussian multi-frame classier. The system shows accuracy levels of 100% on un-occluded video sequences. The addition of occlusion only reduced the accuracy to 90%. Non perfect segmentations with the addition of occlusions reduced the accuracy to 44%. The system can be improved with further development of an advanced tracker.

In 3D head motion based analysis, the principle involving faster vertical motion than horizontal motion in a fall was proposed by Jansen and Deklerck in [13]. The method uses information extracted from images obtained using three dimensional visual approaches in combination with a context model. The contextual model interprets the fall occurrence differently. It depends on the time, location and duration of the fall event.

2.4.5 Discussion on Vision Based Approaches

Vision based systems tend to deal with intrusion better than other approaches. Recent research in computer vision on surveillance indeed provides a practical and complex framework. Most of the emphasis in the context of surveillance in computer vision is dedicated to methods with the ability of real time execution using standard computing platforms and low cost cameras. The methods with capability of dealing with robustness still leave a wide open area for further research and development. Video analysis of human behaviour containing semantic description of the activities belongs to higher level abstraction and lower level represents the segmentation of motion along with feature extraction in computer vision.

Current posture related methods are classified depending on the use of a model. 3D techniques are not mostly automatic and usually require manual initialisation. Generally model dependent methods obtain posture relatively easily and are robust to occlusion to an extent after labelling the body parts. Many of the body modelling techniques are 2D models. Comparatively other non model based techniques compute the posture using features.

Models, learned through extended observation such as the interpretation of human activities in a scene, provide contextual representation of the activities. These models provide recognition and summary of the events and activities. Several techniques have been developed to learn these models automatically as manual techniques are useful to an extent. The range involved dealing with the complexity and abstraction of comprehensive activity and event analysis to fall detection automatically. However interpreting human behaviour and resulting pattern analysis depends on the choice of level of abstraction.

In 3D head motion analysis methods, principle of faster vertical motion than horizontal motion during a fall is applied. The head is initially located and then using filters 3D head position is estimated. The idea of using appropriate thresholds to distinguish fall from other actions is applied by computing vertical and horizontal velocities of the head [60].

Though some of the implementations discussed earlier have shown a diverse pattern when it comes to dealing with image sequences, still there is plenty of room for further development in this area. Fall detection has still not been implemented using the total optical flow of the image sequence or specifically analysing the optical flow of the subject after the subject has been tracked and located. Further higher level abstraction can be applied on calculated optical flow to achieve higher levels of accuracy and robustness.

2.5 Discussion

We have reviewed different techniques for the detection of a fall event. Table 1 lists various characteristics of those approaches. A comprehensive and robust fall detection system should

possess both high sensitivity and good specificity. The existing approaches have not comprehensively satisfied the accuracy as well as robustness of a fall detection system. However, the existing approaches do provide a framework to further develop techniques as well as modify the existing algorithms to achieve better performance.

As discussed earlier, sensor based approaches lack consistency when it comes to providing highly accurate automatic fall detection systems. Higher accuracy levels have been achieved to an extent using multi-dimensional combination of physiological and kinematic parameters. Further research and development should continue in terms of making the design more automatic and without much intervention.

Vision based approaches in comparison to others are certainly the area to look forward to. Most of the existing vision based approaches lack flexibility. These approaches are often case specific and dependent on different scenarios. There is a need for a reliable and robust generic fall detection algorithm. Both ambience and sensor based approaches share a common disadvantage, generally, of subject data not being visually verified by the control or care service provider for accuracy.

Continuous surveillance through vision/camera and sensor based systems also introduces some ethical issues concerning the respect of confidentiality and privacy and also the risk of dependency of the subject on the technology. A common definition of a fall and of a fall detection system would certainly benefit the research community as well as the healthcare industry for the evaluation of fall detection systems. From a research perspective, there are issues relating to the availability of data sets of falls for training as well as evaluation. A comprehensive data set containing different scenarios of falls with different camera angles and with both static and moving cameras should be publicly available for the scientific community for the development and research purposes.

Approach	Category	Cost	Intrusion	Accuracy	Setup
Wearable Devices	Tri-Axial	Cheap	Yes	Scenario Dependent	Easy
	Posture	Cheap	Yes	Scenario Dependent	Easy
	Inactivity	Cheap	yes	Scenario Dependent	Easy
Ambient	Audio	Cheap to Medium	yes	Scenario Dependent	Easy / Medium
	Video	Cheap to Medium	yes	Scenario Dependent	Easy / Medium
Vision based	Body Shape Change	Medium	Low /dependent	Higher / Non Specific	Medium
	Posture	Medium	Low /dependent	Higher / Non Specific	Medium
	Inactivity	Medium	Low /dependent	Higher / Non Specific	Medium
	Spatiotemporal	Medium	Low /dependent	Higher / Non Specific	Medium
	3D Head Change	Medium	Low /dependent	Higher / Non Specific	Medium

Table 2.1: *Brief comparison of different categories of fall detection approaches*

Chapter 3

Background Subtraction

In this chapter the methods of background subtraction is reviewed and our implementation of background subtraction is demonstrated. Background subtraction is a primary step in majority of the surveillance related applications in computer vision. The successful extraction of foreground objects via background subtraction can lead to a considerable reduction of computational time of the rest of the algorithm due to reduced searching regions. A typical detection system is constructed through searching regions containing motion, reducing noise, lighting changes and object tracking in a video sequence. The most common technique used for separating an object's motion from the rest of the scene with static camera setup is background subtraction.

Background subtraction produces a very useful medium of object segmentation with static background for applications used for surveillance and tracking etc. Distinctive object segmentation and precise tracking exhibits a greater challenge with respect to low computational complexity. Obtaining a background estimate with statistical computation is the basis of the idea. Combinations of semantic procedures have been used traditionally to reduce noise within the background subtraction process. Such operations are useful in successfully isolating foreground objects but the constancy around the borders of the segmentation can be affected by a noisy background. Background subtraction can significantly improve computational time and efficiency due to limited implementation regions for searching and noise reduction. Environmental changes, shadowing and highlight effects are factors that need catering for when dealing with indoor surveillance applications. Factors such as occlusions, variations in appearance, fast moving objects and denser moving

objects also require attention to detail to an extent to get a good quality of foreground mask after the subtraction of background.

Most background subtraction methods are based on a hypothesis of fixed background. The moving objects are extracted using this assumption. Further assumptions such as colour differencing are applied. That is, the moving object's colour is considered to be different than the fixed background. The following equation describes the most commonly used background subtraction techniques;

$$X_t(s) = \begin{cases} 1 \; if \; d(I_s, t, B_s) > \tau \\ 0 \quad otherwise \end{cases} \tag{3. 1}$$

"Where X_t is the motion label field at time t (also called motion mask), d is a distance between I_s the video frame at time t at pixel s and B_s the background at pixel s; τ is a threshold." [127]

3.1 Related Work

The critical and foundational step in video surveillance, pedestrian detection and tracking etc is the classification of the moving subjects from the video sequence. The typical method of detecting the moving subjects used is "Background Subtraction". Background subtraction is generally implemented by comparing each video frame with a reference or background model. Diverging pixels from the background in the respective frame are acknowledged as moving subjects. The acknowledged foreground pixels can be further developed to track and locate the subjects. The accuracy of the extracted foreground is in terms of its corresponding relation to the actual movement in the video sequence as background subtraction is implemented mostly as the first step in designing computer vision algorithms. There are several methods of background subtraction available but it is still complicated to classify moving subjects in a complex environment [66]. Many popular

methods specifically model the background with the assumption of a bootstrapping phase where the presentation of the image consists of the background frames only.

3.1.1 GMM Based Approaches

Jwu-Sheng Hu et al [69] presents background subtraction involving shadow and highlight removal for indoor environmental surveillance. Precise extraction of Foreground regions is performed despite illumination variations and dynamic background. This current Gaussian mixture model (GMM) based background subtraction model is advancement of their earlier work [90] on colour-based probabilistic background model (CBM). Short-term colour-based background model (STCBM) and the long-term colour-based background model (LTCBM) are further extracted. A gradient-based version of the probabilistic background model (GBM) is built using the two extracted colour based models to improve its flexibility. A cone-shape illumination model (CSIM) as a new dynamic cone-shape boundary in the RGB colour region is presented to differentiate pixels among shadow, highlight, and foreground, thus solving the problem of Shadows and highlights from changes in illumination. The method which is a combination of the CBM, GBM, and CSIM distinguishes the background which can used to detect abnormal conditions. This method improves the ability to handle illumination variation both locally and globally. The accuracy of the background subtraction improves with improved illumination handling.

An adaptive background mixture model is proposed by Stauffer et al [70] by modelling each pixel as a mixture of Gaussians and using an on-line approximation to update the model for tracking. The Gaussians are computed through simple analysis every time the features of the Gaussians are updated. Non corresponding pixel computations to the background Gaussians are assembled through connected components. Finally a multiple hypothesis tracker is implemented for frame to frame tracking of the connected components.

Scenes containing visually overlapping subjects as well as presence of higher number of subjects affected the accuracy of the tracking system. The tracker was consistent with the fast lighting variations. A generic solution was proposed by Sun et al [72] in the form of a hierarchical Gaussian mixture model (HGMM) to handle sharp changes during background subtraction. The GMM is quite capable of dealing with gradual illumination changes. A HGMM works in a top-down style. State models of different scales are exercised because sharp changes cannot be alone identified at the pixel level. The intermediate scale's GMMs are constituted in a similar way to identify sharp changes in the partial scene of each of the frames. The lowest scale's implementation of pixel-wise GMM is quite capable of dealing with gradual illumination changes.. State models of different scales are used because sharp changes cannot be reliably identified at a pixel level. The intermediate scale's GMMs are constituted in a similar way to identify sharp changes in the partial scene on each of the frame. The lowest scale's implementation of pixel-wise GMMs is performed. All state models are arranged in a hierarchical mode accordingly and the HGMM responds quickly to sharp changes, although the system shows more robustness on estimation of background on a small sample set. Zivkovic [73] developed an efficient and simple adaptive approach for background subtraction using Gaussian mixture probability density. The parameters are constantly updated and the number of components of the mixture is constantly adapted by using recursive equations as an online procedure.

A boosted version of Gaussian Mixture Model (GMM) was proposed by Tang et al in [76]. The proposed version is developed with respect to spatio-temporal constraints. Furthermore, a method is devised to detect shadows and reduce noise. The GMM has the drawback of slower updating rate and initialisation procedure, which consumes times and space. In this implementation, GMM is boosted with the assumption of each channel's own variation within RGB colour space. One learning rate is used in the updated version to

achieve a quicker correspondence of the variation value. Also a threshold value is changed from 0.002 to a larger value as previously proposed for accurate background selection and faster place, thus making this version more efficient. The process of shadow elimination works on the assumption of no change in chromaticity of the shadow area and lower levels of pixel intensities in the shadow area. The threshold method is applied to compute the intensity of the current pixel compared with the background pixel. Noise reduction involves two steps: the first step uses neighbourhood information to eliminate small isolation point; the second step uses a bounding box to locate an subject's movement and all pixels outside bounding box are considered to be part of the background. Tang et al in [77] carried out further development to their previous work in [76]. The idea of using an energy function (Graph cut) is applied to retrieve the fragments which are wrongly segmented as the background. A foreground weight is added with single Gaussian models to build the energy function. Therefore, the background weight is updated when a pixel is not considered to be a part of the foreground. The method is further developed with the introduction of space sampling method to lower the computational load. Resolution of the image is lowered and GMM is applied. The binary mask at the end of the processing is zoomed it back to its original resolution using an interpolation method.

[78] Suo et al used a GMM to present their model number adaptive method to increase efficiency. An updating method involving adapting the learning rate is used for accurate segmentation of slow moving subjects and subjects that stop for a while during motion. Gaussian distributions are used to model each pixel by the Gaussian mixture model. The corresponding parameters of all the Gaussians are initialised with the same values. A small number of Gaussians for a non-modelled pixel have higher value of weight. The others have almost equal values of weights at the end of the sample sequence of the frames. After resetting the model number the computation is reduced considerably. The updating method

proposes a higher value of the learning rate of the distribution representing background and lowered value of the distribution representing the moving subjects. The updating method rapidly updates the background model when the subject stops moving and still distinguishes it as foreground with appropriate segmentation but considers it as a part of background if subject stops for a longer duration of time. In [83] Zivkovic et al analysed pixel-level background subtraction through recursive equations. The training model is updated in order to adapt to changes with a choice of a suitable time adaptation period. The estimates of the mean and the estimates of the variances describe the Gaussian components. To limit the influence of old data, an exponentially decaying envelope is applied. The proposed algorithm uses an on-line clustering method. Additional clusters of small weights exhibit intruding foreground subjects. The background model is approximated through the initial largest clusters. The subject is considered as part of the background if it doesn't move for longer duration as its weight value is increased. A simple non-parametric adaptive density estimation method is also presented. Cheng et al in [84] & [86] used a finite Gaussian mixture model to propose a flexible method to estimate the background model. MAP (maximum a posteriori) is considered to be a flexible criterion for the estimation of the parameters. The EM (Expectation Maximisation) procedure is applied for MAP estimation of mixture models parameters. The EM algorithm is based on the interpretation of identically distributed samples as incomplete data. The method achieves the optimal number of mixture components by driving the unnecessary components to elimination and concurrently estimating the mixture parameters. The GMM models both background scene and the foreground without discrimination. Some components are responsible for modelling the background scene and some for the foreground. Any pixel of the frame with standard deviation more than the threshold of 2.0 from any foreground components is recorded as part of background subtraction. The background model incorporates any changes to the scene

background through recourse to an online background model update. The update is same as for background model learning. A Multiple Gaussian component mixture through an online estimate is used to deal with the illumination change. The system effectively performs on outdoor, indoor and cluttered scenes.

Zhou et al modified the Gaussian mixture model (GMM) in [85] for background subtraction proposed by Stauffer et al in [70] and combined it with optical flow techniques with the support of temporal differencing to detect moving subjects. The modified approach uses the true mean value to accurately detect moving subjects. Background subtraction is achieved with pixel values more than 1.0-1.5 standard deviations away from the background distributions being marked as foreground. A range of coarse to fine views of the image is achieved by constructing a Gaussian pyramid for both the source and target images. Initial motion contains smaller pixel displacements and leads to an estimation at the coarsest level. The process of refining motion parameters is performed and continues until the finest level. Motion estimates are corrected through a temporal differencing approach. A simple fill-hole step is performed in post processing to deal with interior holes created when subject moves to a location previously occupied by a smaller subject after change detection. The approach provides reasonably precise located boundaries.

3.1.2 Non GMM Based Approaches

Mahadevan et al [67] proposed their idea for background subtraction in highly dynamic scenes by drawing inspiration from biological vision. Background subtraction is formulated as the complement of saliency detection. A saliency map is computed that classifies the background pixels at each location. A collection of spatio-temporal pieces are derived from each window as surround and center windows are centered at the location. The parameters of DT (an autoregressive generative model that represents the appearance of the stimulus) are

computed to obtain densities from center, surround and total windows. The saliency map is implemented with locations of less saliency threshold, which are allocated to the background. The background subtraction method depends on the approximate distinction of motion between the center and surround locations. This method doesn't require training phase and is completely unsupervised. [68] Bayona et al reviewed the existing background subtraction approaches with stationary foreground and divided them into different categories. Various stationary subject detection methods using background subtraction technique are discussed with their comparison in typical surveillance scenarios. Different methods are selected from allocated categories and evaluation is performed to analyse the advantages and disadvantages of those approaches. The analysis shows higher accuracy of stationary subjects detection using background subtraction in simpler scenarios whereas the evaluation analysis is heterogeneous for more complex scenarios. Generally methods based on the characteristics of the background subtraction model present low accuracy.

Haritaoglu et al [71] designed detection and tracking system to detect and track multiple people and monitor their activities in an outdoor environment. The system operates on monocular gray-scale video images/images from an infrared camera. The system learns and obtains the background model statistically to detect foreground subjects, even if there are moving foreground subjects in the field of view. The system consists of two steps to eliminate moving pixels from the background model. The first step involves the identification of stationary pixels through pixel-wise median filtering which is performed typically during 20-40 seconds to distinguish moving pixels. At each pixel location, standard deviation and median value of intensities are computed. The second step involves the formulation of an initial background model via the processing of stationary pixels. The system uses two different methods to update the background to adapt to illumination changes and physical changes in the background scene i.e. pixel-based update and object-based update

respectively. The system dynamically constitutes a change map to decide whether a pixel-based or an object based update method is required. The change map is composed of three main components: detection support map, motion support map and change history map. The system shows the capability to simultaneously track multiple people even with occlusion.

A texture based technique was presented by Heikkila et al in [74] for background modelling. Each and every background pixel is identically modelled which also leads to a high speed parallel implementation. Local Binary Patterns (LBP), a gray-scale invariant texture primitive statistic, is chosen due to its characteristics as a measure of texture. The background model for the pixel is composed of an adaptive LBP histograms group. The calculation of LBP histograms is performed as a feature vector over a circular region around the pixel. The histogram intersection is used during the evaluation phase. The measure possesses an inherent property to compute the common part between histograms. It explicitly disregards the features only occurring on one of the histograms. The proximity measure's threshold is a user defined parameter. Background processes do not necessarily generate all of the model histograms. The histogram's higher weight reflects the probability of being a background histogram. The final stage involves the updating step. The model histograms are sorted in descending order of their weights. The histograms with higher weights are selected as background histograms. The system deals considerably with illumination variations found common in natural scenes. The method efficiently computes the proposed features.

The idea of comparison between video frames and a stationary background model is presented in [75] by Migdal et al. The connection between the moving subjects and spatio-temporal occupancies, enforced by moving subjects, is capitalised upon by the model. To achieve this, Markov Random Fields (MRFs) are used and developed during the background subtraction process. MRFs present Probabilistic Graphical Models. The edges and nodes in MRFs present random variables and contingent affiliation between them respectively. Three

MRFs with different characteristics are used in total. Maximum A Posteriori (MAP) estimation contributes achieving segmentations. The computation of MAP estimation can be performed by applying the [89] Gibbs Sampler method.

[79] Unger et al developed an algorithm to perform background subtraction for an unconstrained motion of the camera, rotation, zoom and lens distortion. The algorithm is based on global motion estimation and a weighted summation of motion compensated images. A Quartic model is used to model lens distortion and complex motion. This model is an extended version of the Affine Transformation model that has higher order values to express tilt and distortion. Optical flow estimation is applied for the calculation of the point correspondences as an input to the estimation of the global motion model. The background image's pixel value is computed through calculations of a weighted sum of the pixels trajectory over time. Moving regions with small motion deviation, as predicted by global motion model, have a higher probability of being part of the background as compared to regions with higher motion deviation. Motion deviation of the background is modelled as a Rayleigh distribution [90]. Post processing method uses colour segmentation on the original image. The luminance of the background subtracted image is averaged within the segments. An iterative diffusion method shares the background information among similar regions based upon similarity calculations and neighbouring dependencies between adjacent segments. A region-based method for foreground detection and shadow removal was employed by Izadi et al in [81] in video sequences. The approach is based on processing two foregrounds resulting from colour and gradient based background subtraction techniques. The method is constructed on two assumptions. Moving subjects occupy almost up to three quarter of the scene and the inclusion of a background initialisation within the system. Background subtraction is performed in the colour domain with inspiration taken from the GMM presented in [83] by Zivkovic and van der Heijden. The modified version was

originally proposed by Stauffer et al in [70]. A single Gaussian distribution is applied to model each pixel's gradient value in the gradient based background subtraction process. A Sobel kernel is used to filter the gray scale image at each frame. The Gaussian model is updated using each pixel's gradient magnitude. Two binary maps, one each from colour based and gradient based background subtraction techniques are extracted. A median filter is used to reduce noise. A circular structuring element of 3-pixel diameter is used as descriptive close filtering to smooth outer edges and fill the gaps. The final image is generated by adding all non shadow regions to the enhanced image. The system performs significantly well for situations including non-stationary background, camouflage and shadows in video sequences.

A compressed form of background model for a long image sequence was represented by [82] Kim et al through sampling background values at each pixel and quantising them into codebooks. The model also has the capability to cope with local and global illumination changes. Code words contain clustered samples at each pixel based on brightness bounds and colour distortion metrics. The background is encoded on a pixel-by-pixel basis. For a pixel to be part of the background it has to meet the criteria of meeting brightness range of the codeword and colour distortion has to be less than the detection threshold. A colour model is designed to compute the evaluation of colour distortion and brightness distortion. The pixel is considered as foreground if matching codeword is not found. The algorithm is further improved by layered modelling/detection and adaptive codebook updating. The adaptive layered modelling contributes towards the acquisition of changes to the background. Variance with suitable learning rates and exponential smoothing of the codeword helps dealing with illumination changes. The system based on the codebook algorithm shows good characteristics on several background modelling problems.

[87] Ko et al developed a background subtraction and modelling approach through the analysis of temporal variation in colour distributions/intensity. The conventional approach

follows spatio-temporal variations of regional and point statistics in isolation. The approach acts as a hybrid between texture and pixel-wise comparison. For each pixel location including spatial and temporal proximity values, a background model is built. The distribution is computed and contains feature values of the texture of the location. It is used in classification to compare it with the background model. The background model for a pixel is represented by a non-parametric density estimate. Each pixel is treated independently in the distribution. The distance between foreground model distribution and the corresponding background model distribution for that location is calculated to distinguish labels for background and foreground using a threshold value. The background is updated over time to accommodate any changes. The spatial consistency plays an important role in dramatically improved performance of the system. The algorithm is suggested to perform appropriately with large consistent changes in background. The approach also deals well with foreground subjects that have similarity in appearance to the background. Moreover, the system caters for subjects that do not dominate the scene with sudden motion and also for stationary periods. The evaluation of numerous background subtraction approaches is performed in [88] by Parks et al and several post-processing techniques are proposed to improve the performance of foreground masks that result from foreground detection. Noise filtering is suggested to be applied to the foreground mask to remove erroneous blobs. The blob processing is examined to improve the blobs detected in the foreground mark through morphological closing to fill internal holes and small gaps and area thresholding to remove blobs that are insufficiently large to be of interest. The problem of ghosting, the incorrect detection of foreground when subject restarts its motion after stopping its earlier motion and eventually becoming part of the background scene, is addressed to an extent by reducing blobs with nearest zero value of the average optical flow as ghost blobs are motionless. Conditional background updating

using the foreground mask is performed to enhance the detection and to reduce the background model becoming polluted with foreground pixel information.

3.2 Discussion

An efficient and accurate background subtraction technique remains crucial in determining the performance of the applications. The discussed methods earlier present a wide range with respect to the category of datasets used as well as applications. GMM based approaches have turned out to be one of the most reliable techniques to overcome and deal with noisy datasets. GMM based methods present decent robustness to video sequences containing significant background motion.

3.3 Implementation of Background Subtraction

Several surveillance related applications in computer vision require background subtraction for foreground segmentation for post processing at higher level abstraction. The process of segmentation can be constructed initially using simplistic solutions such as using static background. Many existing background subtraction techniques have advantages and disadvantages when it comes to implementation in terms of computational load and performance. The implementations published by Seth Benton in [128] have been applied to discuss the efficiency and quality of the results.

3.3.1 Frame Differencing

The simplest way of implementing background subtraction in terms of computational load etc is using the frame differencing technique. As the name suggests, it involves

subtraction of frames. The observed frame (current frame) is subtracted from the previous possible adjacent frame. A threshold value is used to compare difference in pixel intensities yielding a foreground mask. If the observed pixel value is greater than the threshold value then the observed pixel is assigned to the background. Background $B_{(x,y,t)}$ is estimated to be the previous frame. The following equation represents the idea of frame differencing using threshold value in its most crude form;

$$B_{(x,y,t)} = F_{(x,y,t-1)} \tag{3.2}$$

$$\left|F_{(x,y,t)} - F_{(x,y,t-1)}\right| > T_s \tag{3.3}$$

Where $F_{(x,y,t)}$ and $F_{(x,y,t-1)}$ represent current and previous frames respectively. Threshold value is represented by T_s.

Assigning a suitable threshold value remains an open question in not only frame differencing technique but in all other approaches too. The threshold value can be assigned analytically. It can be assigned keeping in mind the noise in the video sequence to an extent. It can be changed with respect to the quality of the results.

This approach is very easy to implement with the greater advantage of modest computational load. There are no complex calculations involved in order to implement this method. The background model cannot be expected to stay the same for long periods of time. This technique carries another value added advantage of adaptability. The background is adaptable when it comes to changes such as illumination changes etc. As the background is constructed using previous frame therefore changes are adapted rapidly.

The uniformly distributed pixel values tend to become part of the background, thus affecting the quality of the results greatly. As this approach solely relies on frame differencing and relies on subject's movement in terms of difference in pixel values therefore if the subject pauses for a period of one frame, it becomes part of the background.

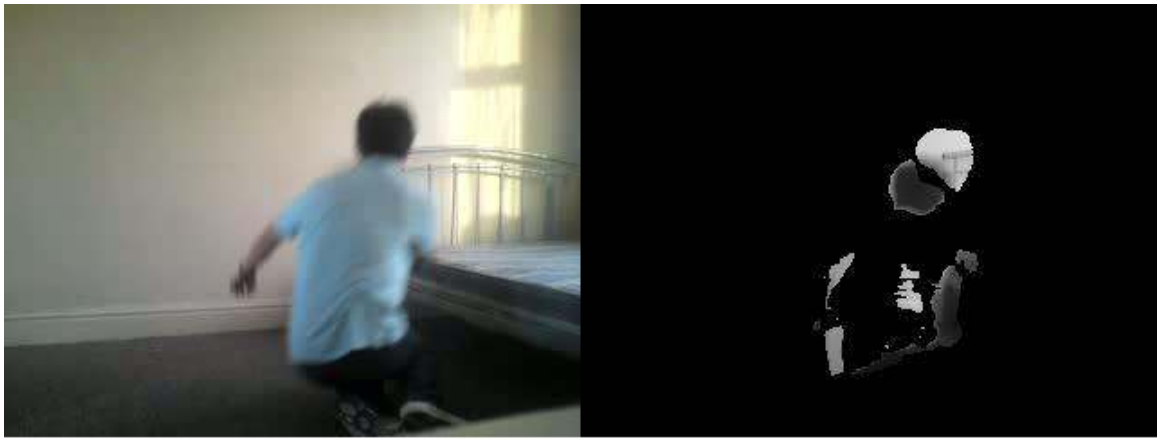

Figure 3.1 Background Subtracted Image using Frame Differencing

Figure 3.1 shows the result of background subtraction applied to one of image sequence from our fall dataset. The output doesn't have a very clear foreground mask. We used threshold value of 50 to avoid noise.

3.3.2 Mean Filtering

Mean filtering can be applied to improve the results obtained from the simple frame differencing method. Background $B_{(x,y,t)}$ is considered as the mean of previous *n* frames. The estimated background model is represented as;

$$B_{(x,y,t)} = \frac{1}{n}\sum_{i=0}^{n-1} F_{(x,y,t-i)} \tag{3.4}$$

$$\left|F_{(x,y,t)} - \frac{1}{n}\sum_{i=0}^{n-1} F_{(x,y,t-i)}\right| > T_s \tag{3.5}$$

Mean filtering has relatively higher memory requirements which also adds factor of time on computational load. A running average can resolve the issue of memory requirement to an extent. The mean background estimate can be represented by;

$$B_{(x,y,t)} = \frac{t-1}{t} B_{(x,y,t-1)} + \frac{1}{t} F_{(x,y,t)} \tag{3.6}$$

3.3.3 Approximate Median Filtering

The idea of median filtering is constructed by buffering the previous frames of the video sequence. The median of the buffered frames is computed that yields the background estimate. Then frame differencing described earlier is applied to obtain background subtraction. Therefore approximate median filtering refers the median of previous frames in a video sequence to establish a statistical background model for background subtraction. Median of the previous *n* frames is used with the assumption of background reappearing in the video sequence. The background estimate can be represented as;

$$B_{(x,y,t)} = Median\{F_{(x,y,t-i)}\} \tag{3.7}$$

$$\left|F_{(x,y,t)} - Median\{F_{(x,y,t-i)}\}\right| > T_s \tag{3.8}$$

$$i \in \{0, \ldots, n-1\}$$

The computational load can further be improved by applying the method of the approximate median filtering. The background pixel value is incremented by 1 if the observed pixel value in the current frame is greater than the corresponding background pixel. Similarly The background pixel value is decremented by 1 if the observed pixel value in the current frame is smaller than the corresponding background pixel. The approximate median estimate is achieved eventually when half the pixel values are greater than the background and half are less than the background. The time of convergence depends on frame rate and movement in the video sequence.

The median procedure is very memory and time consuming operation due to memory requirements associated with this procedure which limits its applications. The performance exhibited by the median filter is comparable to complex approaches. The approach shows adequate levels of robustness.

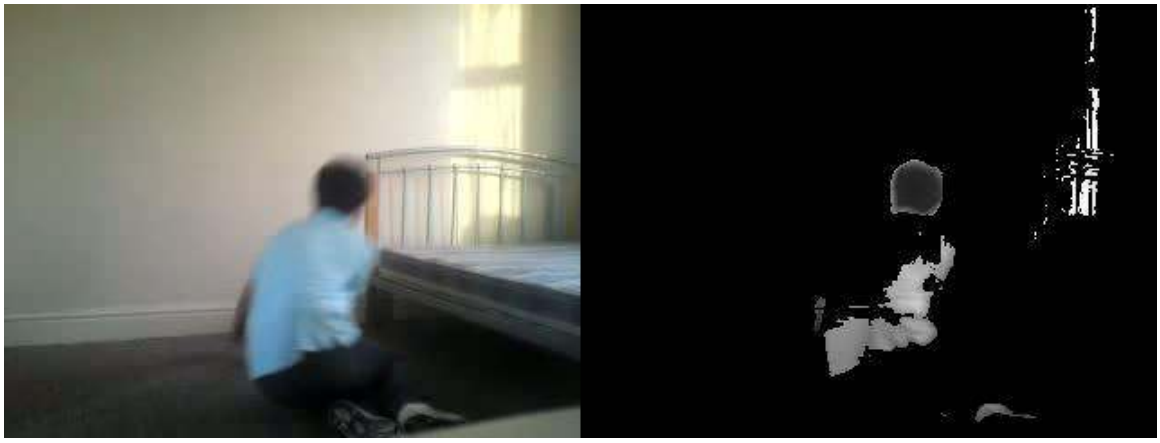

Figure 3.2 Background Subtracted Image using Approx. Median Filtering

Figure 3.2 shows the result of background subtraction applied to one of image sequence from our fall dataset using approximate median filtering. The output doesn't have a very clear foreground mask and also include the noise due to the illumination changes. We used threshold value of 50 to reduce the effect of noise.

3.3.4 Mixture of Gaussians

Mixture of Gaussians (MoG) is amongst the high complexity approaches. The approach described by Stauffer et al in [70] presented an adaptive background mixture model by modelling each pixel as a mixture of Gaussians and using an on-line approximation to update the model for tracking. The background model is parametric. An online model containing mixture of adaptive Gaussians is constructed to approximate this process.

3.3.4.1 Online Mixture Model

MoG functions represent every pixel's location. A Probability distribution function for the observed current pixel is given by;

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} * \eta(X_t, \mu_{i,t}, \sum_{i,t}) \quad (3.9)$$

Where *K* is the number of distributions and its value depends on the computational capacity and memory constraints. Usually 3 to 5 is used. $X_t$is the current pixel's value, $\omega_{i,t}$ is an estimate of the weight (portion of the data accounted for by this Gaussian) of the *ith* Gaussian in the mixture at time t, $\mu_{i,t}$ is the mean value of the *ith* Gaussian in the mixture at time t, $\sum_{i,t}$ is the covariance matrix of the *ith* Gaussian in the mixture at time t and η is a Gaussian probability density function.

$$\eta(X_t, \mu_{i,t}, \sum_{i,t}) = \frac{1}{(2\pi)^{\frac{n}{2}}(|\sum|)^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t-\mu_t)^T\Sigma^{-1}(X_t-\mu_t)} \quad (3.10)$$

3.3.4.2 Model Adaptation

Gaussians are updated using on-line K-means approximation. To find a match, all new pixel values are compared with K Gaussian distributions. A match is assumed to be a pixel value within 2.5 standard deviations (2.5σ) of a distribution. The μ and σ parameters for matched distributions are updated as follows;

$$\mu_{i,t+1} = (1-\rho)\mu_{i,t} + \rho X_{t+1} \quad (3.11)$$

&

$$\sigma^2_{i,t+1} = (1-\rho)\sigma^2_{i,t} + \rho(X_{t+1} - \mu_{i,t+1})^2 \quad (3.12)$$

Where

$$\rho = \alpha\, \eta(X_t|\mu_{i,t}, \sigma_k) \quad (3.13)$$

and α is learning rate. The prior weights of all distributions are adjusted as follows;

$$\omega_{i,t+1} = (1-\alpha)\omega_{i,t} + \alpha(M_{i,t+1}) \quad (3.14)$$

where $M_{i,t+1}$ = 1 for the model for the matching Gaussian and $M_{i,t+1}$= 0 for non matching. If no match is found then the current value replaces the least probable distribution as its mean value, a low prior weight and initial high variance.

3.3.4.3 Background Model Estimation

Heuristically, the Gaussian distributions which have the most supporting evidence and the least variance correspond to the background. The value of ω/σ orders the Gaussians. The value is increased with increased evidence and decreased variance. The background model is constructed on choosing first *B* distributions.

$$B = argmin_b(\sum_{i=1}^{b}\omega_i > T) \quad (3.15)$$

T above represents a minimum region of expected background.

This approach requires intelligent parameter optimisation as well as initialisation of the Gaussians as it cannot deal with rapid illumination changes. All parameters have a significant impact on the performance of the method. The model provides fast recovery when background reappears and has an automatic pixel-wise threshold.

Figure 3.3 Background Subtracted Image using Mixture of Gaussians

Figure 3.3 shows the result of background subtraction using mixture of Gaussians applied to one of image sequence from our fall dataset. The output has considerable noise which affects the quality of the overall result. The connected points in the binary foreground mask gives good outer boundary of the subject but still missed out most of the information. We used threshold index of 0.25 instead of threshold value as per the original settings of the implementation.

3.3.5 Graph Cut

The graph based approach by Howe et al in [129] is implemented to construct a model of the static background. The model can be updated either dynamically frame by frame in a video sequence or offline. The next frame is compared on a per pixel basis with the background model.

The graph is constructed based on the image. A corresponding graph vertex $V_{i,j}$ is created by each pixel in the image. Foreground and background are represented by two additional vertices as source and sink respectively. Six nodes: source, sink and vertices of four connected neighbours in the graph are connected by a typical vertex. Fewer neighbour links are formed for vertices corresponding to the pixels on the image edge. All pixel vertices are connected to source and sink. The weight of the links between the pixel vertices, the source *s* and sink *t* is derived directly from the difference between the current frame and the background at the corresponding pixel $\delta_{i,j}$:

$$\omega(s, p_{i,j}) = \delta_{i,j} \tag{3.16}$$

$$\omega(p_{i,j}, t) = 2\tau - \delta_{i,j} \tag{3.17}$$

All neighbour links between pixel vertices have similar weights. The weights are equivalent to τ times α. The value of α is typically close to 1.0. τ acts parallel to the threshold value in the morphological method. It corresponds to the level where the pixel association to the foreground is greater than the background. Grouping of neighbouring pixels are administered by the value of α. The lower value of α shows weak pixel bonding similar to the result obtained using just threshold for background subtraction.

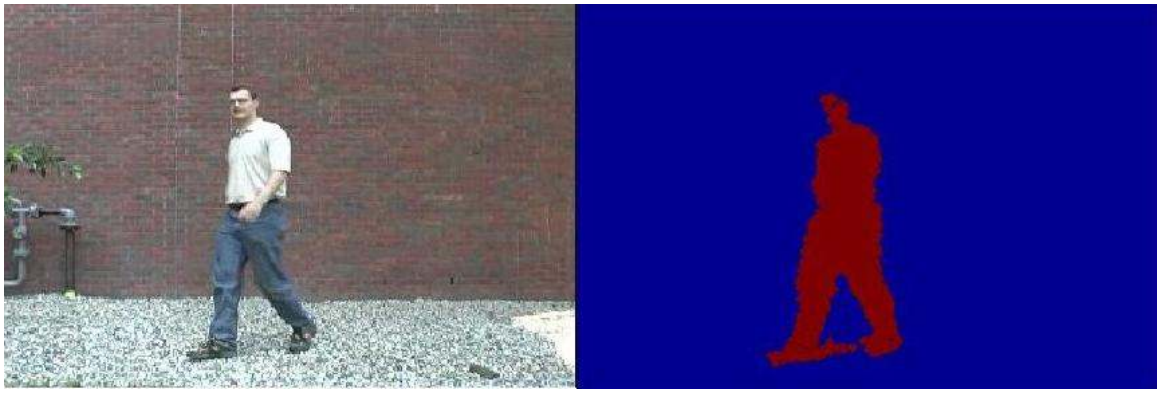

Figure 3.4 Background Subtracted Image using Graph Cut

Figure 3.5 Background Subtracted Image using Graph Cut

Figure 3.4 and 3.5 shows the result of background subtraction using graph cut applied to two of the image sequences. The output has no noise in 3.4 but very visible noise content in 3.5 that affects the quality of the overall result. The connected points in the binary image provide very good quality of the foreground. We used threshold value of 10 as per the original settings of the implementation to yield the best quality of foreground mask and minimum noise. Figure 3.5 clearly proves that the image sequence needs to have suitable lighting conditions for the graph cut to provide good quality foreground mask.

3.4 Conclusion

Graph cut based approach for foreground extraction and background subtraction has exhibited results with reduced noise and acceptable accuracy in comparison with

morphological procedures based techniques. The graph based approach tends to deal with noise reduction by overcoming its effects adequately through information aggregation from a local neighbourhood around every pixel and remaining true to the underlying data. The efficiency of the method points out negatively towards its implementation in comparison with morphological procedures as empirically it runs at lower resolution and lower speed.

The use of background subtraction in the range of applications strongly suggests the likelihood of adoption of the new techniques and will definitely supply significant advantages in a number of fields. The use of graph cut in addition to other modern approaches will surely open the door for new applications. Therefore graph cut draws decent amount of attention from scientific community for further advancement especially in the area of foreground segmentation.

Chapter 4

Optical Flow

In this chapter the Optical Flow methodology is reviewed and our implementation of Optical Flow Estimation is demonstrated. We have applied the optical flow computation to test its accuracy in one of the real world applications as fall detection. We have also demonstrated an evaluation of fall detection using optical flow estimation for the computation of motion involved in the action and represented it as histogram of optical flow (HOF) as feature descriptor. We have computed the accuracy of out HOF descriptor and compared it against the other feature descriptors.

One of the fundamental properties of any video scene is its motion. The most often used representation of image motion is the optical flow field. The estimation of optical flow still remains one of the very important research areas in computer vision. Optical flow is indeed one of the very powerful tools to interpret motion analysis. Estimation of pixel motion in two consecutive frames yields the optical flow computation. A significant amount of information can be obtained from the apparent pixel motion such as brightness pattern due to a relative motion between a camera and the subjects. Several advanced approaches have been proposed starting from the original work of Horn and Schunck [124] as well as Lucas and Kanade [125]. Robust statistics have extended both approaches allowing the treatment of outliers due to occlusions or motion discontinuities, through the matching or smoothing. Advancement in the area of the optical flow research has not only improved the quality of optical flow estimation approaches but also provided an insight into how several techniques can work together through greater understanding of their functional implementation and component/parameter analysis.

4.1 Related Work

Lucas and Kanade's [125] gradient-based method is among the most accurate and computationally efficient methods for optical flow estimation. [91] Brox et al proposed a novel variational approach containing an energy function based upon three assumptions: gradient constancy, brightness constancy and spatio-temporal smoothness constraint to preserve discontinuity. The approach is minimised with a numerical methods and is an integration of earlier published ideas. The use of warping is also known as coarse-to-fine technique. This idea is used to compute the non-linearised optical flow well in image registration. The image correspondence issues are solved through a multi-resolution method. To measure the displacement vector, some small variations are allowed in the grey value through an invariant criteria called gradient of the image grey value. Multi-scale ideas are applied to cater larger displacements. Using these assumptions and ideas, an energy function is derived that penalises deviations from these model assumptions. A piecewise smooth flow field generalises the smoothness assumption. Numerical methods including Euler-Lagrange equations and Taylor expansions are applied based on two nested fixed point iterations. The coarse-to-fine technique as single minimisation problem is developed. Energy functions are minimised in image registration through multi-resolutions. The angular errors are significantly small for optical flow estimation due to parameter variations insensitivity. [92] Lim et al described a method based on Lucas-Kanade's gradient based method to achieve estimates of optical flow at a standard frame rate using high frame rate sequences. The optical flow estimates at standard rate are computed after estimates are further processed, aggregated and filtered. The optical flow estimates between two consecutive high speed frames is computed using the Lucas-Kanade method. A final estimate of optical flow between two consecutive standard frame rate images is built using this estimate. The error accumulation is

defended through warping and refining. Computational complications are reduced by warping gradients instead of the frame. A higher sampling frame rate is used to compute the small displacements between intermediate frames in the case of large displacements. Similarly in the case of small displacements, lower sampling frame rate is used to increase the efficiency. An approach of recovering dense optical flow field map from two images was proposed by Alvarez et al in [93]. The approach takes into account the symmetry across the images, possible occlusions and discontinuities in the flow field. The energy function is computed. The Nagel-Enkelmann operator [94] is applied to counteract smoothness of the flow map across boundaries of the images. The system is embedded into multi-resolution framework. The symmetry across the two images is reintroduced by simultaneous calculation of the flow from image1 to image2 and image 2 to image1. Partial differential equations are used to add explicit terms that deny the compatibility of the two optical flow computations. The system is capable of detecting large occlusions to ensure compatibility is achieved most of the time. Brox in [95] improved the classic optic flow estimation techniques of Lucas and Kanade and of Bigun et al through a diffusion method to build a nonlinear structure tensor (ST). The conventional benefits of the linear ST are kept and additionally its problem of subject delocalization is also tackled. The linear ST is derived by smoothing each component through a Gaussian kernel with standard deviation. This approach appropriately reduces the noise. Diffusivity is kept at its maximum except at locations where discontinuities in the magnitude exist. The image gradient drives the diffusion. The new structure tensor adapts the diffusivity and constitutes a nonlinear diffusion process. The nonlinear ST improves the results of other methods where a linear ST is used. Nir et al [96] used a spatio-temporal model with varying coefficients multiplying a set of base functions at each pixel to introduce an estimation of optical flow. A general space-time model represents the flow field. The piecewise constant coefficients describe the true optical flow. The coefficient variations don't

induce most of the local spatio-temporal changes in the flow rather the variations are mostly affected by the changes in the basis functions, thus making the coefficient regularisation more meaningful. The proposed smoothness term carries a penalty for spatio-temporal variations in the coefficient functions. The suggested over parameterisation models produce effective performance of optical flow recovery through the smoothness penalty term.

An analysis containing the spatial and temporal statistics of natural optical flow fields is performed in [97] by Roth et al. A system is devised to exploit the spatial statistics computed. A database of realistic optical flow fields from natural and manmade scenes (3D camera motions recovered from hand-held and car-mounted video sequences) is constructed. The difference in image coordinates under which a scene is viewed contribute towards its optical flow. Various characteristics are learnt by histogram patterns of statistics analysis and such properties are attributed to each type of motion i.e. translational camera motion and rotational camera motion. The spatial statistics of optical flow are computed in overlapping patches and using the approach of Fields-of-Experts (FoE) approach in [98] by Roth et al which applies a Markov random field to model the prior probability of optical flow fields. Dense optical flow computations are performed when the new optical flow prior is integrated into a recent optical flow method by Bruhn et al [99] and the results are compared quantitatively. A comparison between new optical flow prior and previous robust priors is performed. The quantitative improvement of flow accuracy is demonstrated through the new optical flow prior. Efros et al [100] performed action recognition using a motion descriptor based on optical flow estimates in a figure centric spatio-temporal volume of stabilised human subjects and combined it with an associated measure of similarity. The Lucas-Kanade method is applied to compute the optical flow. The optical flow is processed as spatial pattern consisting of noisy estimates. The spatio-temporal motion descriptor is formulated by aggregating and smoothing noisy measurements of the optical flow. Optical flow was

employed by Brox et al [101] in order to compute additional point correspondences for 3D pose tracking. The supplementing of optical flow improves pose tracking. It also increases robustness of pose estimation by producing additional correspondences (2D and 3D) and resolves ambiguous situations. It also produces an improved initial pose estimate through motion estimation by means of the optical flow correspondences that can cater for large displacements. The approach deals smartly with textured, homogeneous, cluttered, blurring subjects or noise artifacts. Bruhn et al in [102], an extension of their earlier work in [99], combined global Horn and Schunck's approach with the local Lucas–Kanade method to introduce the combined local-global (CLG) system for optic flow estimation. The CLG system uses the minimisation of the energy functional to compute the optic flow field. The system combines the dense flow fields of Horn–Schunck with the high noise robustness of Lucas–Kanade. Euler–Lagrange equations are implemented to minimise the energy function for the recovery of the optical flow field. Higher frequencies on coarser grids are the representation of lower frequencies on the finest grid, where they can be amply reduced. The system, through a multi-grid approach, delivers accurate results much faster. The optical flow estimation was regularised by Xiao et al in [103] through a flexible multi-cue driven adaptive bilateral filter. Even with highly desirable motion discontinuities the filter is able to compute the smoothly varying optical flow field. A two step, filter based updating model is applied on the conventional one step variational updating model. The occlusion detector detects the occlusion area by exploiting the natural property of the occlusion between two frames. The data and occlusion energy is balanced by the occlusion term in the variational framework. The multi-cue driven bilinear filter containing occlusion function and a one dimensional Gaussian substitutes the conventional anisotropic diffusion tensor in the variational framework. The multi-cue driven bilinear filter disables the influence of the occluded region completely during the diffusion process and also reduces the influence based on motion

dissimilarity. The system produces accurate optical flow field and performs effective occlusion handling.

McCane et al generated motion fields in [104] from polyhedral subjects using real scenes. A test data set is proposed to benchmark the optical flow methods containing complex synthetic sequences and natural sequences with ground truth. The ground truth of synthesised motion fields is obtained through tracing the ray of two adjacent frames. The method used is proposed by Mason et al in [105]. The 3D location of the intersection, the corresponding pixel location and the subject is recorded when a ray collides with an subject. Each subject and each subject-ray intersection is applied with a transformation. The new pixel location is yielded geometrically. The ground truth motion vector is computed through subtraction. The quantitative evaluations of the synthetic and natural sequences achieved are relatively consistent. Dalal et al developed a human/person detector in [106] combining motion/appearance descriptors with histograms of oriented differential optical flow to detect stationary and moving humans. Several motion coding methods are applied on data sets containing video sequences of human subjects with possible background and camera motion. The Histogram of Oriented Gradients (HOG) descriptor [107] is used to compute the visual appearance thus covering the stationary part of the person detector. Motion boundary coding is performed through Motion Boundary Histograms (MBH). The local orientations of motion Edges are captured by imitating the stationary image HOG descriptors. The calculation of relative limbs' movement is performed using Internal Motion Histogram (IMH) descriptor.. A spatiotemporal differencing scheme is evaluated to compute the image intensity. Optical flow is implemented using a simple constant brightness assumption. A robust descriptor is obtained through oriented histogram voting by motion and appearance channels. The overall performance is improved when the features are used in combination with static appearance descriptors. The analysis of convergence results for the Horn and Schunck optical-flow

estimation method are presented by Mitiche et al [108]. The exercise of linear system of equations ordered to make definite symmetric positive matrix is performed explicitly. The exercise results in convergence of the iterative point-wise and block-wise Gauss–Seidel and relaxation approaches. The correspondence of block Jacobi, Gauss–Seidel, and relaxations approaches to the tri-diagonal block decomposition of the matrix also shows the convergence. Wedel et al in [109] proposed an improvement variant of Zach et al's [110] method for the original duality based TV-L^1 optical flow. The total variation (TV) regularization is employed to preserve discontinuities in the flow field. The intensity value artifacts due to illumination changes are modelled through a structure-texture decomposition approach. The approach is based on an assumption of shadows appearing only in structural part, which includes the main large subjects. The total variation based model of Rudin et al in [111] for image de-noising is used to accomplish the structure-texture decomposition. An image is observed as structural and textual content. It correlates to main large subjects and fine scale details. The convergence to unfavourable local minima is avoided by employing a coarse-to-fine approach embedded with the energy minimisation method. The improvements increase the accuracy of the optical flow estimation considerably. Brox et al in [112] & [119] combined the advantages of energy minimization methods and descriptor matching strategies to perform optical flow estimation with large displacements. The hierarchical segmentation of the image is used to extract regions as a better image coverage is rendered by the hierarchical segmentation. Regions correspond highly with separately moving structures than conventionally used blobs or corners. A sparse set of hypotheses for correspondences on regions is obtained through descriptor matching. In order to compute a dense accurate flow field, these correspondences from sparse descriptor matching are integrated into a variational scheme to influence the variational approach. Local information from the raw image data and a smoothness prior is combined with it to avoid the local optimization to get stuck in a local

minimum underestimating the true flow. The local optimization to the correct large displacement is achieved through global nearest neighbour matching. Exploiting image information compared to the interpolated point correspondences yields an accurate flow field. Line and optical flow histograms were used in [113] by Ikizler et al to present a concise illustration for human action recognition. The dense optical flow representation and global temporal information together with shape descriptors based on the distribution of lines is employed. Canny edges are implemented to calculate the probability of boundaries. The densest region of high response features is identified to localise human figures and the Hough transform is applied to fit straight lines into the boundaries. Orientations and spatial locations are processed with histogram to put together spatial information of the human body. A dense block based optical flow is extracted through previous frame matching. Orientation histograms of these optical flow estimates are generated. The performance of the algorithm is further enhanced by a simple feature i.e. the overall velocity of the subject in motion. The algorithm reduces classification time substantially.

Sun et al in [114] presented a complete probabilistic model of optical flow. The model encompasses spatial statistics, brightness inconstancy statistics and the relationship between flow boundaries and the image intensity structure of the flow field. The steered derivatives of optical flow are modelled. First model of spatial smoothness is achieved on the assumption of flow fields being independent of the reference image. The Steerable Random Field model [115] is generalised to a steerable model of optical flow by capturing the oriented smoothness of the flow fields. Before training the model, a Gaussian scale mixture (GSM) [116] representation is employed for empirical determination of scales and global variance. A further generalization of the constancy assumptions (gradient constancy and brightness constancy) is proposed. The constancy of responses to several general linear filters (derivative filters and Gaussian smoothing filter) is modelled. The generalized data term with

the steered model of flow improves performance substantially. Regularisation is applied in the images in [117] to compute scene flow by Vedula et al. The motion at every point in the scene is represented by a 3D flow field. Scene flow computation is based on assumption of optical flow calculation for each camera initially. Three different ways are discussed to combine optical flows i.e. known scene geometry from single camera, known scene geometry from multiple cameras and unknown scene geometry from multiple cameras. Scene flow from single camera is mainly computed for theoretical interest. Unknown scene geometry from multiple cameras helps constraining scene structure from the inconsistencies in multiple optical flows. In the known scene geometry from multiple cameras, the geometry of the scene is contained in the volume and is segmented into voxels. The voxel coloring algorithm [118] is applied to calculate the occupancy of each voxel. Various images are projected by each voxel. The occupancy of each voxel also depends on the consistency of measurements from the various images. In the unknown scene geometry from multiple cameras, the measure of likelihood for each voxel is calculated. The co-planarity measure is not greatly influenced by outliers therefore the visibility is ignored. The method can be further developed by arranging camera to compute efficient scene flow. In [120] Chaudhry et al used a histogram of oriented optical flow (HOOF) to represent each frame of a video and performed classification of human action recognition through HOOF time-series. HOOF features are non-Euclidean. The HOOF features are independent of the scale and the direction of motion. System parameters are learnt using kernels on the original histograms. The temporal evolution of HOOF features is modelled using Non-Linear Dynamical systems (NLDS). The Binet-Cauchy trace kernel for NLDS is computed as "*the expected value of an infinite series of weighted inner products between the outputs after embedding them into the high-dimensional (possibly infinite) space*" [120]. Different actions are classified by exploiting the temporal evolution of these histograms. This is based on an assumption of each action inducing a time-series of HOOF

with specific dynamics and different action inducing different dynamics. The comparison on the dynamics of HOOF time series is performed for action recognition. The application of HOOF features framework yields significant results. Optical flow was applied in [121] as a feature to extract information by Andrade et al from the crowd video data to build an event detector for emergencies in crowds. The adaptive mixture of Gaussians is implemented to model the background. A Gaussian spatio-temporal filter is implemented to reduce noise. The resulting mask is then used with the optical flow estimation. The robust dense optical flow algorithm by Black et al [122] is implemented to compute optical flow estimates. Smooth optical flow estimates are achieved at the motion boundaries. A median filter is applied for further noise reduction and to reduce flow vectors in the model. This reduced observation noise and only flow vectors inside foreground subjects are analysed. The optical flow features are encoded with Hidden Markov Models to allow for the detection of emergency or abnormal events in the crowd. Díaz et al presented a hardware implementation of the optical flow estimation in [123] through a pipelined optical flow processing system based on a field-programmable gate array (FPGA) device. It allows change of configuring parameters to adapt the sensor to illumination conditions. The design of circuit included a customised digital signal processing system in a single chip of high computational power. It is built on pipeline resources and comprehensive intrinsic parallelism. Frames are received through the camera and stored in memory banks. Smoothing is applied through Spatial Gaussian filter. Temporal derivation and spatio-temporal smoothing of images is performed via the infinite-impulse-response (IIR) temporal filter. After the calculation of spatial derivatives least square matrices are computed for the estimation of integration of neighbourhood velocities. Finally velocity estimation is performed through arithmetic operations using a customized floating-point unit. The computation bit width increases throughout the pipeline structure. The

computing scheme can be further altered to improve the design as the modularity of the system enables it do so.

4.2 Discussion

Most of the currently implemented approaches have strong resemblance with Horn and Schunck's original method of optical flow estimation. The original algorithms of Horn and Schunck and Lucas-Kannade are both very competitive with respect to adequate implementation. The introduction of coarse-to-fine estimation solves the issues of filter constancy of high order as larger displacements. Illumination changes have been reduced to an acceptable extent. Other added functionalities include bi-cubic interpolation based warping, temporal averaging of image derivatives and graduated non-convexity to minimize non-convex energies.

4.3 Implementation of Optical Flow Estimation

A wide variety of approaches have been presented to cater for the extraction of motion analysis from image sequences in the field of computer vision. Optical flow is the illusion of motion generated by the brightness patterns within an image scene. Relative motion between the subject and the observer produces optical flow. The optical flow can be computed using various approaches. The velocity vector fields represent the optical flow once it is computed. The image intensity conservation based approaches have attracted greater attention in order to enhance computational efficiency.

Several algorithms dealt with the issue of the inherent aperture using the idea of brightness constancy assumption and applied additional suppositions to build an optical flow estimate. Lucas and Kanade in [125] applied a constant motion model for parameter solution

over image regions to deal the with the aperture problem. The functional minimisation using mathematical tools from calculus was first employed by Horn and Schunk in [124] to solve optical flow problems. The idea of solving dense global optical flow fields was introduced by their pioneering work. They presented a quality function with two terms: deviations from the brightness constancy equation is penalised by a data term and variations in the flow field is penalised by the smoothness term. Many approaches with improvements have been presented using their work.

The optical flow computation via several methods over a video sequence in computer vision has been widely used in various applications especially in surveillance related implementations and motion detection etc. Several algorithmic implementations need to be efficient enough to be used in real time applications. Typical optical flow approaches are less efficient when it comes to implementing them as there is a significant amount of computation and memory use as well as time. Despite much research effort invested in addressing optical flow computation it remains a challenging task in the field of computer vision. The following implementations of few of the most common optical flow approaches have been applied to discuss the efficiency and accuracy including the approach with the use of median filtering.

4.3.1 Optical Flow Computation

The typical methods for optical flow computation use images at time t and $t + \Delta t$ through Taylor series. Spatial and temporal derivatives are employed. The intensity of each of the voxel in a given image I is represented as:

$$\boldsymbol{I(x, y, t)} \tag{4.1}$$

Using the assumption of constant image intensity of each visible scene point over time, we can deduce the following:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \tag{4.2}$$

We can develop the Taylor series based on assumption of movement to be small;

$$I(x+\delta x, y+\delta y, t+\delta t) = I(x,y,t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + H.O.T. \quad (4.3)$$

Where higher order terms are represented by one term as *H.O.T*. *H.O.T* can be ignored:

$$\frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = 0 \quad (4.4)$$

Dividing by δt:

$$\frac{\partial I}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial I}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial I}{\partial t}\frac{\delta t}{\delta t} = 0 \quad (4.5)$$

This results in:

$$\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial t} = 0 \quad (4.6)$$

Velocity components of optical flow associated with the observed voxel are represented by V_x and V_y with respect to X-axis and Y-axis. The derivatives of the image at $I(x,y,t)$ is represented by $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$ and $\frac{\partial I}{\partial t}$. Which can also be represented by I_x, I_y and I_t.

$$I_x \cdot V_x + I_y \cdot V_y = -I_t \quad (4.7)$$

Equivalently

$$\nabla I^T \cdot \vec{V} = -I_t$$

4.3.2 Lucas Kanade Optical Flow Computation

Lucas Kanade's method of optical flow calculation [125] is based on dividing image into patches and computing optical flow estimates for each of them thus making it a local/sparse method. It is essentially a differential two frames algorithm as it requires minimum number two frames to function. The algorithm also presents the hypothesis of uniform velocity for all the pixels in the neighbourhood. Using a *m* x *n* window (m > 1) centered at the pixel *P* we can deduce the following:

$$I_{x_1}V_x + I_{y_1}V_y = -I_{t_1}$$

$$I_{x_2}V_x + I_{y_2}V_y = -I_{t_2}$$

$$\vdots$$

$$I_{x_n}V_x + I_{y_n}V_y = -I_{t_n} \tag{4.8}$$

This can be rewritten in the form of matrices:

$$\begin{bmatrix} I_{x_1} & I_{y_1} \\ I_{x_2} & I_{y_2} \\ \vdots & \vdots \\ I_{x_n} & I_{y_n} \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} -I_{t_1} \\ -I_{t_2} \\ \vdots \\ -I_{t_n} \end{bmatrix} \tag{4.9}$$

Using the least squares methods a possible solution could be given as shown in the following:

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum {I_{x_i}}^2 & \sum I_{x_i} I_{y_i} \\ \sum I_{x_i} I_{y_i} & \sum {I_{y_i}}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum I_{x_i} I_{t_i} \\ -\sum I_{y_i} I_{t_i} \end{bmatrix} \tag{4.10}$$

For a given window the equation 4.7 can be rewritten with the sums running from *i* =1 to n:

$$f(u,v) = \iint (I_x u + I_y v + I_t)^2 \, dx \, dy \tag{4.11}$$

The integration is performed with respect to the dimension of the window. Using the assumption of uniform velocity all over window with non dependence of *u* and *v* on *dx* and *dy*, we can deduce the following system:

$$u \iint {I_x}^2 \, dx \, dy + v \iint I_x I_y \, dx \, dy + \iint I_x I_t \, dx \, dy = 0 \tag{4.12}$$

$$v \iint {I_y}^2 \, dx \, dy + u \iint I_x I_y \, dx \, dy + \iint I_y I_t \, dx \, dy = 0 \tag{4.13}$$

Several parameters need attention in terms of tuning in order to obtain results at adequate levels and to reduce noise.

- **Window Size:** As the algorithm is based on the hypothesis of uniform velocity for all the pixels in a given window of size *m* x *n*. This means the size of the window can be chosen as direct consequence, the density of optical flow vectors.
- **Threshold:** In order to reduce the noise content between two frames the size of the temporal derivative of a pixel can be chosen to distinguish its movement. A temporal derivative of a pixel less than threshold will determine the non-movement of the pixel.

- **Smoothing:** Gaussian smoothing is represented by a matrix with the width of *w* x *w*, where *w* is the neighbourhood window. It is used in order to weight the sums in equation 1.9 such that the components far from the central pixel count less than those that are closer.

$$G[i][j] = -\frac{j-w-1}{2\pi\cdot\sigma^3}\cdot e^{\left(-\frac{(i-k-1)^2+(j-k-1)^2}{2\cdot\sigma^2}\right)} \quad (4.14)$$

Where σ represents the standard deviation in the above Gaussian function. The pixels far from the center can be considered in terms of their importance depending on the value of the σ. A video sequence with large spatial derivatives yields numerical problems of stability issues, as the algorithm deals with very small and very large numbers in the case of large spatial derivatives.

Figure 4.1 Two adjacent frames from fall action sequence

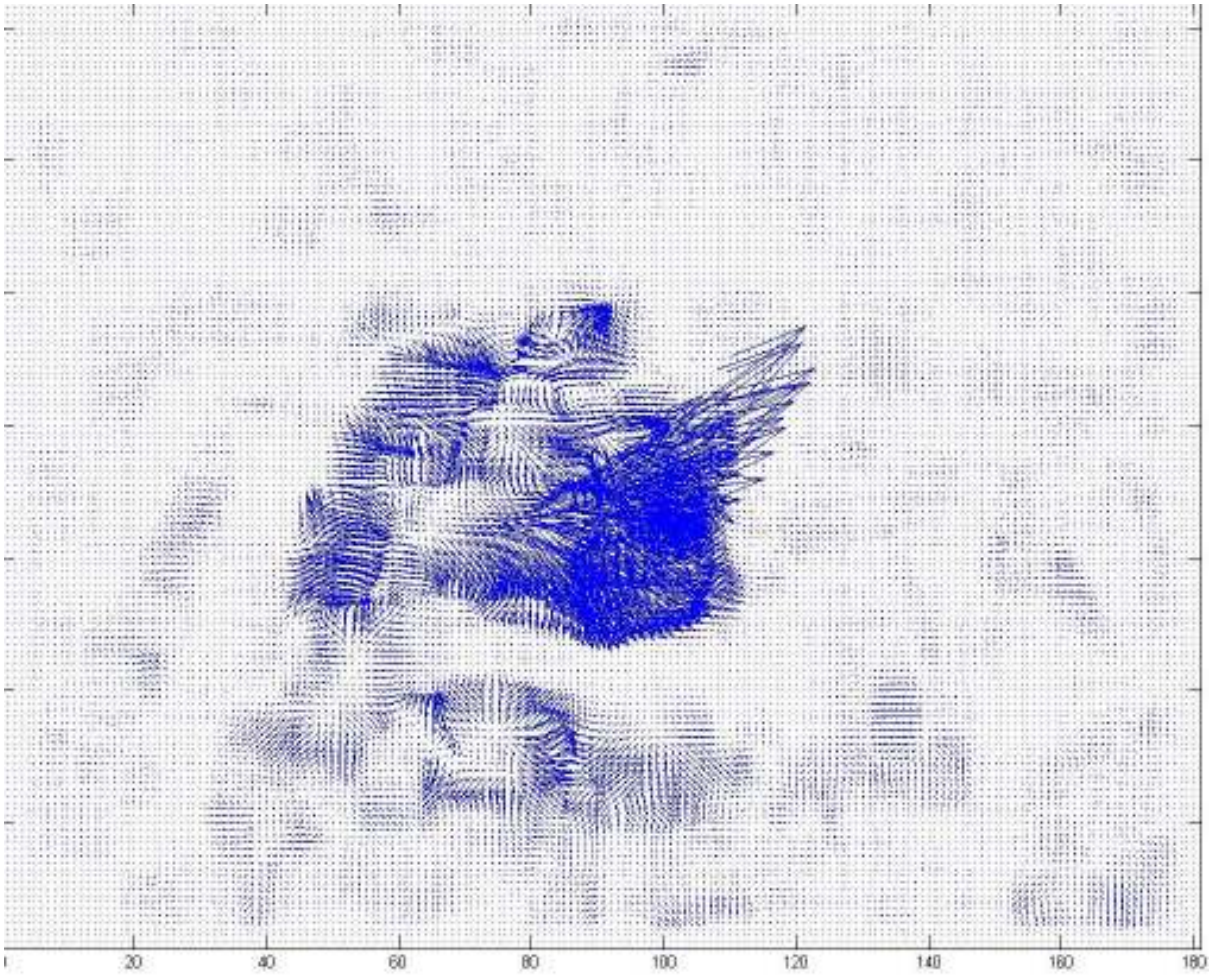

Figure 4.2 Lucas-Kannade Optical Flow of two adjacent frames in Figure 4.1

Figure 4.2 shows the representation of optical flow computation of two consecutive frames shown in figure 4.1. There is apparent noise in the representation, which seems to be part of the motion of the subject as well as the background. The overall optical flow representation requires further post processing to yield a better optical flow representation.

4.3.3 Horn-Schunk Optical Flow Computation

The use of variation methods was initiated by Horn and Schunk [124] for optical flow field computation. This approach presents a global constraint of smoothness over the optical flow field computation. The global energy function is formulated for the optical flow field. The energy function is minimised. The energy function presents two assumptions: constancy

imposed over the grey image by the grey value constancy (this means grey value of the image is not variable over time) and global smoothness of optical flow field. An additional assumption was presented in this approach in the form of a constraint. The variation in the optical flow remains smooth and, therefore, velocity components in the neighbourhood are uniform. The energy function can be presented as;

$$\min_{(u,v)} \int_x \int_y E(u,v)\, dx\, dy \tag{4.15}$$

Where

$$E = E_d + \alpha E_r \tag{4.16}$$

$$E_d(u,v) = (I_x u + I_y v\ + I_t)^2 \tag{4.17}$$

E_d is the data term and E_r is a regularization term. The parameter α is a positive scalar quantity to trade off the goodness of E_d with E_r. This relates to a standard choice of E_r to be the following isotropic regularisation term.

$$E_r(u,v) = \frac{1}{2}(|\nabla u|^2 + |\nabla v|^2) \tag{4.18}$$

Two elliptic partial differential equations can be derived using variation calculus.

$$\alpha \Delta u - I_x\left(I_x u + I_y v\ + I_t\right) = 0 \tag{4.19}$$

$$\alpha \Delta v - I_y\left(I_x u + I_y v\ + I_t\right) = 0 \tag{4.20}$$

The two velocity components *u* and *v* make this coupled system of equations symmetric. The equations can be simultaneously solved to determine the coupling effect. The following Gauss-Seidel relation is suggested in this approach.

$$u^{k+1} = \bar{u}^k - I_x \frac{I_x \bar{u}^k + I_y v^k + I_t}{\alpha + {I_x}^2 + {I_y}^2} \tag{4.21}$$

$$v^{k+1} = \bar{v}^k - I_x \frac{I_x \bar{u}^k + I_y \bar{v}^k + I_t}{\alpha + {I_x}^2 + {I_y}^2} \tag{4.22}$$

Where *k* is the iteration counter, $\bar{u}$ and $\bar{v}$ are average of the neighbouring points to u and v.

Figure 4.3 Two frames from fall action sequence

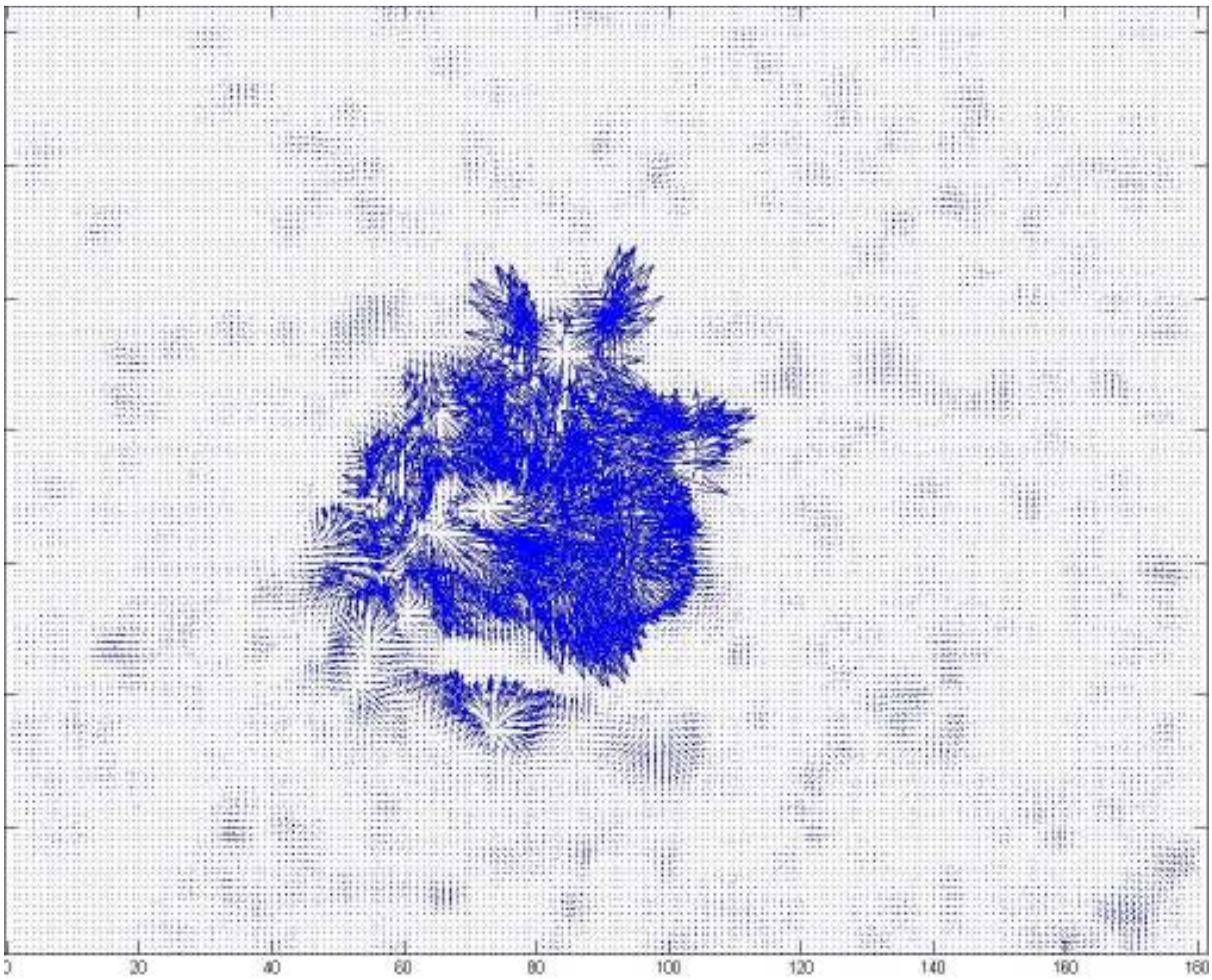

Figure 4.4 Horn-Schunk Optical Flow of two frames in Figure 4.3

Figure 4.4 shows the representation of optical flow computation of two frames shown in figure 4.3. Noise seems to be apparent in the representation which seems to be part of the motion of the subject as well as the background. The overall optical flow representation is requires further post processing to yield a better optical flow representation.

4.3.4 Optical Flow Computation using Variational Approach

Brox et al in [91] implemented a novel variational approach containing an energy function based upon three assumptions: gradient constancy, brightness constancy and spatio-temporal smoothness constraint to preserve discontinuity. The approach is minimised with a numerical methods and is an integration of earlier published ideas. The use of warping is also known as coarse-to-fine technique. This idea is used to compute the non-linearised optical flow in image registration. The image correspondence issues are solved through a multi-resolution method. To measure the displacement vector, some small variations are allowed in the grey value through invariant criteria called gradient of the image grey value. Multi-scale ideas are applied to cater larger displacements. Using these assumptions, an energy function is derived that penalises deviations from these model assumptions.

Let x = $(x, y, t)^T$ and w = $(u, v, 1)^T$, the global deviations from the grey value constancy assumption and the gradient constancy assumption are measured as;

$$E_d(u,v) = \int_\Omega (|I(\mathrm{x}+\mathrm{w}) - I(\mathrm{x})|^2 + \gamma|\nabla I(\mathrm{x}+\mathrm{w}) - \nabla I(\mathrm{x})|^2)\, dx \qquad (4.23)$$

where γ represents a weight between both assumptions. An increasing concave function $\Psi(s^2)$ is implemented as outliers can impose influence on the estimation due to quadratic penalisers. It leads to a robust energy function;

$$E_d(u,v) = \int_\Omega \Psi(|I(\mathrm{x}+\mathrm{w}) - I(\mathrm{x})|^2 + \gamma|\nabla I(\mathrm{x}+\mathrm{w}) - \nabla I(\mathrm{x})|^2)\, dx \quad (4.24)$$

A piecewise smooth flow field generalises the smoothness assumption through a smoothness term.

$$E_{smooth}(u,v) = \int_\Omega \Psi(|\nabla_3 u|^2 + |\nabla_3 v|^2)\, dx \quad (4.25)$$

The spatio-temporal smoothness is performed using assumption based on spatio-temporal gradient $\nabla_3 = (\partial x, \partial y, \partial t)$. The total energy function is the weighted sum of the data term and the smoothness term:

$$E(u,v) = E_d + \alpha E_{smooth} \quad (4.26)$$

Numerical methods including Euler-Lagrange equations and Taylor expansions are applied based on two nested fixed point iterations. The coarse-to-fine technique as single minimisation problem is developed. Energy functions are minimised in image registration through multi-resolutions. The angular errors are significantly small for optical flow estimation due to parameter variations insensitivity.

4.3.5 Optical Flow Computation combining Classical Formulations with Modern Optimising Techniques

The accuracy of optical flow estimation algorithms has improved gradually as is evident from the results on the Middlebury optical flow benchmark [126]. The classical optical flow approaches in combination with modern optimization and implementation techniques perform well. The algorithm described in [130] by Sun et al is implemented to calculate accurate optical flow estimate. A baseline method is devised which is termed as "classical". The classical method is based on original Horn-Schunk formulation. The method and model is systematically varied using different state of the art techniques.

The classical optical flow subject function in its spatially discrete form is defined as;

$$E(u,v) = \sum_{i,j}\{\rho_D\left(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j})\right) + \lambda[\rho_S(u_{i,j} - u_{i+1,j}) + \rho_S(u_{i,j} - u_{i,j+1}) + \rho_S(v_{i,j} - v_{i+1,j}) + \rho_S(v_{i,j} - v_{i,j+1})]\} \quad (4.27)$$

*"where **u** and **v** are the horizontal and vertical components of the optical flow field to be estimated from images I_1 and I_2, λ is a regularization parameter, ρ_D and ρ_S are the data and spatial penalty functions respectively. Three different penalty functions are considered: (1) the quadratic HS penalty $\rho(x) = x^2$; (2) the Charbonnier penalty $\rho(x) = \sqrt{x^2 + \epsilon^2}$ [99], a differentiable variant of the L1 norm, the most robust convex function; and (3) the Lorentzian $\rho(x) = log\ (1 + \frac{x^2}{2\sigma^2})$, Note that this classical model is related to a standard pair-wise Markov random field (MRF) based on a 4-neighborhood."[130]*

4.3.5.1 Baseline Method

The structure texture decomposition method by Rudin-Osher-Fatemi (ROF) in [132] is applied following the approach in [131] to cater for lighting variations. The input sequences are pre-processed and texture and structure components are linearly combined (20:1 proportion). The parameters' setting is followed according to the description in [131].

A standard incremental [99] [133] multi-resolution approach is applied to compute optical flow fields with larger displacements. The optical flow estimation at coarse level is used to warp the second image toward the first at the next finer level. The flow increment is computed between the warped second image and the first image. The value of standard deviation of the Gaussian anti-aliasing filter is set to be $\frac{1}{\sqrt{2d}}$ where *d* is the down-sampling factor. Every level is down-sampled recursively from its nearest lower level. The down-sampling factor of 0:8 in the final stages of optimisation, is used according to the settings in [114] during the construction of the pyramid. The number of pyramid levels is established

adaptively in such a way that the top level has a width or height of around 20 to 30 pixels. The flow increment is calculated at every pyramid level by using 10 warping iterations.

The data term is linearised at every warping step that contains the following type of computing terms;

$$\frac{\partial}{\partial x} I_2(i + u_{i,j}^k, j + v_{i,j}^k) \tag{4.28}$$

where $\frac{\partial}{\partial x}$ represents the partial derivative in the horizontal direction, u^k and v^k are the current optical flow estimates at iteration *k*. A 5-point derivative filter $\frac{1}{12}[-1\ 8\ 0\ -8\ 1]$ is applied to calculate the derivatives for the second image. Current optical flow estimate through bi-cubic interpolation is applied to warp the second image and its derivatives toward the first [131]. The spatial derivatives of the first image are calculated and averaged with the warped derivatives of the second image [134]. $\frac{\partial I_2}{\partial x}$ is replaced with it. The corresponding spatial and temporal derivatives of pixels moving out of the image boundaries are set to zero. A *5* x *5* median filter is applied for the removal of outliers from the calculated optical flow fields after each warping iteration [131]. A graduated non-convexity scheme [135] [99] is used for the Charbonnier (Classic-C) and Lorentzian (Classic-L) penalty function. A quadratic subjective is linearly combined in varying proportions with a robust objective, from fully quadratic to fully robust. A single regularization weight λ is used for both the quadratic and the robust objective functions as opposed to [114].

4.3.5.2 Development of Baseline method into Improved Model

The median filtering heuristic can be further formulated as an explicit objective function. An over-smoothed result is achieved when median is computed on a neighbourhood

centered on a corner or thin structure which is dominated by the surrounding. This can be resolved through the examination of the non local term. For a given pixel, if other pixels in the area, belonging to the same surface, are known then they can be assigned higher weights. The introduction of a weight factor into the non local term imposes the modification to the objective function [136] [137]:

$$\sum_{i,j}\sum_{(i',j')\epsilon N_{i,j}} w_{i,j,i',j'}\left(\left|\hat{u}_{i,j}-\hat{u}_{i',j'}\right|+\left|\hat{v}_{i,j}-\hat{v}_{i',j'}\right|\right) \tag{4.29}$$

Where $w_{i,j,i',j'}$ shows the highly likely relationship of i',j' to the same surface as i,j. $w_{i,j,i',j'}$ can be approximated. The weights are defined according to their colour value distance, spatial distance and their occlusion state [138] [103] [139] as:

$$w_{i,j,i',j'} \propto exp\left\{-\frac{|i-i'|^2+|j-j'|^2}{2\sigma_1^2}-\frac{|I(i,j)-I(i',j')|^2}{2\sigma_2^2}\right\}\frac{o(i',j')}{o(i,j)}, \tag{4.30}$$

Eq. (22) in [138] is used to compute the occlusion $o(i,j)$. $I(i,j)$ is the colour vector in the lab space, σ_1 = 7 and σ_2 = 7. Higher weights represent the brighter values. In [141] the intervening contour defines affinities among neighbouring pixels for the local Lucas and Kanade algorithm [125]. Only the motion of sparse points is estimated using this scheme and dense optical flow field is then interpolated. Eq. 4.30 is solved approximately using the formula (3.13) in [140] for $\hat{u}$ and $\hat{v}$ as the following weighted median problem;

$$\min_{\hat{u}_{i,j}}\sum_{(i',j')\epsilon N_{i,j}\cup\{i,j\}} w_{i,j,i',j'}\left|\hat{u}_{i,j}-u_{i',j'}\right|, \tag{4.31}$$

The method Classic+NL-Full is for all the pixels. The solution becomes the median only in the case of all weights being equal. The faster version of this method without accuracy loss can be adopted. Motion boundaries can be detected for a given current optical flow estimate using a Sobel Edge detector. These edges can be dilated with a *5*x*5* mask to compute optical flow boundary regions. The weighting in Eq. 4.30 is applied in these regions using a *15*x*15* neighbourhood. Equal weights are applied to calculate the median in the non boundary regions using a *5*x*5* neighbourhood.

Figure 4.5 Two frames from fall action sequence

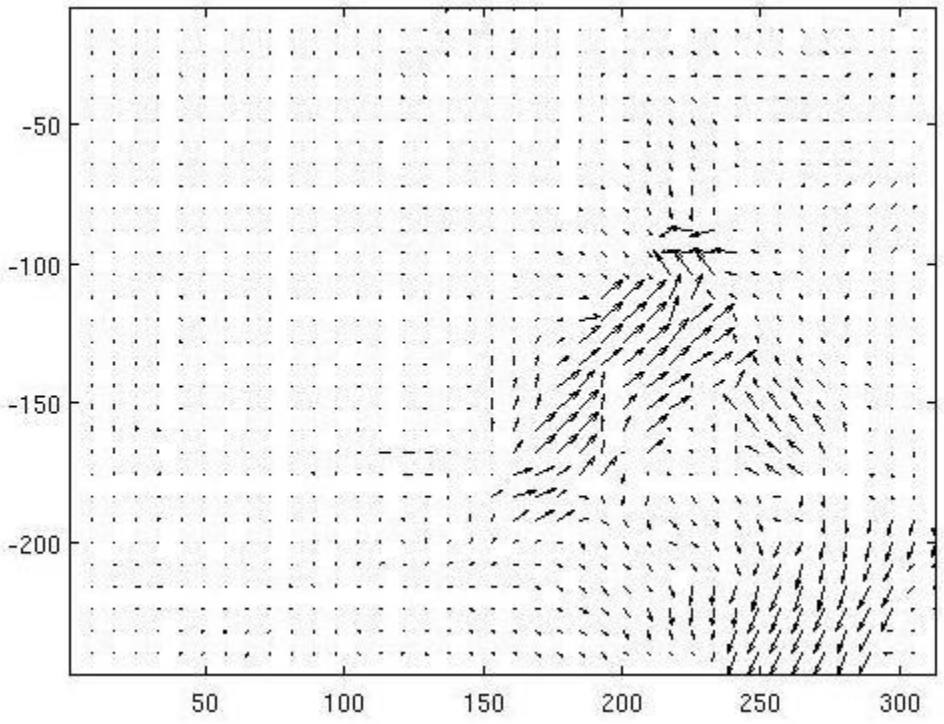

Figure 4.6 Optical Flow of two frames in Figure 4.6

Figure 4.6 displays the representation of optical flow computation of two frames shown in figure 4.5. The representation shows very accurate vector magnitudes and direction. The approach caters for the larger movements appropriately. The overall representation is

quite easy to draw a general estimate of optical flow field. There is hardly any noise in the representation. The overall optical flow representation is quite accurate.

4.4 Fall Detection Evaluation using different types of Feature Representation

Using simulated falls from [148] and some video sequences from our own dataset under supervised conditions and other actions as non falls from KTH dataset [155], the ability to distinguish between falls and non falls (other actions) was analysed and compared against two other approaches. Data analysis in terms of feature representation through descriptors was computed and classification was performed using MATLAB. The dataset of fall actions included; forward falls, backward falls and lateral falls left and right, performed with legs straight and flexed. An evaluation is performed using HOG3D, MHI and HOF for fall detection to analyse the results and draw the accuracy levels.

4.4.1 Descriptors

4.4.1.1 HOG3D

Three dimensional histogram of oriented gradients (HOG3D) is used in our evaluation as one of the descriptors for feature representation. It was proposed by Klaser et al [149]. The descriptor is based on histograms of 3D gradient orientations. It is also regarded as an extension of Scale Invariant Feature Transform (SIFT) descriptor [150]. An integral video representation is used to compute the gradients. The spatio-temporal gradients are uniformly quantised through regular polyhedrons. A combination of shape and motion information at the same time is used by the descriptor. A n_x x n_y x n_t cell scheme is used for a given 3D

patch. Normalisation is performed after concatenation of gradient histograms of all cells. The implementation is used from [151] and recommended parametric settings are applied. The number of spatial and temporal cells $n_x = n_y = 4$ and $n_t = 3$, descriptor size Δ_x (σ) = Δ_y (σ) = 8σ, Δ_t (σ), Δ_t (τ) = 6τ and the polyhedron type for quantizing orientations used is "Icosahedrons". [152]

4.4.1.2 MHI

MHI is a Motion-History Image formed to represent as how motion evolves. H_τ represents pixel intensity as a function of the temporal history of the motion at that point in an MHI. A simple replacement and decay operator is used:

$$H_\tau(x,y,t) = \begin{cases} \tau & if\ D(x,y,t) = 1 \\ \max(0, H_\tau(x,y,t-1) - 1) & otherwise, \end{cases} \quad (4.32)$$

Where $D(x,y,t)$ is a binary image sequence indicating regions of motion. The more recent motion pixels show brighter levels in a resulted scalar-valued image. The MHI can be thresholded above zero to generate Motion Energy Image (MEI). The direction of the motion in MHI is implicitly represented. Therefore the relation between the construction of the MHI and the direction of motion is important. A statistical description of the MHI is computed using moment based features. The settings in [153] are applied to achieve suitable shape discrimination in a translation and scale invariant manner. The MHI is flatten into a feature vector as the representation of the video sequence.

4.4.1.3 HOF

An accurately estimated optical flow from video sequence is one of the best ways to extract motion. A HOF descriptor is used following the approach in [154] and [152] by Laptev et al and Wang et al respectively. The image sequence is divided spatially and

temporally into sub regions as the number of spatial and temporal cells $n_x = n_y = 3$ and $n_t = 2$. The number of bins used in each region is 5. The total length of the descriptor is 90. 5-bin optical flow histograms (HOF) are computed. Normalized histograms are concatenated. The two dimensional histogram is calculated for each segment. The HOF descriptor settings are followed as described in [152], [151]. There is resemblance between HOF and HOG

4.4.2 Experimental Setup

The datasets used for the evaluation is described below. The features are computed using three types of descriptors described above. The feature based classification is performed using support vector machine (SVM).

4.4.3 Data Sets

We used two datasets. A very well-known dataset "KTH" and the other data set were taken from [148] and [154]. We also added some video sequences of our own simulated fall action.

4.4.3.1 KTH Dataset

There are six different types of human actions in the KTH dataset: jogging, walking, hand waving, hand clapping, boxing and running. 25 subjects perform each action several times. The video sequences contain four different scenarios: indoors, outdoors, outdoors with different clothes and outdoors with scale variation. The background is fixed (static) in some sequences and homogenous in most. There are 2391 video sequences in total in the dataset. We used 100 sequences for the training set and 45 for the testing set. At least 10 video sequences were taken from each action class for the training set and 5 video sequences for the

testing set. The training and evaluation is performed on a SVM classifier and accuracy is computed as performance measure.

4.4.3.2 Fall Action Data Set

There are fourteen different scenarios of human fall actions in the fall action dataset such as fall from walking, fall from sitting, fall from standing, backward fall, side way fall (left and right) etc. The background is static in all video sequences. Different subjects fall and the action is viewed from different viewpoints. The background is cluttered in the majority of the sequences including tables, chairs, etc. We have selected in total 46 fall video sequences for training set and 23 video sequences for testing with most of the fall types catered for. The training and evaluation is performed on a SVM classifier and accuracy is computed as performance measure.

4.4.4 Experimental Results

Experimental results in a tabular form are presented for three descriptors we used. We have compared their accuracy.

Table 4.1 System Detection Evaluation Results

	HOF	HOG3D	MHI
Accuracy	94%	82%	77%

The HOF descriptor clearly achieved higher accuracy compared to HOG3D and MHI as shown in table 4.1. Figures 4.7, 4.9 and 4.11 show some of the images from the video sequences used from the fall action dataset and their respective estimated optical flow in figures 4.8, 4.10 and 4.12, which was used to compute HOF features.

Figure 4.7 Fall Image Sequences

Table 4.2 System Detection Evaluation Format

System Detection / Fall Action	Fall	Non Fall
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

The table 4.2 presents tabular definition of TP, FP, FN and TN.

Table 4.3 Evaluation Results

System Detection / Fall Action	Fall	Non Fall
Positive	21 (TP)	3 (FP)
Negative	2 (FN)	42 (TN)

The table 4.3 shows the accuracy breakdown of our fall detection system. The fall detection system only detected 2 false negatives (FN) of fall action as non fall out of 23 action fall sequences and detected only 3 false positives (FP) of non fall action as fall out of 45 non fall action sequences.

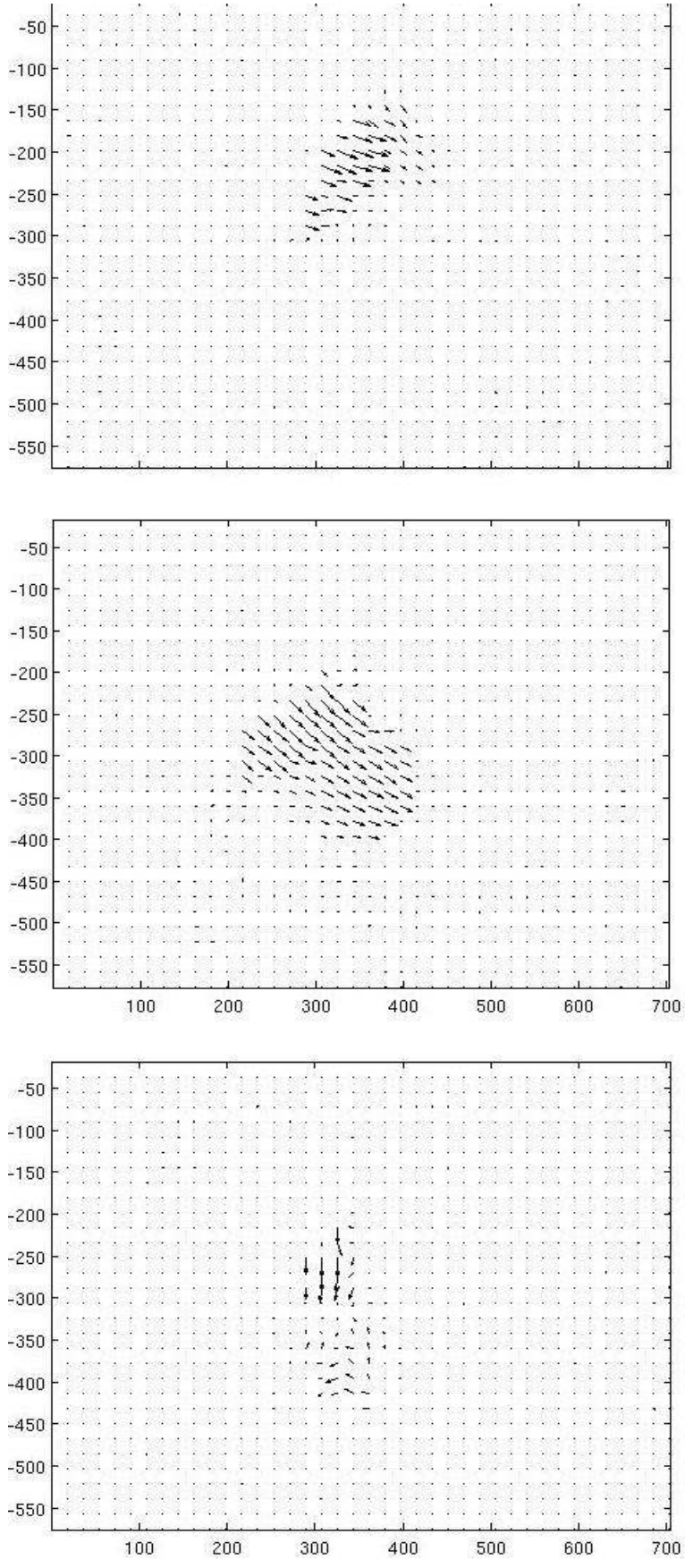

Figure 4.8 Optical Flows of the last two frames of Fall Image Sequences in Figure 4.7

Figure 4.9 Fall Image Sequences

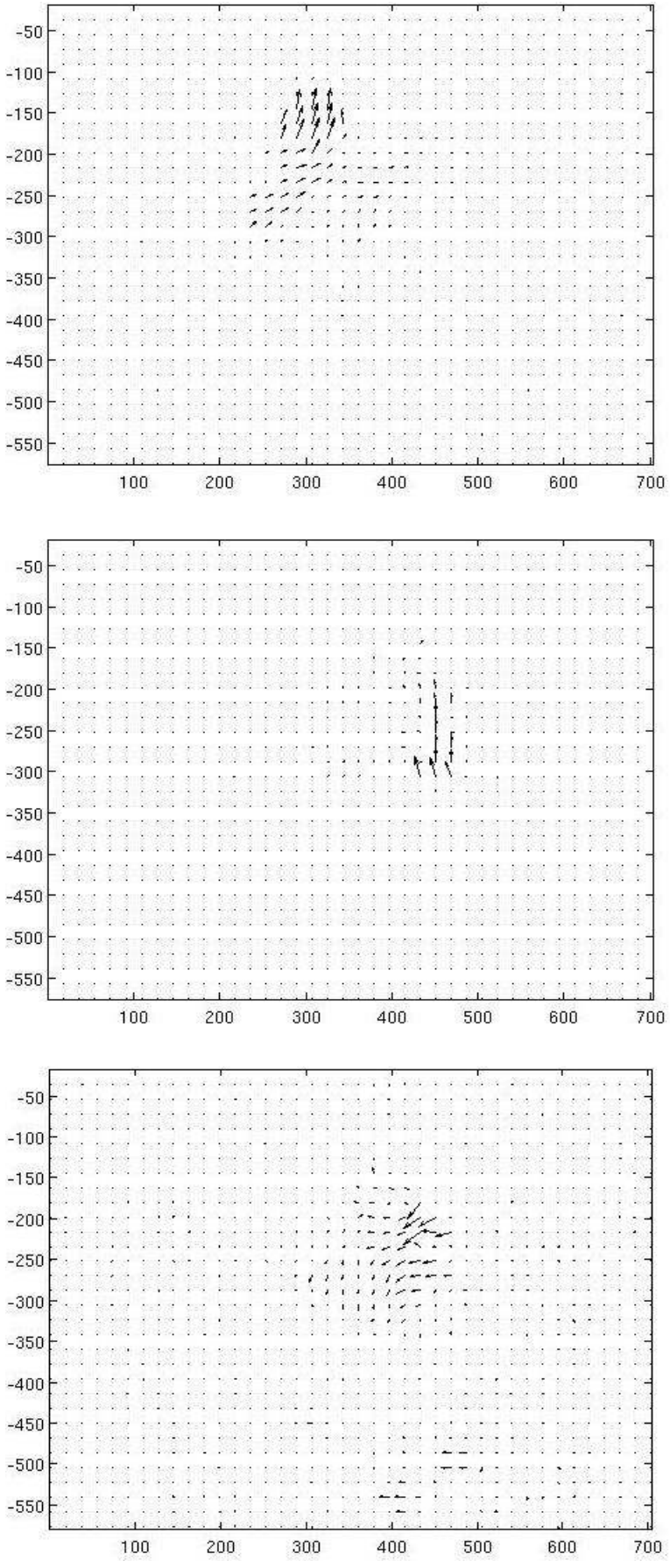

Figure 4.10 Optical Flows of the last two frames of Fall Image Sequences in Figure 4.9

Figure 4.11 Fall Image Sequences

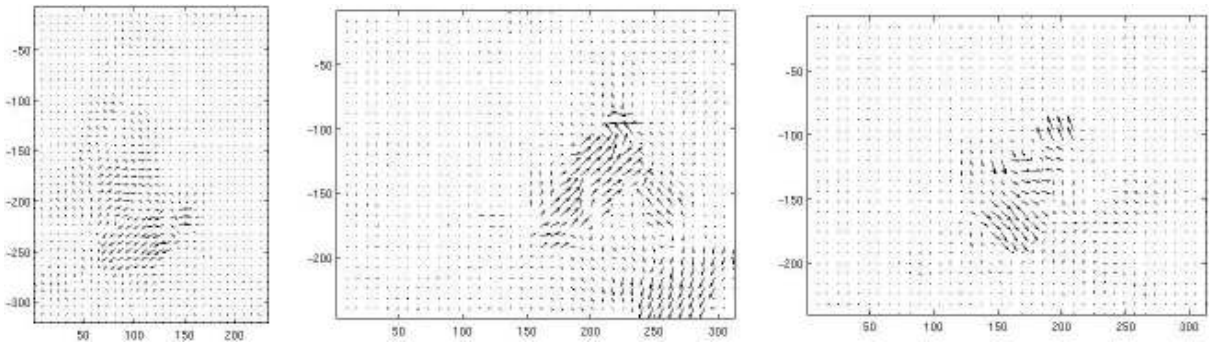

Figure 4.12 Optical Flows of the last two frames of Fall Image Sequences in Figure 4.11

We use a set of following criteria described in [156] to evaluate our system, including sensitivity and specificity.

$$sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{TN + FN}$$

Table 4.4 Final System Performance

Accuracy	94%
Sensitivity	91%
Specificity	93%

4.5 Discussion

Significant statistical improvements that lead to state of the art accuracy levels can be achieved through mere key adjustments. The application of median filter heuristics to intermediate optical flow estimates during warping and incremental estimation has shown fruitful results. Therefore this heuristic is of prominent interest as it has added robustness to the methods and improved the overall accuracy. It also enhanced the accuracy of the recovered optical flow fields, thus increasing the overall energy of the objective function. The new term integrates information over large spatial neighbourhoods robustly. The idea of outlier removal through median filtering after incremental estimation has shown very encouraging results. Not only is the energy of the final output of the optical flow estimation amplified through median filtering, but also median filtering has improved the overall accuracy by reducing noise in the flow estimates after warping iteration. The connections between median filtering and de-noising can be further exploited and large region of spatial neighbourhood can be regularised.

The challenges for optical flow algorithms focus on factors like motion discontinuities and large displacements etc. Solid progress has been made in the field of optical flow

estimation. Current methods applied on the Middlebury optical flow benchmark [126] show higher levels of accuracy. Most optical flow approaches are built on specific data term, prior term, and optimization procedure for the flow field computation. The algorithm must be able to deal with the aspects involved in making the optical flow complex and inherently ambiguous. The aspects make the optical flow intrinsically ill-posed such as texture-less regions and the aperture problem. Suitable penalty functions are required for prior and data terms to deal with factors like occlusions, noise, non-rigid motion and motion discontinuities. Optimisation procedures falling into local minima due to small subjects and larger displacement require attention. Furthermore illumination changes, motion blur, mixed pixels etc are some of the other commonly know aspects.

Most of the currently implemented approaches have strong resemblance with Horn and Schunck's original method of optical flow estimation. The original algorithms of Horn and Schunck and Lucas-Kannade are both very competitive with respect to adequate implementation. The introduction of coarse-to-fine estimation solves the issues of filter constancy of high order as larger displacements. Illumination changes have been reduced to an acceptable extent. Other added functionalities include bi-cubic interpolation based warping, temporal averaging of image derivatives and graduated non-convexity to minimize non-convex energies. The above described added functionalities permit to implement optimised models. Also the neighbours in a broadened image block can be weighted adaptively. Surfaces and boundaries can be further explored in the temporal domain to extract additional explicit explanations. Using classical formulation and going beyond can yield advanced growth.

Chapter 5

Conclusion

5.1 Conclusions

We have addressed several issues in human fall action detection analysis using optical flow estimation. The thesis is composed of two essential parts. An overview of fall action detection and optical flow estimation at the levels ranging from fundamental approaches to higher abstraction techniques in chapter 2 and chapter 4, the implementation of optical flow computation algorithms and using the most accurate algorithm of optical flow for fall detection evaluation in chapter 4. Chapter 2 contains our contribution of a detailed survey of currently implemented approaches and techniques for fall detection in all areas of research and industry including computer vision. Chapter 4 consists of our contribution of using optical flow technique for fall detection evaluation. In the following, we summarize the contents and findings of each individual chapter.

Chapter 1 contained the introduction, motivation, our contributions and the structure outline of our research study. It introduced and described the facts and background of our research area "Fall Detection". Then we have described the motivation behind this research study using the available facts within our research area and industry. We have also briefly discussed the ongoing trends and frequency of research of our research area in the research community and industry. We have also provided brief information about our contributions in the thesis. Finally we described the structure of thesis in an outline.

Chapter 2 presented one of the contributions of our research study. We have presented a comprehensive survey of fall detection approaches and techniques used by the research community and the industry. Different types of fall action are introduced, followed

by the classification of fall detection methods into three categories i.e. Wearable devices, Ambient and Vision based. We have reviewed three different categories of fall detection approaches. At the end of the chapter, we have discussed, concluded and pointed out a way forward as the use of optical flow for the future research in the area of fall detection. This chapter has laid down an ideal foundation for our research study.

Chapter 3 comprised of an overview of currently implemented background subtraction techniques. We initiated our research study with background subtraction as generally used procedure in surveillance related applications. The background subtraction can lead to minimised information for post processing for computational efficiency. We have implemented some of the background subtraction techniques and discussed their results. The dataset requirements in terms of lighting conditions, resolution and illumination changes etc were not met by our dataset once we constructed our dataset later in conjunction with the other datasets used in our research study. Therefore we haven't used the background subtraction later in our research study.

Chapter 4 presented our other contribution of this research study. We have described an overview of optical flow estimation techniques. We have implemented several optical flow estimation techniques such as Lucas-Kannade, Horn-Schunk, Variational approaches etc and discussed the quality of the results achieved. We have also recommended the technique built on classical formulations with modern optimising procedures that yielded the best quality of the experimental results. We have applied this recommended technique to demonstrate an evaluation of fall detection. We compared the experimental results achieved against the other techniques. The experimental results clearly demonstrated the superiority of our recommended approach in terms of higher achieved accuracy of 94%.

5.2 Future studies

The analysis of fall detection can be further exploited for post processing. The fall action belongs to a category of actions which are not majorly intentional rather abnormal. The fact that fall is a natural but abnormal action draws greater deal of attention towards its complications. The study of such complicated actions can lay down ideal foundation for abnormal activity detection. This can further lead to perfect initiation of grounds for abnormal behaviour analysis in the area of surveillance, thus making the accurate automatic monitoring of such complex activities and behaviours possible in the real time industry applications. Despite the high accuracy achieved in our fall detection evaluation and experimental results, there are still few aspects that need further investigation.

5.2.1 Fusion of Data

The current surveillance research community and industry are focusing their research studies and applications using multiple cameras and multiple subjects. This requires further analysis of optical flow estimation techniques. Other than using visual information, audio, ultrasonic and infrared data can also be used to detect and analyse information especially in the area of surveillance as these additional sensors can provide significant and unique information. This stands out as one of the very complicated and challenging tasks in terms of fusion of data and its post processing. One of the problems in using such complicated and comprehensive data will be data fusion and standards to classify the use of information.

5.2.2 Embedded Applications

The optical flow estimation technique used in our experimental setup and evaluation does have higher memory requirements, thus adding to the computational load. Computational efficiency remains a very important criterion with respect to real time

applications within the industry. Further post processing analysis can be carried out to look for the factors affecting efficiency. Efficient approaches can be further exploited on suitable embedded platforms for experimentation. The optimised algorithms can be mapped using digital signal processing media processors to facilitate smart video surveillance e.g. accurate automatic surveillance in a care home etc. The system can be developed using standard hardware and simple technology for end user.

References

[1] S. J. McKenna and H. Nait-Charif, *Summarising Contextual Activity and Detecting Unusual Inactivity in a Supportive Home Environment*, 17th IEEE International Conference on Pattern Recognition, Vol. 4, pp. 323-326, 2004.

[2] A. M. Tabar, A. Keshavarz and H. Aghajan, *Smart Home Care Network using Sensor Fusion and Distributed Vision-based Reasoning,* 4th ACM International Workshop on Video Surveillance and Sensor Networks, 2006.

[3] N. Noury, T. Herd, V. Rialle, G. Virone, E. Mercier, G. Morey, A. Moro, and T. Porcheron, *Monitoring Behaviour in Home using a Smart Fall Sensor and Position Sensors,* 1st Annual International Conference On Micro Technologies in Medicine and Biology, pp. 607-610, 2000.

[4] H. Foroughi, A. Naseri, A. Saberi, and H. S. Yazdi, *An Eigenspace-Based Approach for Human Fall Detection Using Integrated Time Motion Image and Neural Network,* 9th IEEE International Conference on Signal Processing (ICSP), pp. 1499-1503, 2008.

[5] H. Foroughi, B. S. Aski, and H. Pourreza, *Intelligent Video Surveillance for Monitoring Fall Detection of Elderly in Home Environments,* 11th IEEE International Conference on Computer and Information Technology (ICCIT), pp. 219-224, 2008.

[6] S. Luo and Q. Hu, *A Dynamic Motion Pattern Analysis approach to Fall Detection*, IEEE International Workshop on Biomedical Circuits & Systems, pp. 1-5, 2004.

[7] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell and B. G. Celler, *Implementation of a Real-Time Human Movement Classifier Using a Triaxial Accelerometer for Ambulatory Monitoring,* IEEE Transactions on Information Technology in Biomedicine, Vol. 10(1), pp. 156-157, 2006.

[8] J. Chen, K. Kwong, D. Chang, J. Luk, and R. Bajcsy, *Wearable Sensors for Reliable Fall Detection*, 27th IEEE Annual Conference of Engineering in Medicine and Biology (EMBS), pp. 3551-3554, 2005.

[9] S. G. Miaou, P. H. Sung and C. Y. Huang, *A Customized Human Fall Detection System Using Omni-Camera Images and Personal Information*, 1st Trans-Disciplinary Conference on Distributed Diagnosis and Home Healthcare (D2H2), pp. 39-42, 2006.

[10] M. Alwan , P. J. Rajendran, S. Kell, D. Mack , S. Dalal, M. Wolfe, and R. Felder, *A Smart and Passive Floor-Vibration Based Fall Detector for Elderly*, IEEE International Conference on Information & Communication Technologies (ICITA), pp. 1003-1007, 2006.

[11] J. Tao, M. Turjo, M. F. Wong, M. Wang and Y. P. Tan: *Fall Incidents Detection for Intelligent Video Surveillance,* Fifth IEEE International Conference on Information, Communications and Signal Processing, pp. 1590-1594, 2005.

[12] N. Thome and S. Miguet, *A HHMM-Based Approach for Robust Fall Detection,* 9th IEEE International Conference on Control, Automation, Robotics and Vision (ICARCV), pp. 1-8, 2006.

[13] B. Jansen and R. Deklerck, *Context aware inactivity recognition for visual fall detection,* IEEE Pervasive Health Conference and Workshops, pp. 1-4, 2006.

[14] C. Rougier, J. Meunier, A. St-Arnaud and J. Rousseau, *Fall Detection from Human Shape and Motion History using Video Surveillance,* 21st IEEE International Conference on Advanced Information Networking and Applications Workshops (AINAW'07), pp. 875-880, 2007.

[15] C. W. Lin and Z. H. Ling, *Automatic Fall Incident Detection in Compressed Video for Intelligent Homecare*, 16th IEEE International Conference on Computer Communications and Networks (ICCCN), pp. 1172-1177, 2007.

[16] M. Kangas, A. Konttila, I. Winblad and T. Jämsä, *Determination of simple thresholds for accelerometry-based parameters for fall detection,* 29th IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBS), pp. 1367-1370, 2007.

[17] S. Srinivasan, J. Han, D. Lal and A. Gacic, *Towards automatic detection of falls using wireless sensors*, 29th IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBS), pp. 1379-1382, 2007.

[18] Y. Lee, J. Kim, M. Son, and M. Lee, *Implementation of Accelerometer Sensor Module and Fall Detection Monitoring System based on Wireless Sensor Network*, 29th IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBS), pp. 2315-2318, 2007.

[19] M. R. Narayanan, S. R. Lord, M. M. Budge, B. G. Celler and N. H. Lovell, *Falls Management: Detection and Prevention, using a Waist mounted Triaxial Accelerometer,* 29th IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBS), pp. 4037-4040, 2007.

[20] G. Wu and S. Xue, *Portable Preimpact Fall Detector with Inertial Sensors,* IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 16, pp. 178-183, 2008.

[21] C. C. Wang, C. Y. Chiang, P. Y. Lin, Y. C. Chou, I. T. Kuo, C. N. Huang and C. T. Chan, *Development of a Fall Detecting System for the Elderly Residents*, 2nd IEEE International Conference on Bioinformatics and Biomedical Engineering, ICBBE, pp. 1359-1362, 2008.

[22] Z. Fu, E. Culurciello, P. Lichtsteiner and T. Delbruck, *Fall Detection using an Address-Event Temporal Contrast Vision Sensor*, IEEE Transactions on Biomedical Circuits and Systems, Vol. 2, pp. 88-95, 2008.

[23] C. J. Robinson, M. C. Purucker and L. W. Faulkner, *Design, Control, and Characterization of a Sliding Linear Investigative Platform for Analyzing Lower Limb*

Stability (SLIP-FALLS), IEEE Transactions on Rehabilitation Engineering, Vol. 6, pp. 334-350, 1998.

[24] A. Sixsmith and N. Johnson, *A Smart Sensor to Detect the Falls of the Elderly*, IEEE Pervasive Computing, IEEE CS and IEEE ComSoc, Vol. 3, pp. 42-47, 2004.

[25] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, *Probabilistic Posture Classification for Human-Behavior Analysis*, IEEE Transactions on Systems Man and Cybernetics - Part A: Systems and Humans, Vol. 35, pp. 42-54, 2005.

[26] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovelland and B. G. Celler, *Implementation of a Real Time Human Movement Classifier Using a Tri-axial Accelerometer for Ambulatory Monitoring*, IEEE Transactions on Information Technology in Biomedicine, Vol. 10, pp. 156-167, 2006.

[27] J. M. Kang, T. Yoo and H. C. Kim, *A Wrist-Worn Integrated Health Monitoring Instrument with a Tele-Reporting Device for Telemedicine and Telecare*, IEEE Transactions on Instrumentation and Measurement, Vol. 55, pp. 1655-1661, 2006.

[28] C. F. Juang and C. M. Chang, *Human Body Posture Classification by a Neural Fuzzy Network and Home Care System Application*, IEEE Transactions on Systems, Man and Cybernetics Part A: Systems and Humans, Vol. 37, pp. 984-994, 2007.

[29] A. H. Khandoker, D. T. H. Lai, R. K. Begg and M. Palaniswami, *Wavelet-Based Feature Extraction for Support Vector Machines for Screening Balance Impairments in the Elderly*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 15,pp. 587-597, 2007.

[30] G. Wu and S. Xue: *Portable Preimpact Fall Detector with Inertial Sensors*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 16, pp. 178-183, 2008.

[31] N. Thome, S. Miguet and S. Ambellouis, *A Real-Time, Multiview Fall Detection System*: *A LHMM-Based Approach, Circuits and Systems for Video Technology*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 18, pp. 1522-1532, 2008.

[32] S. Fleck and W. Strasser, *Smart Camera Based Monitoring System and Its Application toAssisted Living*, 4th IEEE Workshop on Embedded Systems Security, Vol. 96, pp. 1698-1714, 2008.

[33] G. Shi, C. S. Chan, W. J. Li, K. S. Leung, Y. Zou and Y. Jin; *Mobile Human Airbag System for Fall Protection Using MEMS Sensors and Embedded SVM Classifier*, IEEE Sensors Journal, Vol. 9, pp. 495-503, 2009.

[34] Y. Zigel, D. Litvak and I. Gannot; *A Method for Automatic Fall Detection of Elderly People Using Floor Vibrations and Sound Proof of Concept on Human Mimicking Doll Falls*, IEEE Transactions on Biomedical Engineering, Vol. 56, pp. 2858-2867, 2009.

[35] M.A. Estudillo-Valderrama, L.M. Roa, J. Reina-Tosina and D. Naranjo-Hernandez, *Design and Implementation of a Distributed Fall Detection System - Personal Server*, IEEE Transactions on Information Technology in Biomedicine, Vol. 13, pp. 874-881, 2009.

[36] T. Tamura, T. Yoshimura, M. Sekine, M. Uchida and O. Tanaka, *A Wearable Airbag to Prevent Fall Injuries*, IEEE Transactions on Information Technology in Biomedicine, Vol.13, pp. 910-914, 2009.

[37] H. Ghasemzadeh, R. Jafari and B. Prabhakaran, *A Body Sensor Network With Electromyogram and Inertial Sensors: Multimodal Interpretation of Muscular Activities*, IEEE Transactions on Information Technology in Biomedicine, Vol. 14, pp. 198-206, 2010.

[38] H. Ghasemzadeh, V. Loseu and R. Jafari, *Structural Action Recognition in Body Sensor Networks: Distributed Classification Based on String Matching*, IEEE Transactions on Information Technology in Biomedicine, Vol. 14, pp. 425-435, 2010.

[39] H. Rimminen, J. LindstroNm, M. Linnavuo and R. Sepponen, *Detection of Falls Among the Elderly by a Floor Sensor Using the Electric Near Field*, IEEE Transactions on Information Technology in Biomedicine, Vol. 14, pp. 1475-1476, 2010.

[40] F. Bianchi, S. J. Redmond, M. R. Narayanan, S. Cerutti and N. H. Lovell, *Barometric Pressure and Triaxial Accelerometry Based Falls Event Detection*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 18, pp. 619-627, 2010.

[41] C. F. Lai, S. Y. Chang, H. C. Chao and Y. M. Huang, *Detection of Cognitive Injured Body Region Using Multiple Tri-axial Accelerometers for Elderly Falling*, IEEE Sensors Journal, Vol. 11, pp. 763-770, 2011.

[42] M. J. Mathie, A. C. F. Coster, N. H. Lovell and B. G. Celler, *Accelerometry: Providing an Integrated, Practical Method for Long-term, Ambulatory Monitoring of Human Movement*, Journal for Physiological Measurement (IOPScience), Vol. 25, 2004.

[43] G. Wu, *Distinguishing Fall Activities from Normal Activities by Velocity Characteristics*, Elsevier Journal of Biomechanics, Vol. 33, pp. 1497-1500, 2000.

[44] V. Vishwakarma, C. Mandal and S. Sural, *Automatic Detection of Human Fall in Video*, Springer-Verlag 2nd international conference on Pattern recognition and machine intelligence (PReMI'07), 2007.

[45] P. Boissy, S. Choquette, M. Hamel and N. Noury, *User-Based Motion Sensing and Fuzzy Logic for Automated Fall Detection in Older Adults*, Telemedicine and e-Health Mary Ann Liebert, Vol. 13, pp. 683-694, 2007.

[46] K. H. Wolf, A. Lohse, M. Marschollek and R. Haux, *Development of a Fall Detector and Classifier based on a Tri-axial Accelerometer Demo Board*, IEEE Journal for Pervasive Computing & UbiWell workshop for Healthcare Applications, 2007.

[47] P. A. Bromiley, P. Courtney and N. A. Thacker, *Design of A Visual System for Detecting Natural Events by the Use of an Independent Visual Estimate: A Human Fall Detector*, TINA Vision Publications & Empirical Evaluation methods in Computer Vision, pp. 61-87, 2002.

[48] B. U. Toreyin, E. B. Soyer, I. Onaran, and A. E. Cetin, *Falling Person Detection Using Multi-sensor Signal Processing*, 15th IEEE Signal Processing and Communications Applications Conference, SIU & EURASIP Journal on Advances in Signal Processing, Vol. 8, 2007 & 2008.

[49] T. Zhang, J. Wang, L. Xu and P. Liu; *Using Wearable Sensor and NMF Algorithm to Realize Ambulatory Fall Detection*, Springer-Verlag, Vol. 4222, pp. 488-491, 2006.

[50] C. Rougier and J. Meunier; Demo: *Fall Detection Using 3D Head Trajectory Extracted From a Single Camera Video Sequence*, First International Workshop on Video Processing for Security (VP4S), 2006.

[51] T. Zhang, J. Wang, P. Liu and J. Hou, *Fall Detection by Embedding an Accelerometer in Cellphone and Using KFD Algorithm,* International Journal of Computer Science and Network Security (IJCSNS), Vol. 6, 2006.

[52] B. Kaluza and M. Lustrek, *Fall Detection and Activity Recognition Methods for the Confidence Project: A Survey*, *12th International* Multi-conference Information Society, vol. A, pp. 22-25, 2009.

[53] R. Cucchiara, A. Prati and R. Vezzani, *A Multi-Camera Vision System for Fall Detection and Alarm Generation*, Expert Systems Journal, Vol. 24, pp. 334-345, 2007.

[54] X. Zhuang, J. Huang, G. Potamianos and M. Hasegawa-Johnson, *Acoustic Fall Detection Using Gaussian Mixture Models and GMM Super-Vectors*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.69-72, 2009.

[55] B. U. Toreyin, Y. Dedeoglu and A. E. Cetin, *HMM Based Falling Person Detection Using Both Audio and Video*, 14th IEEE Signal Processing and Communications Applications Conference, 2006 & Springer-Verlag, Vol. 3766, pp. 211-220, 2005.

[56] M. Kangas, I. Vikman, J. Wiklander, P. Lindgren, L. Nyberg and T. Jamsa, *Sensitivity and Specificity of Fall Detection in People Aged 40 Years and Over*, Gait Posture, ELSEVIER 29, pp. 571-574, 2009.

[57] C. Doukas, I. Maglogiannis, F. Tragkas, Dimitris, Liapis and G. Yovanof, *Patient Fall Detection Using Support Vector Machines*, International Federation for Information Processing (IFIP), SpringerLink, Vol. 247, pp. 147-156, 2007.

[58] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz and M. Aud, *Linguistic Summarization of Video for Fall Detection Using Voxel Person and Fuzzy Logic*, Computer Vision and Image Understanding (CVIU), ELSEVIER, Vol. 113, pp. 80-89, 2009.

[59] M. N. Nyan, F. E. H. Tay, A. W. Y. Tan and K. H. W. Seah, *Distinguishing Fall Activities from Normal Activities by Angular Rate Characteristics and High-Speed Camera Characterization*, Medical Engineering & Physics, ELSEVIER 28, pp. 842–849, 2006.

[60] X. Yu, *Approaches and Principles of Fall Detection for Elderly and Patient*, 10th IEEE International Conference on e-Health Networking, Applications and Services (HealthCom), pp. 42-47, 2008.

[61] N. Noury, A. Fleury, P. Rumeau, A. K. Bourke, G. Ó. Laighin, V. Rialle and J. E. Lundy, *Fall Detection-Principles and Methods*, 29th IEEE International Conference on Engineering in Medicine and Biology Society (EMBS), pp. 1663-1666, 2007.

[62] C. Griffiths, C. Rooney and A. Brock, *Leading Causes of Death in England and Wales – How Should We Group Causes?* Health Statistics Quarterly 28, Office for National Statistics, 2008.

[63] L Hazelhoff, J Han and P.H.N. de With, *Video-Based Fall Detection in the Home Using Principal Component Analysis*, Advanced concepts for Intelligent Vision Systems (ACIVS), SpringerLink, Vol. 5259, pp. 298-309, 2008

[64] C Rougier, J Meunier, A St-Arnaud, and J Rousseau, *Robust Video Surveillance for Fall Detection Based on Human Shape Deformation*, IEEE Transactions on Circuits and Systems for Video Technology (CSVT), Vol. 21, pp. 611-622, 2011

[65] A Williams, D Ganesan and Hanson, Aging in place: *Fall detection and localization in a distributed smart camera network*, 15th International Conference on Multimedia, pp. 892-901, 2007

[66] S. Cheung and C. Kamath, *Robust techniques for background subtraction in urban traffic video, Journal on Applied Signal Processing*, EURASIP, Vol. 05, pp. 2330-2340, 2005

[67] V. Mahadevan and N. Vasconcelos, *Background Subtraction in Highly Dynamic Scenes*, IEEE conference on computer vision and Pattern Recognition, CVPR, pp. 1-6, 2008

[68] Á. Bayona, J. SanMiguel and J. Martínez, *Comparative evaluation of stationary foreground object detection algorithms based on background subtraction techniques*, Sixth IEEE international Conference on Advanced Video and Signal based Surveillance, AVSS, pp. 25-30, 2009

[69] J. Hu and T. Su, *Robust Background Subtraction with Shadow and Highlight Removal for Indoor Surveillance*, IEEE International Conference on Intelligent Robots and Systems, pp. 4550-4550, 2006 & Journal on Advances in Signal Processing, EURASIP, Vol. 07, 2007

[70] C. Stauffer and W. Grimson, *Adaptive background mixture models for real-time tracking*, IEEE International Conference on computer vision and Pattern Recognition, CVPR 1999 & IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, pp. 747-757 2000

[71] I. Haritaoglu, D. Harwood and L. Davis, *Real-Time Surveillance of People and Their Activities*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, pp. 809-830, 2000

[72] Y. Sun and B. Yuan, *Hierarchical GMM to handle sharp changes in moving object detection*, IET Electronics Letters, Vol. 14, pp. 801-802, 2004

[73] Zoran Zivkovic, *Improved Adaptive Gaussian Mixture Model for Background Subtraction*, 17th International Conference on Pattern Recognition, ICPR, Vol. 2, pp. 28-31, 2004

[74] M. Heikkila and M. Pietikainen, *A Texture-Based Method for Modelling the Background and Detecting Moving Objects*, IEEE Transactions on Pattern Analysis and Machine Learning, Vol. 28, pp. 657-662, 2006

[75] J. Migdal and W. Grimson, *Background Subtraction Using Markov Thresholds*, IEEE Workshop on Motion and Video Computing, WACV/MOTIONS, Vol. 2, pp. 58-65, 2005

[76] Z. Tang and Z. Miao, *Fast Background Subtraction and Shadow Elimination Using Improved Gaussian Mixture Model*, IEEE International Workshop on Haptic Audio Visual Environments and their Applications, HAVE, pp. 38-41, 2007

[77] Z. Tang and Z. Miao, *Fast Background Subtraction Using Improved GMM and Graph Cut*, Congress on Image and Signal processing, CISP, pp. 181-185, 2008

[78] P. Suo and Y. Wang, *An Improved Adaptive Background Modelling Algorithm Based on Gaussian Mixture Model*, International Conference on Signal Processing, ICSP, pp. 1436-1439, 2008

[79] M. Unger, M. Asbach, and P. Hosten, *ENHANCED BACKGROUND SUBTRACTION USING GLOBAL MOTION COMPENSATION AND MOSAICING*, 15th IEEE International Conference on Image Processing, ICIP, pp. 2708-2711, 2008

[80] Y. Benezeth, P. Jodoin, B. Emile, H. Laurent and C. Rosenberger, *Review and Evaluation of Commonly-Implemented Background Subtraction Algorithms*, 19th International Conference on Pattern Recognition, ICPR, pp. 1-4, 2008

[81] M. Izadi and P. Saeedi, *Robust Region-Based Background Subtraction and Shadow Removing Using Colour and Gradient Information*, 19th International Conference on Pattern Recognition, ICPR, pp. 1-5, 2008

[82] K. Kim, T. Chalidabhongse, D. Harwood and L. Davis, *Real-time foreground–background segmentation using codebook model,* Real Time Imagine, ELSEVIER, Vol. 11, pp. 172-185, 2005

[83] Z. Zivkovic and F. Heijden, *Efficient adaptive density estimation per image pixel for the task of background subtraction*, Pattern Recognition Letters, ELSEVIER, Vol. 27, pp. 773-780, 2006

[84] J. Cheng, J. Yang, Y. Zhou and Y. Cui, *Flexible background mixture models for foreground segmentation, Image and Vision Computing*, ELSEVIER, Vol. 24, pp. 473-482, 2006

[85] D. Zhou and H. Zhang, *Modified GMM Background Modelling and Optical Flow for Detection of Moving Objects*, IEEE International Conference on Systems, Man and Cybernetics, Vol. 3, pp. 2224 – 2229, 2005

[86] J. Cheng, J. Yang, and Y. Zhou, *A Novel Adaptive Gaussian Mixture Model for Background Subtraction*, Pattern Recognition and Image Analysis, SpringerLink and IbPRIA, Vol. 3522, pp. 193-216, 2005

[87] T. Ko, S. Soatto, and D. Estrin, *Background Subtraction on Distributions*, 10th European Conference on Computer Vision & LNCS Springer-Verlag, Vol. 5304, pp. 567-580, 2008

[88] D. Parks and S. Fels, *Evaluation of Background Subtraction Algorithms with Post-processing*, 5th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS, pp. 192-199, 2008

[89] S. Geman and D. Geman, *Stochastic Relaxation, Gibbs distributions and the Bayesian Restoration of Images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984

[90] J. Ohm and H. Luke, *Signal ubertragung*, vol. 9, Springer-Verlag, 2005

[91] T. Brox, A. Bruhn, N. Papenberg and J. Weickert, *High Accuracy Optical Flow Estimation Based on a Theory for Warping*, 8th European Conference on Computer Vision & Springer LNCS, vol. 4, pp. 25-36, 2004

[92] S. Lim and Abbas El Gamal, *Optical Flow Estimation using High Frame Rate Sequences,* International Conference on Image Processing, Vol. 2, pp. 925-928, 2001

[93] L. Alvarez, R Deriche, T. Papadopoulo and J. Sanchez, *Symmetrical Dense Optical Flow Estimation with Occlusions Detection*, European Conference on Computer Vision (ECCV) and SpringerLink, Vol. 2350, pp. 721-725, 2002

[94] H. Nagel and W. Enkelmann, *An Investigation of Smoothness Constraint for the Estimation of Displacement Vector Fields from Image Sequences,* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, pp. 565–593, 1986

[95] T. Brox and J. Weickert, *Nonlinear Matrix Diffusion for Optic Flow Estimation*, Pattern Recognition, SpringerLink, Vol. 2449, pp. 446-453, 2002

[96] T. Nir, A. Bruckstein and R. Kimmel, *Over-Parameterized Variational Optical Flow*, International Journal of Computer Vision, SpringerLink, Vol. 76, pp. 205-216, 2007

[97] S. Roth and M. Black, *On the Spatial Statistics of Optical Flow*, 10th IEEE International Conference on Computer Vision (ICCV), 2005 and International Journal of Computer Vision, SpringerLink, Vol. 74, pp. 33-50, 2007

[98] S. Roth and M. J. Black, *Fields of experts: A Framework for Learning Image Priors,* IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 860–867, 2005

[99] A. Bruhn, J. Weickert, and C. Schnorr, *Lucas-Kanade meets Horn-Schunck: Combining local and global optic flow methods*, International Journal of Computer Vision, SpringerLink, Vol. 61, pp. 211–231, 2005

[100] A. Efros, A. Berg, G. Mori and J. Malik, *Recognizing Action at a Distance*, 9th IEEE International Conference on Computer Vision (ICCV), Vol. 2, pp. 726-733, 2003

[101] T. Brox, B. Rosenhahn, D. Cremers and H. Seidel, *High Accuracy Optical Flow Serves 3-D Pose Tracking: Exploiting Contour and Flow Based Constraints*, European Conference on Computer Vision (ECCV) and LNCS, Springer-Verlag, Vol. 3952, pp. 98-111, 2006

[102] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger and C. Schnorr, *Real-Time Optic Flow Computation with Variational Methods*, IEEE Transactions on Image Processing, Vol. 14, pp. 608-615, 2005

[103] J. Xiao, H. Cheng, H. Sawhney, C. Rao, and M. Isnardi, *Bilateral Filtering-based Optical Flow Estimation with Occlusion Detection*, European Conference on Computer Vision and LNCS, SpringerLink, Vol. 3951, pp. 211-224, 2006

[104] B. McCane, K. Novins, D. Crannitch and B. Galvin, *On Benchmarking Optical Flow*, Computer Vision and Image Understanding, ELSEVIER, Vol. 84, pp. 126-143, 2001

[105] D. Mason, B. McCane, and K. Novins, *Generating motion fields for the evaluation of optical flow algorithms on complex scenes*, Computer Graphics International (CGI), pp. 65–69, 1999

[106] N. Dalal, B. Triggs, and C. Schmid, *Human Detection Using Oriented Histograms of Flow and Appearance*, European Conference on Computer Vision and LNCS, SprinherLink, Vol. 3952, pp. 428-441, 2006

[107] N. Dalal and B. Triggs, *Histograms of Oriented Gradients for Human Detection*, Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), pp. 886–893, 2005

[108] A. Mitiche and A. Mansouri, *On Convergence of the Horn and Schunck Optical-Flow Estimation Method*, IEEE Transactions on Image Processing, Vol. 13, pp. 848-852, 2004

[109] A. Wedel, T. Pock, C. Zach, H. Bischof and D. Cremers, *An Improved Algorithm for TV-L^1 Optical Flow*, Visual Motion Analysis, LNCS, Springer-Verlag, Vol. 5064, pp. 23-45, 2009

[110] C. Zach, T. Pock, and H. Bischof, *A Duality Based Approach for Real time TV-L^1 Optical Flow*, Proceedings of the 29th DAGM conference on Pattern recognition, Springer-Verlag, LNCS, vol. 4713, pp. 214-223, 2007

[111] L. Rudin 1, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D 60: Nonlinear Phenomena, ELSEVIER, Vol. 60, pp. 259-268, 1992

[112] T. Brox, C. Bregler and J. Malik, *Large Displacement Optical Flow*, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp.41-48, 2009

[113] N. Ikizler, R. Cinbis and P. Duygulu, *Human Action Recognition with Line and Flow Histograms*, 19th IEEE International Conference on Pattern Recognition (ICPR), pp. 1-4, 2008

[114] D. Sun, S. Roth, J. Lewis, and M. Black, *Learning Optical Flow*, European Conference on Computer Vision (ECCV) and LNCS, Springer-Verlag, Vol. 5304, pp. 83-97, 2008

[115] S. Roth and M. Black, *Steerable Random Fields*, 11th IEEE International Conference on Computer Vision (ICCV), pp. 1-8, 2007

[116] M. Wainwright and E. Simoncelli, *Scale Mixtures of Gaussians and the Statistics of Natural Images*, Neural Information Processing Systems (NIPS), pp. 855–861, 1999

[117] S. Vedula, S. Baker, P. Rander, R. Collins and T. Kanade, *Three Dimensional Scene Flow*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, pp. 475-480, 2005

[118] S. Seitz and C. Dyer, *Photorealistic Scene Reconstruction by Voxel Coloring*, International Journal of Computer Vision, SpringerLink, vol. 35, pp. 151-173, 1999

[119] T. Brox and J. Malik, *Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, pp. 500-513, 2011

[120] R. Chaudhry, A. Ravichandran, G. Hager and R. Vidal, *Histograms of Oriented Optical Flow and Binet-Cauchy Kernels on Nonlinear Dynamical Systems for the Recognition of Human Actions*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, 2009

[121] E. Andrade, S. Blunsden and R. Fisher, *Hidden Markov Models for Optical Flow Analysis in Crowds*, 18th International Conference on Pattern Recognition (ICPR), Vol. 1, pp. 460-463, 2006

[122] M. Black and P. Anandan, *A framework for the Robust Estimation of Optical Flow*, 4th International Conference on Computer Vision, pp. 231–236, 1993

[123] J. Díaz, E. Ros, F. Pelayo, E. Ortigosa and S. Mota, *FPGA Based Real Time Optical Flow System*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 16, pp. 274-279, 2006

[124] B. Horn and B. Schunck, *Determining Optical Flow*, Artificial Intelligence, Vol. 17, pp. 185–203,1981

[125] B. Lucas and T. Kanade, An Iterative Image Registration Technique with an Application to Stereo Vision, 7th International Joint Conference on Artificial Intelligence, pp. 674–679, 1981

[126] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black and R. Szeliski, *A Database and Evaluation Methodology for Optical Flow*, International Journal of Computer Vision, SpringerLink, Vol. 92, pp. 1-31, 2007 & 2011

[127] Y. Benezeth, P. Jodoin, B. Emile, H. Laurent and C. Rosenberger, *Review and Evaluation of Commonly Implemented Background Subtraction Algorithms*, 19th International Conference on Pattern Recognition (ICPR), pp. 1-4, 2008

[128] Seth Benton, *Background Subtraction Implementations*, http://www.sethbenton.com/background_subtraction.html

[129] N. Howe and A. Deschamps, *Better Foreground Segmentation through Graph Cuts*, ARVIX (Cornell University Library), 2004

[130] D. Sun, S. Roth, and M. Black, *Secrets of Optical Flow Estimation and their Principles*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2432-2439, 2010

[131] A. Wedel, T. Pock, C. Zach, D. Cremers, and H. Bischof, *An improved algorithm for TV-L1 optical flow*, Dagstuhl Motion Workshop, 2008

[132] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear Total Variation based Noise Removal Algorithms*, Phys. D, Vol. 60(1-4), pp. 259–268, 1992

[133] M. Black and P. Anandan, *The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields*, Computer Vision and Image Understanding (CVIU), Vol. 63, pp. 75–104, 1996

[134] B. Horn, *Robot Vision*, MIT Press, 1986

[135] A. Blake and A. Zisserman, *Visual Reconstruction*, The MIT Press, Cambridge, Massachusetts, 1987

[136] A. Buades, B. Coll, and J. Morel, *A Non-Local Algorithm for Image De-noising*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, pp. 60-65, 2005

[137] G. Gilboa and S. Osher, *Nonlocal Operators with Applications to Image Processing*, Multiscale Model Simul., Vol. 7, pp. 1005–1028, 2008

[138] P. Sand and S. Teller, *Particle Video: Long-Range Motion Estimation using Point Trajectories*, International Journal of Computer Vision (IJCV), Springer-Link, Vol. 80, pp. 72–91, 2008

[139] K. Yoon and I. Kweon, *Adaptive Support-Weight Approach for Correspondence Search*, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 28, pp. 650–656, 2006

[140] Y. Li and S. Osher, *A New Median Formula with Applications to PDE based De-Noising,* Commun. Math. Sci., Vol. 7, pp. 741–753, 2009

[141] X. Ren, *Local Grouping for Optical Flow*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008

[142] F. Healey**,** S. Scobie**,** D. Oliver**,** A. Pryce**,** R Thomson**,** and B. Glampson, *Falls in English and Welsh hospitals: a national observational study based on retrospective analysis of 12 months of patient safety incident reports*, Quality and Safety in Health Care*,* vol. 17, pp. 424-430, 2007

[143] *Simple Strategies can Reduce Falls and Liability: Women and Elderly Fall more Frequently*, Rehab Continuum Report, 2004

[144] *Oldies, Depressed People more likely to take a Tumble*, Thaindian News, 18 June 2008

[145] L. Kowalczyk, *Spending on Health Care rises 7 Percent in Hospitals, Drug costs contribute to faster acceleration 12*, The Boston Globe, 8 Jan. 2002

[146] D.L. Gray-Miceli, *A Nursing Guide to the Prevention and Management of Falls in Geriatric Patients in Long-term Care Settings*, Medscape Today, 19 May 2005

[147] *Patient Alarm & Fall down Safety Alert*, [Survival Store], [cited 2009 Jan 21], Available at: HTTP: http://www.survivalstore.com/r6s15lbb4.html

[148] E. Auvinet, C. Rougier, J.Meunier, A. St-Arnaud and J. Rousseau, *Multiple cameras fall dataset*, Technical Report 1350, Montreal University, 2010

[149] A. Klaser, M. Marszałek, and C. Schmid, *A Spatio-Temporal Descriptor based on 3D Gradients*, British Machine Vision Conference (BMCV), 2008

[150] David G. Lowe, *Distinctive Image Features from Scale-Invariant key Points*, International Journal of Computer Vision (IJCV), Vol. 60, pp. 91-110, 2004

[151] http://lear.inrialpes.fr/software

[152] H. Wang, M. Ullah, A. Klaser, I. Laptev and C. Schmid, *Evaluation of Local Spatio-Temporal Features for Action Recognition*, British Machine Vision Conference (BMCV), 2009

[153] Ming-Kuei Hu, Visual *Pattern Recognition by Moment Invariants*, IRE Transaction on Information Theory, Vol. 8, pp. 179-187, 1962

[154] I. Laptev, M. Marszalek, C. Schmid and B. Rozenfeld, *Learning Realistic Human Actions from Movies*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, 2008

[155] C. Schuldt, I. Laptev, and B. Caputo, *Recognizing Human Actions: A local SVM Approach*, International Conference on Pattern Recognition (ICPR), 2004

[156] S. Siegel and N. Castellan, *Nonparametric Statistics for the Behavioural Sciences*, 2nd Ed., McGraw-Hill, 1988