A STUDY ON DETECTION
OF RISK FACTORS
OF A TODDLER'S FALL INJURIES
USING VISUAL DYNAMIC MOTION CUES

A thesis submitted for the degree of Doctor of Philosophy

by

Hana Na

School of Engineering and Design
Brunel University, Uxbridge, Middlesex, United Kingdom

February 2009

Abstract

The research in this thesis is intended to aid caregivers' supervision of toddlers to prevent accidental injuries, especially injuries due to falls in the home environment. There have been very few attempts to develop an automatic system to tackle young children's accidents despite the fact that they are particularly vulnerable to home accidents and a caregiver cannot give continuous supervision. Vision-based analysis methods have been developed to recognise toddlers' fall risk factors related to changes in their behaviour or environment.

First of all, suggestions to prevent fall events of young children at home were collected from well-known organisations for child safety. A large number of fall records of toddlers who had sought treatment at a hospital were analysed to identify a toddler's fall risk factors. The factors include clutter being a tripping or slipping hazard on the floor and a toddler moving around or climbing furniture or room structures.

The major technical problem in detecting the risk factors is to classify foreground objects into human and non-human, and novel approaches have been proposed for the classification. Unlike most existing studies, which focus on human appearance such as skin colour for human detection, the approaches addressed in this thesis use cues related to dynamic motions. The first cue is based on the fact that there is relative motion between human body parts while typical indoor clutter does not have such parts with diverse motions. In addition, other motion cues are employed to differentiate a human from a pet since a pet also moves its parts diversely. They are angle changes of ellipse fitted to each object and history of its actual heights to capture the various posture changes and different body size of pets. The methods work well as long as foreground regions are correctly segmented.

Acknowledgements

I would like to thank the following people:

- My supervisors, Dr Shengfeng Qin and Professor David Wright for direction and support;
- The school of Engineering and Design at Brunel University for sponsoring the research project with a scholarship;
- Dr Usman Khan and Mr Basil Badi for helping with written English;
- My colleagues, Ms Yuan Yin, Mr Hui Yu, and Mr Zhao Yue for support.

Dedication

I dedicate this thesis and its contents to my family, Mum, Dad, and my two sisters.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AE: Accident & Emergency, 64

CAPT: Child Accident Prevention Trust, 13, 74, 143, 147

CDC: The Centers for Disease Control and Prevention, 15, 16, 148

DLT: direct linear transformation, 51, 54

DOFs: degrees of freedom, 28

DTI: Department of Trade and Industry, 154

ECSA: European Child Safety Alliance, 14, 15, 74, 147

GPS: Global Positioning System, 4

HASS: Home Accident Surveillance System, 64

HSC: Home Safety Council, 148

IAC: image of the absolute conic, 56, 57

KidsafeNSW: Kidsafe New South Wales, 147

KidsafeQLD: Kidsafe Queensland, 147

KidsafeWA: Kidsafe Western Australia, 143, 147

MORT: Management Oversight and Risk Tree, 66

MUARC: Monash University Accident Research Centre, 15, 74, 147

NAMIC: National Association of Mutual Insurance Companies, 148

NSC: National Safety Council, 74

OECD: Organisation for Economic Co-operation and Development, 154

PCA: principal components analysis, 21

ROI: region of interest, 21, 84, 85, 86, 87, 89, 117

RoSPA: Royal Society for the Prevention of Accidents, 3, 64, 65, 67, 74, 154, 159, 161, 163, 165

SafekidsNZ: Safekids New Zealand, 147

SVM: support vector machine, 5, 31

VISU: Victorian Injury Surveillance Unit, 15

Chapter 1
Introduction

1.1 Unintentional Accidents to Children

Young children are not able to assess risks for themselves. They also have poor coordination and balance and need to touch and explore to learn about the world around them. These factors all mean that children are particularly vulnerable to accidents [1]. The actual data tell the same story. For example, on average over two million children per year in the United Kingdom are taken to hospital after having an accident [2], and approximately two hundred children per day are hospitalised and one child dies as a result of unintentional injuries in Australia [3]. Around half of these accidents happen at home and young children aged under five years are most vulnerable to injuries in the home environment where they spend most of their time [4].

Unintentional or accidental injury means injury occurring as a result of an unplanned and unexpected event, which arises at a specific time from an external cause [5]. The five major causes of injuries in the home are falls, burns and scalds, poisoning, drowning, and animal injuries [3]. Falls account for over 40 per cent of all home accidental injuries of children [1].

1.2 Research Problems

Physical injury is the main cause of death and a major cause of ill-health and disability in childhood [5]. As young children are unable to assess risks, the best way to prevent their accidental injuries is continuous supervision and instruction by their parents. This, however, is not always practical or possible. An automatic risk detection approach is proposed here to assist parents' supervision for prevention of accidental injuries to young children.

Young children aged under five are most vulnerable to accidental injuries in the home environment, and falls are the main cause of their injuries. Figure 1 shows that more than half of fall accidents happen in indoor home areas, mainly in the lounge, study, living/dining/play area and inside stairs. Figure 2 shows that children at the ages of one and two are particularly exposed to many fall accidents. This is because children of this age, who are toddlers, learn to walk, climb, and begin to run but are not so good at stopping or swerving to avoid people or objects [6]. In the light of these data, this research specifically targets detection of a toddler's fall risks in an indoor home environment in order to help parents supervise the toddler when they temporarily withdraw their attention.

The general activity of a toddler, who spends time awake at home, would be to play with a parent or a toy. As the fall risk detection aims to aid the supervision of parents, who cannot watch their child all the time, the usual scenario for a toddler is to play with toys on the floor of a playroom. In the case of a family with a pet, the toddler could play with the pet, or the pet could stay around the toddler. Therefore, the main research problems in detecting a toddler's fall risks are as follows:

- Identification of key risk factors related to a toddler's fall injury in the home environment;
- Human detection: to separate a toddler from other clutter such as toy or pet in order to focus on the toddler;
- Status recognition: to collect information related to the toddler useful in recognising risks of injuries from falling such as position, movement, and relative state to other clutter and environments.

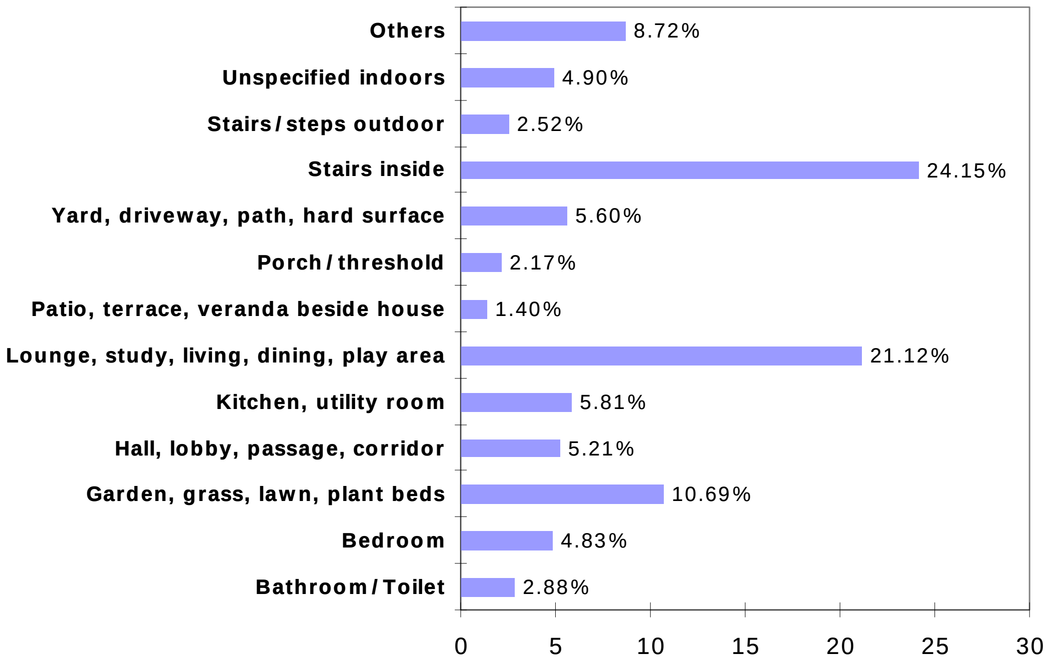


Figure 1. Falls of Children aged up to Five by Location within Home in 2002[1]

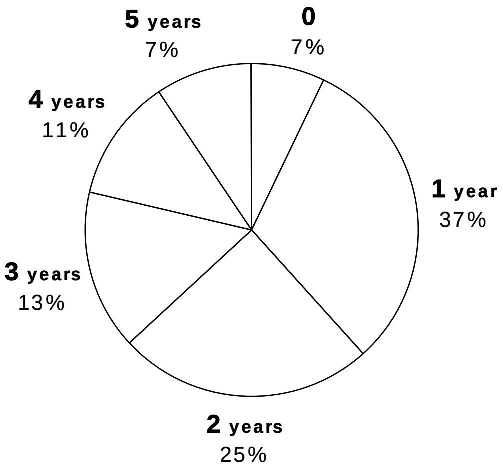


Figure 2. Falls at Home by Age in 2002[1]

[1] A total of 6353 fall accidents at home of young children treated in a hospital were analysed. The details of the accidents were obtained by email from the Royal Society for Prevention of Accidents (RoSPA) after a personal request to its information centre.

First, specific risk factors need to be identified for automatic detection. In finding the identified factors, detection of subject toddlers and recognition of their status are major technical challenges.

1.3 Available Technologies

In relation to the latter two research problems introduced as technical challenges in Section 1.2, existing technical solutions are reviewed in terms of advantages and drawbacks to find a suitable method to solve the problems.

1.3.1 Human Detection

A. Wearable Device

The simplest and most reliable way to distinguish a human from other objects with similar characteristics would be the use of wearable devices. When a target human wears a wireless transmitter, a sensor can be used to detect the transmission within the activity area of the human. If multiple people individually wear a transmitter emitting a unique signal, different people can be recognised separately.

For example, Active Badge [7] is a small device which is worn on office personnel and transmits a unique infra-red signal every ten seconds to locate individuals within a building. Each office within the building is equipped with one or more networked sensors to detect these transmissions. The small low-cost infra-red emitters enable the detection of a large number of personnel. For outdoor locating, Global Positioning System (GPS) technology has been widely used owing to its precise microwave signals. Smart Sight [8] is an intelligent tourist system intended to break the language barrier and provide navigation assistance. The assistance is based on the tourist location derived from a GPS receiver and speech and gesture input from a microphone and a camera. The major drawback of this wearable technology is that the detection requires the user to wear the device all the time.

B. Vision-Based Analysis

One of the well-known methods to free a target human from continuously wearing a device is to analyse images captured by a camera in consequence of the decreasing rate of price to performance of computing for image processing [9]. The human detection in images can be based on motion, appearance, or shape [10].

Sidenbladh [11] exploited motion information of walking humans by calculating optical flow for image windows and using a support vector machine[2] (SVM) to detect the walking motion in videos. Zhao and Nevatia [13] used a colour histogram, which is a kind of appearance model that detects humans by measuring the similarity of foreground pixels to the model. As an instance of the shape-based human segmentation, Leibe and colleagues [14] stored outlines of walking humans as a number of templates and matched them with the foreground edges of input image over different scales.

As in the above methods, diverse characteristics of humans can be applied to differentiate humans from other foreground objects in images. The results sometimes can have problems by detecting non-human objects looking like a human or by some original weakness of image processing such as sensitivity to light changes.

[2] Support vector machines are learning systems for classification and regression that use a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from optimisation theory that implements a learning bias derived from statistical learning theory [12].

1.3.2 Status Recognition

A. Wearable Device

The kinds of sensors used on a target human depend on the types of human status to be recognised. As the work in this thesis relates to fall accidents, this section reviews automatic fall detection approaches for the elderly. The approaches have been actively proposed because early detection of falls, which are a major health hazard for the elderly [15], is an important step in preventing serious injuries and deaths and other obstacles to their independent living [16].

Wearable devices for fall detection generally integrate accelerometers and/or tilt sensors and a wireless interface [17]. This is because such devices are inexpensive and fairly reliable in fall detection and can easily be embedded into existing community alarm and response systems [16]. The algorithm of Hwang and colleagues [18] distinguishes between falling and daily life activity based on signals from an accelerometer, a tilt sensor, and a gyroscope worn on the chest. The small device of Chen and colleagues [19] is worn on the waist and composed of two dual-axis accelerometers for detecting a fall and radio signal strength for locating the victim.

Although wearable devices have such advantages, the proper functioning of the recognition relies on correct as well as continuous wearing of the device [17]. When people wearing the accelerometer-based device sit down quickly or drop the device by accident, it will be recognised as a fall.

B. Sensors in Environment

Sensors can be embedded in the environment instead of being worn by a target human in order to obtain some information about the human. The systems of Tamura and colleagues [20] and Harada and colleagues [21] recognise an infant's posture and body movement respectively using 16 temperature sensors and 384 pressure sensors distributed in the bed in order to prevent infants' sudden cot death. Alwan and colleagues [22] detected human falls by monitoring the vibration patterns from a piezoelectric sensor coupled to the floor surface.

Apart from image sensors, the information which sensors installed in the environment can sense with respect to a human, is considerably limited. This is because the information can be collected when the human has a physical effect on the environment such as body heat and pressure or speaks out in the case of sound sensors. Moreover, a massive number of sensors may need to be distributed to cover all the areas necessary to recognise the human's status.

C. Vision-Based Analysis

Vision-based human motion analysis has attracted great interest owing to its promising application in many areas such as visual surveillance, perceptual user interface, content-based image storage and retrieval, video conferencing, athletic performance analysis, virtual reality, and so on [23]. In general, the stage of human activity recognition in vision-based human motion analysis focuses explicitly on human activities and the interactions between humans by dealing with the entire human body for holistic information or body parts for more subtle actions [10]. For instance, Sixsmith and Johnson [24] analysed InfraRed thermal images of a human body to detect falls of the elderly. For automatic gait recognition, Foster and colleagues [25] used a time-varying signal from a sequence of silhouettes of a walking subject, and Wang and colleagues [26] used joint-angle trajectories of lower limbs. Simple human activities can be recognised by observation of the motion of objects without knowledge of their bodies. The algorithms of Stauffer and Grimson [27] detect unusual situations based on features such as position, speed, direction, and size.

There are a large number of different approaches based on image analysis, and they depend on the goal of the researcher and their applications for human activity recognition [10]. Image analysis technology has considerable potential for recognising diverse and complex human activities, but this research area contains a number of difficult and often ill-posed problems such as inferring the pose and motion of a highly articulated and self-occluding, non-rigid, 3D object from images. This complexity makes the research area challenging [10].

1.3.3 Summary of Available Technologies

This section reviewed general solutions to the technical challenges related to human detection and status recognition. For human detection, a device emitting a signal can be worn by each subject, or images captured by a camera watching the subjects can be analysed. The use of a wearable device produces fairly reliable results but requires each subject to wear the device all the time. Image analysis can free the subjects from this obligation, but its results are relatively error-prone. For status recognition, a sensor can be attached to each subject or installed in the subject's environment. Those sensors can provide reliable information, but the information is fairly limited. In the meantime, images of subjects can provide multiple cues to recognise various situations of the subjects.

1.4 Research Aims

Toddlers are most vulnerable to fall injuries in an indoor home environment, but there has been no study on automatic detection of children's fall risks. On the other hand, many efforts have been made to tackle the hazard of unintentional falls for elderly people by detecting their falls in the computer recognition area [16-19, 22, 24, 28-42]. This is because the population is ageing and late detection of fall accidents can be fatal to elderly people living independently. Most of the existing approaches to elderly fall detection only focus on detecting fall events without doing anything about prevention. Although some of them collect the fall data to evaluate the subject's personal fall risks for later prevention, there is no prevention against falls during the data collection and also against irregular falls afterwards. Some wearable devices provide prompt protection such as an airbag [43] and an overhead tether [44] when sensing a fall, but they require the user to wear the protection all the time.

Thus, the work in this thesis proposes an automatic approach to recognition of risk factors of a toddler's fall injury in the home environment. The methods could be used to give a nearby caregiver an alert to eliminate the factors before a fall happens. This is different from the studies conducted previously, which focus on detecting fall events and are specifically tailored towards elderly people.

Although vision-based analysis is not as reliable as sensors attached to a human body for estimating human status, it is attractive because it does not require a subject to wear a device constantly, especially important in the case of young children, who would not accept such a continuous restraint. Besides, vision-based analysis can cope with various problems in terms of human detection and status recognition, which are the technical problems of this research presented in Section 1.2. Therefore, this research aims to investigate and develop image analysis methods to recognise risk factors of a toddler's fall injuries in an indoor home environment in real time.

The main objectives of this research are:

- Identification of key factors causing fall injuries;
- Investigation and development of image analysis methods for differentiating a toddler from other foreground objects;
- Investigation and development of image analysis methods for recognising the key risk factors by collecting information on the toddler and the environment.

First of all, fall-prone situations of toddlers are investigated to identify the key factors causing fall injuries. As the recognition of fall risk factors is expected to assist caregivers' supervision to toddlers, this research focuses on the factors which need to be continuously watched rather than those which can be simply prevented by installation of a safety product. Those risk factors can be generated from toddlers' behaviours such as climbing or from their environment such as spills, as illustrated in Figure 3.

(a) (b)

Figure 3. Fall-Prone Situations by (a) Behaviour [45] and (b) Environment [46]

For automatic detection of the identified risk factors, existing image analysis methods need to be studied, and novel methods are developed and tested with respect to separation of a toddler from other foreground objects and collection of information on the toddler and the surroundings. This is based on the assumption that the target toddler is left alone on the scene but a caregiver stays nearby, e.g. next room. As the proposed recognition system is meant to detect risk factors from real-time images, the image analysis methods should not require too much computation to be processed in real time.

1.5 Thesis Layout

Chapter 2 reviews all the literature related to the work for achieving the objectives of this research. First, organisational suggestions to prevent young children's falls are scanned to identify risk factors of a toddler's fall injury. Then previous work based on image analysis is studied regarding detection of the fall risk factors such as human detection, handling of regional merges and splits, animal classification, and camera calibration.

Chapter 3 analyses more than two thousand fall records of toddlers at home, who were treated in hospital, and identifies the risk factors of a toddler's fall injury in the home environment. The organisational suggestions and the real fall stories are filtered to extract fall-injury-prone situations, generated by behavioural or environmental changes since this research aims to aid a caregiver's continuous supervision of those changes.

As the work presented in this thesis proposes novel approaches to separating humans from other foreground figures for risk factor detection, Chapter 4 and Chapter 5 focus on human detection. Clutter classified as the foreground is dealt with and differentiated from humans in Chapter 4 because it can become a tripping or slipping hazard. The dynamic movements of human body parts when walking are exploited for the classification of a human and clutter, and this is different from most previous studies, which use appearance information for human detection. The method is, however, not good enough to discriminate humans from pets, which also move their body parts diversely, and therefore other cues are adopted for this discrimination in Chapter 5.

The last chapter clarifies this research's contributions, limitations, and future work for improved results.

Chapter 2
Literature Review

2.1 Introduction

Published literature is reviewed in this chapter with respect to the research objectives identified in Section 1.4. The objectives are identification of key factors causing fall injuries, and investigation and development of image analysis methods of finding a toddler and collecting information on the toddler or the environment to detect the key risk factors.

For the identification of fall risk factors, suggestions for preventing young children's falls were collected from organisations expert in child safety and these are reviewed in Section 2.2. These suggestions are based on knowledge of the potential for fall accidents, which is believed to help reduce the risk of serious injury [1]. Section 2.3 studies existing image analysis methods for human detection, handling of regional merges and splits, animal classification, and camera calibration since they are related to the toddler classification and information collection in images.

2.2 Suggestions to Prevent Children's Falls

There are several organisations which aim to promote changes in attitudes, behaviours, laws, and the environment to prevent accidental injury to children by providing information, advice, resources, and training [47]. There are also research centres specialised in the study of injury and injury prevention. They provide the public with information on how children generally get injured by falling in the home environment and what should be done to prevent the fall injury. The information about fall injury prevention can be largely divided into two categories: *single modification* and *supervision-based action*. *Single modification* is what needs to be done at a single time to reduce children's fall risks such as installation of a safety product or modification of an environment according to safety standards. *Supervision-based action* is what to do when a change which may cause a children's fall is observed such as clearing clutter when it is found to have potential for tripping. The suggestions for fall injury prevention are detailed by the organisations and the two categories as follows.

2.2.1 Child Accident Prevention Trust

The Child Accident Prevention Trust (CAPT) is a UK national charity committed to reducing the number of children and young people, who are killed, disabled, or seriously injured as a result of accidents. CAPT offers several factsheets and sample copies of leaflets on children's accident risks based on expert knowledge of child accident prevention and practical experience of supporting parents and practitioners; all are available through its official website [48].

For *single modification* to prevent fall injuries at home, CAPT suggests fitting safety gates to the top and the bottom of stairs and safety catches on upstairs windows. CAPT also advises keeping furniture away from windows and checking balcony railings to make sure that children cannot climb on them or fall through gaps. For *supervision-based action*, CAPT recommends parents to wipe up spills as soon as they happen to avoid slipping and to encourage children to put their toys away after use to avoid trips [1, 49, 50].

2.2.2 Safe Kids

Safe Kids Worldwide is a global network of organisations whose mission is to prevent accidental childhood injury, a leading killer of children aged fourteen and under. More than 450 coalitions in 16 countries bring together health and safety experts, educators, corporations, foundations, governments, and volunteers to educate and protect families [51]. This section refers to child injury prevention factsheets and recommendations from Australia, New Zealand, and United Arab Emirates members.

Regarding *single modification* to prevent serious injuries relating to falls, Kidsafe, the Australian member of Safe Kids, advises that safety-gates are an essential preventative choice in areas such as stairs, hallways, and entrances and that the vertical gaps between slats of balustrades should be designed to standard recommendations lest children get trapped [52]. Kidsafe also indicates the use of nursery furniture which complies with Australian standards, straps on high chairs, swings or strollers, a window-guard in front of children's windows, a soft surface where children are learning to walk, and corner bumpers covering sharp corners [53, 54]. Safe Kids United Arab Emirates suggests keeping stairs well-lit and furniture away from windows for other *single modification*s. For *supervision-based action*, Safe Kids Worldwide suggests getting rid of hazards such as folded carpets, electric wires, toys, shoes, and clothing on the floor, and Safekids New Zealand advises discouraging children from climbing on furniture such as chairs and couches [55, 56].

2.2.3 European Child Safety Alliance

The European Child Safety Alliance (ECSA) as a branch of EuroSafe, the European Association for Injury Prevention and Safety Promotion, undertakes activities including research studies on child injury issues, publication of reports, and recommendations that ultimately could enhance the quality of children's lives in Europe [57].

The *single modification* tips from ECSA are to use straps when putting the child into a high chair, swing, changing-table, or stroller and shatter-resistant film on glass surfaces where children could fall. The tips also include the use of a soft carpet beside a child's cot or bed in case the child falls out and corner covers on furniture with sharp corners. ECSA guides installation of approved stair-gates with vertical bars at four-inch intervals at the top and the bottom of stairs and child-resistant window-guards throughout the home. ECSA's *supervision-based action* tip is to remove tripping hazards on the floor and stairs [58].

2.2.4 Monash University Accident Research Centre

Monash University Accident Research Centre (MUARC) is the world's leading injury prevention research centre, which identifies emerging injury problems, determines and evaluates safety strategies, and advises on policies to bring about reductions in injury-related harm [59]. MUARC runs the Victorian Injury Surveillance Unit (VISU), a fundamental resource for injury prevention, and analyses data to identify injury issues, monitor trends, and develop potential countermeasures to injury. VISU also produces the influential *Hazard* publication to provide up-to-date information on current and emerging injury issues and prevention strategies [60]. This section refers to the fall-related editions of *Hazard*.

The recommendations regarding *single modification* from MUARC are installation of stair-gates and window-guards as well as slip-resistant surfaces and impact-absorbing surfaces throughout the home. The recommendations also cover improvements in the design and construction of balconies, stairs, and steps in terms of geometry and visibility with functional handrails provided [61]. The recommendation regarding *supervision-based action* is to discourage children from climbing and playing on furniture as household furniture items are strongly represented in falls of children under five [45, 62].

2.2.5 Others

The Centers for Disease Control and Prevention (CDC) promote health and quality of life by preventing and controlling disease, injury, and disability [63] in the United States of America. CDC presents a document about falls in the home environment in the National Ag Safety Database, which provides information about agriculture-related safety.

The suggestions in the document regarding *single modification* are to tape the edges of rugs down to keep children from skidding and not to stretch electrical cords across rooms. CDC also advises keeping stairs and steps well-lit with sturdy handrails and arranging furniture so that traffic patterns within rooms are as straight and wide as possible. The CDC recommendation particularly for children is installation of window-guards and gates at the top and the bottom of stairways. The suggestions regarding *supervision-based action* are to wipe up spills immediately, to keep floors clear of clutter, and to teach children to pick up their toys after use and never run through the house [64].

Other organisations suggest keeping furniture away from windows and securing children in a restraint system when using a high chair for *single modification*, and they recommend parents to decrease children's climbing temptation for *supervision-based action* [65, 66].

2.2.6 Summary of Suggestions to Prevent Children's Fall Injury

A summary of the organisational suggestions to prevent children's fall in the home environment is shown in Table 1. The suggestions categorised into *supervision-based action* are referred to in Chapter 3 for the identification of fall risk factors requiring continuous supervision for prevention.

Table 1. Summary of Suggestions to Prevent Children's Falls

Organisation	Single Modification	Supervision-based Action
Child Accident Prevention Trust	- Instal stair-gates - Instal window-guards - Put furniture away from windows - Design safe banister	- Wipe spills - Put toys away after use
Safe Kids	- Instal stair/hallway/entrance gates - Design narrow balustrade slat - Use standard nursery furniture - Use a strap in a high chair - Instal window guards - Use a soft surface to walk - Place corner bumpers - Keep stairs well-lit - Put furniture away from windows	- Remove tripping hazards - Discourage climbing furniture
European Child Safety Alliance	- Use straps in a high chair - Use shatter-resistant film on glass - Place a soft carpet beside a bed - Cover sharp corners - Instal stair-gates	- Remove tripping hazards
Monash University Accident Research Centre	- Instal slip-resistant surfaces - Instal impact-absorbing surfaces - Design safe balconies - Design safe stairs - Instal stair-gates - Instal window-guards	- Discourage climbing furniture - Discourage playing on furniture
Others	- Tape edges of rugs down - No electrical cords across rooms - Keep stairs well-lit - Instal sturdy handrails on stairs - Arrange furniture for good navigation - Instal window-guards - Instal stair-gates - Put furniture away from windows	- Wipe spills - Clear clutter on the floor - Decrease climbing temptation

2.3 Related Image Analysis Methods

In order to develop methods for automatic detection of the fall risk factors, which are identified in Chapter 3, related image analysis methods are reviewed in this section. As the fall risk factors are related to a toddler's behaviour and environment, the automatic detection requires to classify a toddler and other foreground objects and collect information related to them in images. First, diverse methods of human detection are studied in Section 2.3.1 for the toddler classification. Since the detected toddler and foreground objects need to be individually tracked to collect useful information, Section 2.3.2 reviews published literature related to handling of regional merges and splits, which is a typical problem of tracking. Considering pets in the toddler classification, existing methods of animal classification are investigated in Section 2.3.3, and camera calibration methods are studied in Section 2.3.4 for use of 3D world information.

2.3.1 Human Detection

Nearly every vision-based analysis of human motion starts with human detection, which aims at segmenting regions corresponding to people from the rest of an image. Human detection is significant since the subsequent processes in vision-based analysis of human motion such as tracking and action recognition, are greatly dependent on it. This process usually involves foreground segmentation and object classification [23], but some approaches do not have the stage of classifying objects or discarding non-human objects after the segmentation. Instead, they go straight to the next process such as tracking with the assumption that all the noteworthy segmented foreground regions correspond to humans. This section only reviews the studies attempting to classify human and non-human objects since the objectives of this research include the detection of not only a toddler but also other foreground objects such as toys and pets.

The studies should not use any markers on the body of the subjects since one of the main reasons why the work in this thesis uses image analysis technology is to free the subject toddler from wearing a device continuously. The whole process of reviewed human detection methods should be automated, but partial manual work is allowable for the initialisation, which is normally required at a single time when the camera in use is installed such as camera calibration.

The reviewed work is limited to that published within the last five years from 2004 to 2008, as there is a large amount of literature related to human detection. The review also excludes approaches to analysing images where a small part of a human such as a face or hands is dominant and the detection only targets the part. This is because the image to be analysed in this work needs to cover a space where the subject toddler spends time for detection of fall-prone situations, and a small part of the toddler could not be captured in such a high resolution.

Hence, the literature to be reviewed with respect to human detection in images was selected according to the following principles:

- Markerless image analysis methods;
- Fully automated human detection methods;
- Publications between 2004 and 2008;
- Studies, which include classification of human and non-human foreground objects;
- Approaches to detecting a human in images where more than half of the human body appears.

The review is categorised by the kinds of cues used for human classification: geometric knowledge, skin colour, 3D human model, sample training, motion information, and others.

A. Geometric Knowledge

Although the human body size varies accordingly to genes and environmental factors, normal and adult people have constantly structured and proportioned body parts. The human body structure and proportion are precisely different from any other object in the world and can be used to detect humans in images.

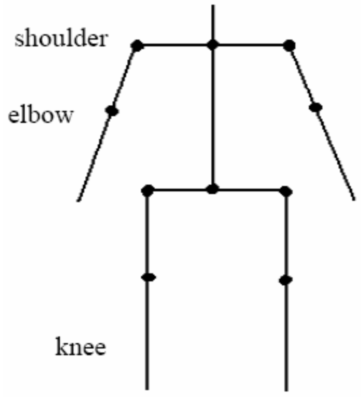


Figure 4. Standard Human Skeleton Model [67]

The method of Fan and Wang [67] starts with detection of differences between two successive images for motion segmentation, assuming that the background is motionless. Then a dummy skeleton consisting of elongated parts such as an arm and a thigh and joints such as a shoulder and a coxa, is extracted from the silhouette of each segmented region. The conspicuous joints of the extracted skeleton are examined and confirmed on the basis of several rules given by the standard human skeleton model shown in Figure 4. The dummy skeleton is compared with the standard skeleton to determine the positions of inconspicuous joints such as elbow and knee joints. As the pre-defined scales of the lower and upper limbs influence locating of the inconspicuous joints, wrong scales in some cases cause errors, affecting the accuracy of the result. The skeleton is useful not only to detect the human body but also to estimate its pose, but the results are error-prone when some joints are hidden or kept out of the image.

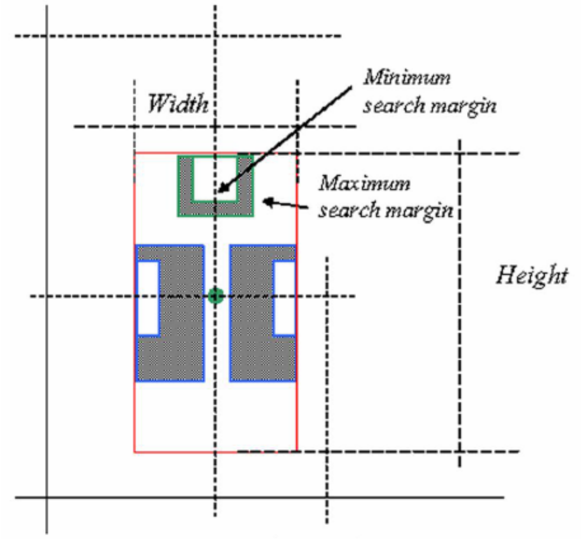


Figure 5. Human Component Search Margins [68]

In order to discard any moving object which is not a human being, Schleicher and colleagues [68] applied a principal components analysis[3] (PCA) algorithm in a hierarchical manner after background subtraction. Owing to the many different orientations and poses of a human body, it is difficult to apply any kind of pattern recognition method directly to the full image. Hence, a PCA method is only applied to some specific regions within the region of interest (ROI) of each detected object. Those regions are located where the main human components, the left and right arms and the head, are supposed to be found, and variable margins are defined to look for them, as shown in Figure 5. This is because the components can be placed in different positions with respect to the main body and can have different sizes. In order to recognise each body component within the correct region, PCA is trained with samples of arms and heads in different poses and orientations. Sample training is another famous method to detect a human and Section D reviews studies which chiefly use a sample training method. Although a

[3] Principal component analysis is a statistical technique that linearly transforms an original set of variables into a substantially smaller set of uncorrelated variables that represents most of the information in the original set of variables [69].

human in frontal views is successfully detected at a very high rate, lower success rates are obtained in back views and lateral views, and long-term occlusions cause a fairly high rate of failed detections.

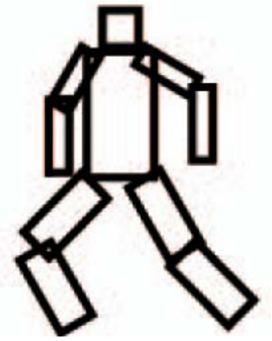


Figure 6. Human Body Pictorial Structure [70]

Ramanan and colleagues [70] present two methods of human detection. One is a *bottom-up* approach that looks for candidate body parts based on edges and motion, the candidates being clustered to find assemblies of parts, which might be people. Each torso cluster is interpreted as a unique person, and one temporal pictorial structure like the one in Figure 6 is instantiated for each cluster to confirm limbs. As this method localises the torso first and then finds the remaining limbs, it has problems in estimating the limbs when the torso localisation is poor. The other method is a *top-down* approach that looks for an entire person in typical poses, assuming that people tend to occupy certain key poses. Given an edge image, a tree pictorial structure in a pose, like the one in Figure 6, is searched for, using rectangle chamfer template costs to construct limb likelihoods. Global constraints are also enforced for human detection such as similar appearances of left and right legs. The *top-down* method works better than the *bottom-up* method, but the subject people need to behave predictably or to be observed for a long time owing to the detection of certain key poses in the *top-down* method.

In order to detect pedestrians near inner-city bus stops, Bird and colleagues [71] used very simple geometric knowledge compared with the above studies. First, the background is subtracted, and blobs are extracted by identifying regions of connected foreground pixels. Then the shape of each blob on the scene is examined at every frame to determine if the blob corresponds to a pedestrian, using heuristics as follows:

- The blob's height is greater than its width;
- The blob is not within ten pixels of the edge of the image;
- The blob has not merged within the past five frames;
- The blob's real world height is between 60 and 80 inches.

For the last heuristic check related to the real world height, the camera in use is calibrated, with the angular layout of most bus stops and the sidewalks around them being used as the required geometric primitives for camera calibration.

Fihl and colleagues [72] employed simple characteristics of human silhouettes. After an image is processed by background subtraction and noise is removed with a median filter, split blobs of a single human owing to some error in the background subtraction are merged into bounding boxes, each representing one human. This merging is done by investigating the size and proximity of the bounding box of each blob. Since the silhouette of a person can roughly be described by an ellipse, the silhouettes in the merged bounding boxes are compared with a simple body model based on an ellipse. This body model defines limits for the ratio between the major and minor axes of the ellipse, the slope of the major axis, the fidelity between the ellipse and silhouette, and the area of the silhouette.

B. Skin Colour

Although the colour of a human body in images from a surveillance camera cannot be consistent because of clothing, a face or hands are not usually covered by clothing and constantly appear in a skin colour. Given this fact, many approaches have tried to search the skin colour to detect humans within an image. As it is difficult to distinguish between a face and hands or between different people only using the skin colour cue, it is common practice to combine it with other human cues such as geometric constraints.

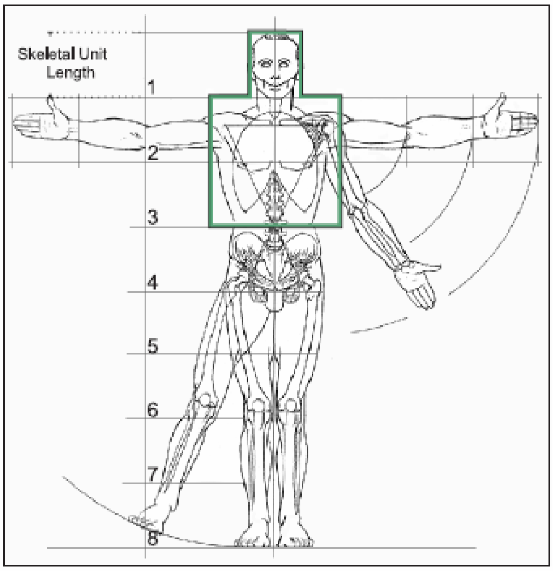


Figure 7. Torso Primitive on Vitruvian Man [73]

Micilotta [73] proposed two different methods of detecting humans, both involving a skin colour and eventually estimating poses in cluttered scenes. The first method uses the torso primitive, shown in Figure 7, to detect a human torso within background-segmented regions and segments skin colour regions, which belong to a face and hands, based on a skin model. The skin model is built with the colour information of the facial region estimated from a detected torso. The second method searches for four body parts, which are a face, a torso, legs, and hands, using sample-trained body parts detectors, and assembles the body parts, which have been correctly detected, using coarse heuristics. Skin colour cues are exploited here to reduce false detections of the body parts. The skin colour models would be naturally susceptible to background clutter like wooden furniture and could be unreliable when the subjects hold their hands behind their backs. The entire process of the second method is a little too slow for real-time operation.

Yang and colleagues [74] fused depth, colour, and motion features for detection of multiple people. The depth and motion features are used to segment moving foreground objects from the background, despite the changes in illumination or camera movement, and also to separate two people appearing close together in a projected image but being at different depths in the physical space. The colour feature concentrates on human face detection to confirm whether the segmented region corresponds to a human. As several body parts like a face, hands, arms, or legs generally appear in the skin colour, only skin colour regions with enough lip colour pixels at the proper position are kept as face region candidates. The major limitation of this system is its low speed owing to the multi-modal fusion.

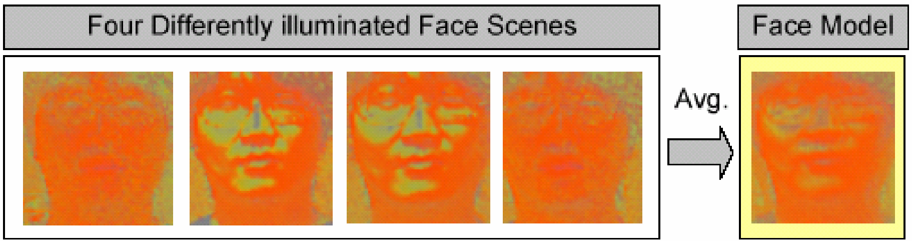


Figure 8. Face Colour Model [76]

There are other ways of analysing colour information for face detection. After finding skin-coloured blobs, Pszczolkowski and Soto [75] applied three criteria to each RGB channel of the blobs to classify face and non-face. The criteria are normalised standard deviation, contrast, and uniformity or energy. Choi and colleagues [76] exploited a face colour model, produced from four differently illuminated scenes of one person, as shown in Figure 8. The red and green colour histogram of the face model is normalised, and the quantisation of the spatial colour histogram enables identification of different faces as well as differentiation of a face from other regions in the skin colour. From the scale of the detected face, the rest body area is estimated.

Some studies combine skin colour and geometric information in order to separate a face from other body parts appearing in a skin colour. Ammouri and Bilodeau [77] monitored medication intake by detecting and tracking a face and hands in images, where the upper body of a single human is visible. After detection of skin-coloured regions, a face region is determined among those regions according to two shape-related rules. The first rule is that the ratio between the width and the height of the region is smaller than 2.25, and this rejects regions that are too wide and narrow, like a forearm. The second one is that the ratio between the surface and the square of the perimeter of the region is larger than 0.02 so as to measure the circularity and find elliptical shape regions. These rules are just good enough to distinguish between a face and arms.

For gesture recognition, Kang and colleagues [78] detected skin blobs from the difference image between successive frames and judged the face and the hands on the basis of size, position and distance of each blob. For a high detection rate, an additional face detector is used to classify regions into faces and non-faces by sample training. Medioni and colleagues [79] applied more complex information in addition to skin colour detection for detection of multiple people. Heads are first detected using image intensity, skin coloured pixel detection, and elliptical edge detection and initialise a 2D articulated upper body model for locating arms and estimating body poses.

C. 3D Human Model

As a human image is a projection of a human body, which is a 3D figure, into a 2D space, the human in the image can look very different depending on the position or the angle of the camera as well as the human pose. Therefore, 3D human information has been adopted in some work to detect a human and further track the motion more accurately.

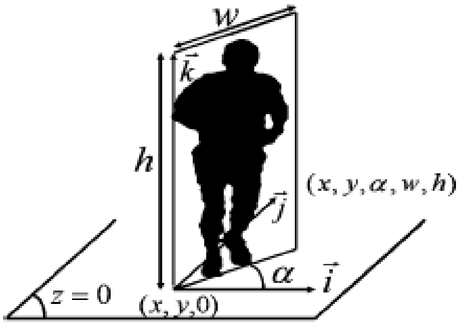


Figure 9. Human Geometric Model [80]

Pham and colleagues [80] proposed a multi-camera system to detect emergency situations and analyse long-term activities which enforce medical follow-up. Silhouettes are extracted by subtracting the background in all camera views, and projections of the 3D human model in Figure 9 are fitted to the detected silhouettes. The human geometric model, a planar rectangle oriented in space, is represented by a 5D vector (x, y, α, w, h), where the couple (x,y) and α respectively indicate the 3D position on the ground and the orientation angle of the model in relation to a world reference and w and h stand for the width and the height of the rectangle. As their aim is to recognise basic actions such as standing, sitting, and lying down, simple rules based on the ratio of the width and the height of the 3D rectangle, produce promising results.

For human posture tracking and classification, Pellegrini and Iocchi [81] exploited a stereo vision camera, the 3D data from which are matched with a 3D human body model. After background subtraction is performed by consideration of intensity and disparity components, small foreground blobs owing to noise are removed. Then each pixel belonging to the extracted blobs is projected in the plan-view, and the foreground segmentation is refined by detection of connected components in the plan-view space. This is possible since the stereo camera in use is calibrated when it is installed. The segmented data are matched with a 3D model for person posture recognition. Since the posture recognition tries to distinguish between the principal postures such as up, sit, bent, on knee, and lying, the 3D model is composed of a head-torso block and a leg block without taking into account arms and hands. As shown in Figure 10, the head-torso block is formed by a set of 700 3D points that represent a 3D surface, and the legs are unified in one articulated body. This model must be adapted to the person being analysed and can be built by assuming that the legs are always in contact with the floor and the ratio is constant between the dimensions in the model parts and the height of the person.

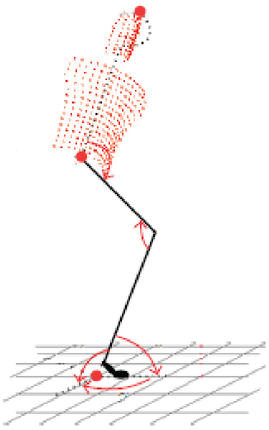


Figure 10. 3D Human Model [81]

Kehl and Gool [82] applied a more complicated 3D human model, covering all the articulated body parts significant in recognition of detailed body poses. Their algorithms start with segmenting the foreground by subtracting the background in each of five camera views and extracting colour edges from the five foreground masks for fast reconstruction of the 3D surface. The extracted data are fitted to the body model in Figure 11, built from superellipsoids and driven by a human skeleton. The skeleton consists of ten joints with a total of 24 degrees of freedom (DOFs), and the superellispoids are a special case of superquadrics, offering a better approximation for complex body parts.

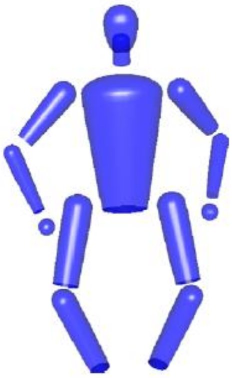


Figure 11. 3D Articulated Human Model [82]

D. Sample Training

A computer can learn human characteristics from a large number of human positive and/or negative samples in order to differentiate humans from other objects. This learning procedure is called training and is also one of the popular vision-based methods for human detection.

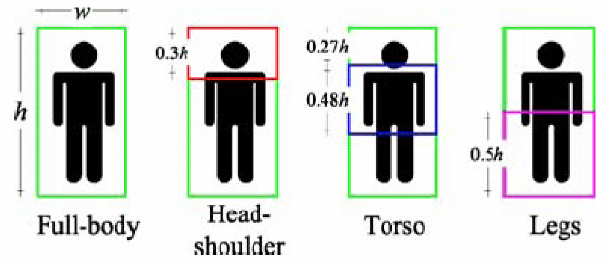


Figure 12. Spatial Relations of Body Parts [83]

For detection and tracking of humans, Wu and Nevatia [83] used three body part detectors trained with respective sample images of a head-shoulder, a torso, and legs. Besides the part detectors, a full-body detector is also learned, and Figure 12 shows the spatial relations of the body parts. The detectors are learned from silhouette-oriented features of 1742 frontal/rear views and 1120 side views of humans and 7000 negative images. Assuming that at least the head of a human in the image is visible, initial human candidates are detected on the basis of the responses of the head-shoulder and full-body detectors. Then each candidate is verified by the torso and legs detectors, considering their depth to overcome occlusions. The depth order is estimated by their y-coordinate, with the assumption that the camera looks down to a ground plane which the humans walk on, and the further the human from the camera, the smaller his or her y-coordinate.

In order to detect pedestrians in crowded real-world scenes with severe overlaps, Leibe and colleagues [14] combined the local information from sampled appearance features with global shape cues. The training images include a wide range of different clothing and accessories such as backpacks, handbags, or books. For the local appearance features, scale-invariant interest points and the patches around them are extracted from 105 training images and their mirrored versions and used for probabilistic top-down segmentations. For the global shape cues, pedestrian silhouettes are extracted from 210 training images, plus the mirrored ones, and fitted to the image to refine the segmentations.

Dalal and Triggs [84] proposed the use of histograms of oriented gradients locally normalised with 2478 human positive training examples and patches from 1218 person-free training photos. This was based on the idea that local object appearance and shape can often be characterised rather well by the distribution of local intensity gradients or edge directions. For the histograms of oriented gradients, the separate gradients of sample images are computed for each colour channel with Gaussian smoothing. Gaussian smoothing also calculates weighted votes in pixels for an edge orientation histogram channel and accumulates the votes into orientation bins over local spatial regions.

Viola and colleagues [85] integrated motion information and image intensity information to detect a walking person. The detector based on the motion information is trained with 2250 consecutive frame pairs of video sequences of street scenes with all pedestrians marked with a box in each frame and 2250 pairs of pedestrian-free frames. The detector based on the appearance information is trained on static patterns of the same video. As a single classifier for humans would require too many features to train with, cascade architecture is used here to make the detector extremely efficient. In cascade architecture, input images are passed to the first classifier for deciding true or false, and a true determination passes the input along to the next classifier in the cascade. The input can be classified as a true example only when all classifiers vote true, while a false determination halts further computation and returns false.

Sidenbladh [11] only employed motion information to train with sample images for human detection in video sequences since the appearance of humans varies hugely owing to clothing, identity, weather, and amount and direction of light in an uncontrolled outdoor environment. The motion cue for the detection is robustly estimated by means of dense optical flow, which consists of the horizontal flow and the vertical flow between a pair of consecutive images in a sequence. The training set contains 443 human flow patterns manually collected from dense flow images of many individuals in different types of environment, and 11688 non-human patterns automatically collected from similar sequences without humans.

In order to detect a salient human in robot vision, Kwak and colleagues [86] generated three feature maps from colour, luminance, and motion features of input images and combined them into a salience map. Moving objects are detected with the salience map, and are then classified into human and non-human by a SVM classifier trained with 150 human candidates and non-human candidates including trees and street lamps. The salience map is used again to determine the most salient human among the verified humans.

Face detection by sample training can help human detection. Schneiderman and Kanade [87] used multiple classifiers to detect faces at any size, location, and pose. Each classifier is based on the statistics of localised parts, and each part is designed to capture various combinations of locality in space, frequency, and orientation. The class-conditional statistics of these part values are collected from samples of face and non-face images so as to build each classifier. In detection, each classifier computes the part values within the image window to look up their associated class-conditional probabilities and then make a decision by comparison of partial likelihood ratio with a threshold. For human detection, Ishii and colleagues [88] proposed detection of faces and heads together in order to also detect people who are not facing a camera. Moving regions are detected by differentiation of three consecutive frames, and four directional (vertical, horizontal, and both diagonal) edge features are extracted from the regions and compared with face, non-face, head, and non-head samples for the face and head detection.

Not only to human detection like the above studies can the sample training method be applied but also to further recognition of human motion or poses. Fanti and colleagues [89] present a hybrid probabilistic model, which is efficient and effective for modelling and recognition of human motion. The probabilistic model combines global variables such as translation of the whole body, and local quantities such as relative position, velocities, and appearances of the body parts. For the recognition of walking motion, the model is trained with 378 frames of a single person walking from right to left, parallel to the camera, assuming that the height of a person in testing images is similar to the one in the training set. The approach of Hochuli and colleagues [90] detects non-conventional human movements using conventional and non-conventional motion sample videos. The foreground objects are segmented by background subtraction and tracked to extract features such as position, speed, changes in direction, and temporal consistency of the bounding box dimension. These features make up feature vectors, and the vectors are matched against the reference feature vectors obtained from sample videos.

In order to estimate the body configuration and pose in 3D space, Mori and Malik [91] stored exemplar 2D views of the human body in a variety of different configurations and perspectives with respect to the camera. On each of these views, the locations of the body joints are manually marked and labelled for future use. The input image is then matched to each stored view, using the technique of shape context matching in conjunction with a kinematic chain-based deformation model. The locations of the body joints are transferred from the exemplar view to the test shape for determining the 3D body configuration.

E. Motion Information

The cues previously presented for human detection are generally based on human appearance such as geometric features of humans and skin colour although some of them are integrated or trained with a non-appearance cue, motion information. This section reviews studies which mainly use motion information for human detection in images.

The method of Antonini and colleagues [92] targets pedestrians to detect and track and uses a behavioural model based on their trajectories. A calibrated monocular camera is used to capture images, and background-subtracted foreground regions are filtered to be 170 cm in height in the real world, with the assumption that the averaged height of human beings equals 170 cm. Then they are tracked to obtain their trajectories, and the trajectories are filtered to be the most human-like according to the behavioural model.

Dimitrijevic and colleagues [93] proposed a template-based approach to detecting human silhouettes in a specific walking pose. The templates consist of short sequences of 2D silhouettes, obtained from motion capture data, and thus the motion information helps distinguish actual people moving in a predictable way from static objects, whose outlines roughly resemble those of humans. The spatio-temporal templates contain silhouettes rendered from six different camera views and at seven different scales and are matched against portions of the input sequence.

For surveillance by pan/tilt/zoom cameras, Davis and colleagues [94] used motion history images as the temporal signature to separate *human activity* from *environmental noise* and *camera noise*. *Human activity* includes a person walking or cycling and moving vehicles, *environmental noise* covers tree shaking, smoke, and reflections, and *camera noise* comprises brick work, building edges, and lamp posts. Hence, *human activity* is defined as any translating object with a minimum spatial size and temporal length.

In order to prevent pedestrians at crossroads from being hit by vehicles, the algorithms of Pai and colleagues [95] detect and track pedestrians by combining the use of a pedestrian model and the walking rhythm of pedestrians after background subtraction. The pedestrian model is based on a non-rigid body, and its main feature is the width-to-height ratio of the human torso, located by matching an ellipse. The torso detection helps to locate the feet part, and the walking rhythm can be measured. Then pedestrians are separated from vehicles through comparison of the non-rigid motion of the lower half of the torso with the rigid motion of a vehicle.

F. Others

This section reviews human detection methods based on cues, which are not be included in the previous categories.

Munoz-Salinas and colleagues [96] took advantage of depth information to detect and track multiple people, using a stereo camera placed at an under-head position. After modelling the background as a geometrical height map of the environment, the foreground is extracted from the background and presented on a plan view map. The plan view map is called an occupancy map that registers in each coordinate the amount of foreground points projected in it. Assuming a person as an object in the occupancy map with sufficient weight, connected components with high occupancy level and appropriate dimensions are detected. Then a face detector from an open source is applied within the areas to confirm the human detection.

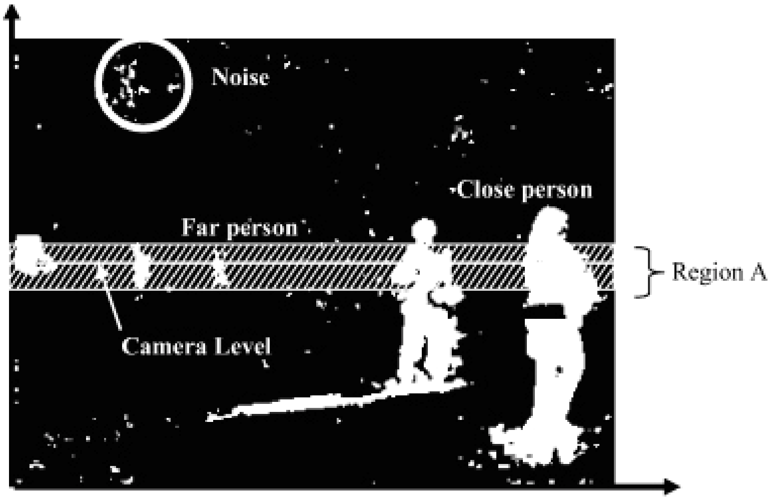


Figure 13. Standing Humans Intersecting Horizon Line (Region A) [97]

The method of Sato and Aggarwal [97] for human extraction is very simple because it is only intended to detect standing humans. After background subtraction, any noise is removed by keeping relatively large blobs and limiting the viewing range within a stripe, covering a small area above and below a horizon line, shown as Region A in Figure 13. The horizon line is set manually to correspond to the height of the camera lens.

G. Summary of Studies in Human Detection

The reviewed studies regarding human detection in images are summarised by the cues for human detection, motion segmentation methods, aims, and drawbacks in Table 2. The cues for human detection are categorised into geometric knowledge, skin colour, 3D human model, sample training, motion information, and others, and the method of motion segmentation preceding human detection is checked in each study. As the subsequent processes such as tracking and action recognition, are greatly dependent on human detection [84] and the cues for human detection should be chosen in the light of all the processes, the aim of each work is also studied. Lastly, the general drawbacks of each cue for human detection are synthesised.

For motion segmentation, there are two general kinds of methods. The first one is background subtraction methods, which capture background images in advance and subtract the background from current images to segment foreground objects. The second is successive image differentiation methods, which compare successive images and obtain their differences. The background subtraction methods are much more popular than successive image differentiation methods.

The general aim of the reviewed studies is human tracking, by which simple activities of people can be recognised such as coming in and out of a specific area although the studies adopting sample training methods generally focus on the human detection task. Owing to the variable appearances of humans, detecting and tracking humans in images is a challenging task [84]. The other studies estimate more detailed information on a human body such as gestures or whole body poses. In order to acquire such detailed information, a 3D model or a stick figure human model, which is fairly complicated, is exploited, or multiple cues are applied together.

As most of the cues used for human detection in images are related to human appearance, they have problems when some body parts are occluded and hidden from camera view. The use of a constant skin colour model can have other problems in that different races have different skin colours and illumination changes cause different skin colours. In order to overcome the occlusion problems, especially the occlusion by another body part in variable poses, 3D human body models attempt to recover 3D human poses and motions, but it is fairly hard to compute at least thirty joints and their kinematical constraints. Many studies propose a sample training method, but sample training requires a large number of human positive and/or negative images.

Table 2. Summary of Studies in Human Detection

Cue for Human Detection	First Author (Year) of Related Studies	Motion Segmentation Method	Aim	Drawback
Geometric Knowledge	Fan (2004)	Successive Image Differentiation	Human Pose Estimation	Occlusions can cause errors in application of geometric knowledge.
	Schleicher (2005)	Background Subtraction	Indoor Human Tracking	
	Ramanan (2007)	N/A	Human Articulation Tracking	
	Bird (2005)	Background Subtraction	Pedestrian Detection	
	Fihl (2006)	Background Subtraction	Human Tracking	
Skin Colour	Micilotta (2005)	Background Subtraction	Human Pose Estimation	Skin colours can vary depending on race or illumination. The face and hands should not be occluded.
	Yang (2005)	Successive Image Differentiation	Human Tracking	
	Pszczolkowski (2007)	N/A	Human Detection	
	Choi (2006)	N/A	Human Tracking	
	Ammouri (2008)	N/A	Human Activity Recognition	
	Kang (2007)	Successive Image Differentiation	Gesture Recognition	
	Medioni (2007)	N/A	Human Pose Estimation	

Table 2. Summary of Studies in Human Detection (Continued)

3D Human Model	Pham (2008)	Background Subtraction	Human Activity Recognition	A 3D human model is computationally complex and expensive.
	Pellegrini (2008)	Background Subtraction	Human Pose Estimation	
	Kehl (2006)	Background Subtraction	Human Tracking	
Sample Training	Wu (2007)	N/A	Human Detection	A large number of sample images are required with manual classification of the samples.
	Leibe (2005)	N/A	Pedestrian Detection	
	Dalal (2005)	N/A	Pedestrian Detection	
	Viola (2005)	N/A	Pedestrian Detection	
	Sidenbladh (2004)	N/A	Human Detection	
	Kwak (2007)	N/A	Human Detection	
	Schneiderman (2004)	N/A	Human Detection	
	Ishii (2004)	Successive Image Differentiation	Human Detection	
	Fanti (2005)	N/A	Human Motion Recognition	
	Hochuli (2007)	Background Subtraction	Human Motion Recognition	
	Mori (2006)	Background Subtraction	Human Pose Estimation	

Table 2. Summary of Studies in Human Detection (Continued)

Motion Information	Antonini (2006)	Background Subtraction	Pedestrian Tracking	Several successive images are needed to get motion information.
	Dimitrijevic (2006)	N/A	Human Pose Detection	
	Davis (2007)	N/A	Human Activity Detection	
	Pai (2004)	Background Subtraction	Pedestrian Tracking	
Others	Munoz-Salinas (2007)	Background Subtraction	Human Tracking	The use of simple cues is acceptable in limited situations.
	Sato (2004)	Background Subtraction	Human Tracking	

Although the use of motion information would not be affected by occlusions as much as the use of appearance-related cues, multiple consecutive images are necessary to obtain the motion information that means it takes time to detect humans. The other simple cues such as an occupancy map or intersection with a horizon line, are very interesting, but they are applicable in limited situations such as humans being with little clutter or standing.

As reviewed in this section, most of the cues used for human detection are related to human appearance, and they generally suffer from occlusions, complex computation, or a large number of sample images. Therefore, the work in this thesis uses dynamic motion cues for toddler detection, as detailed in Chapter 4 and Chapter 5.

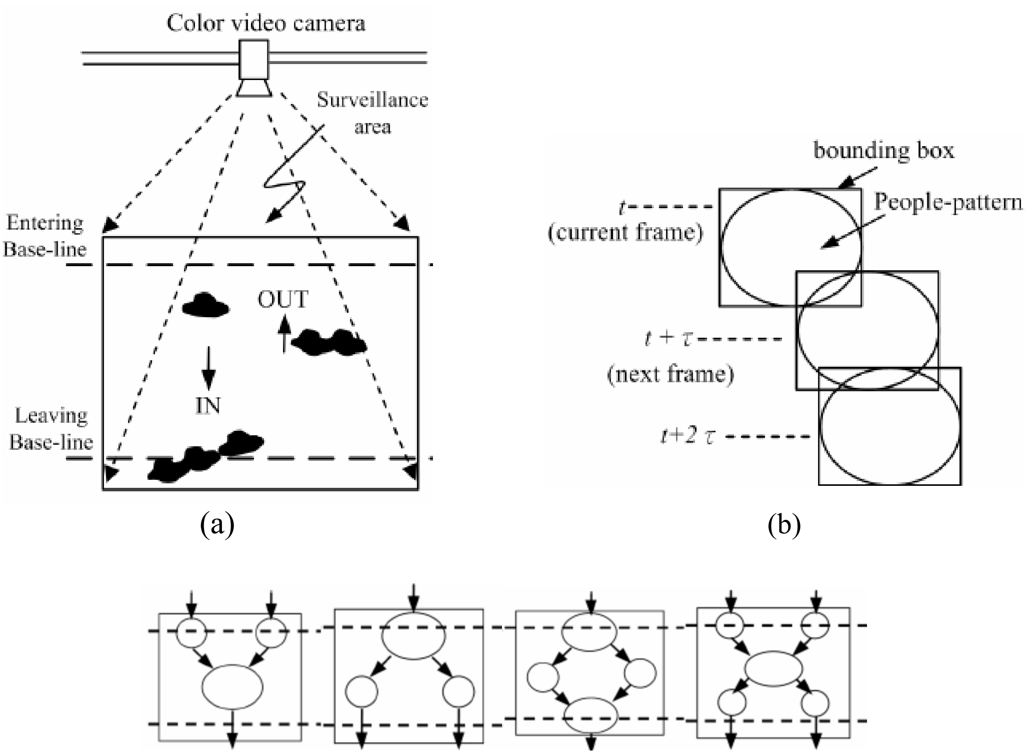
2.3.2 Handling of Regional Merges and Splits

In practice, self-occlusion and occlusions between different moving objects or between moving objects and the background are inevitable [98], and multiple camera systems offer promising methods to reduce ambiguities owing to occlusion. Multiple cameras have been used to choose the best view considering occlusion or to estimate 3D information of each object for coping with occlusion [99-103]. The use of multiple cameras, however, requires complex computation to match identical objects from different cameras or to calibrate the cameras for 3D information.

There are several studies that propose ways to tackle the occlusion problems using a single camera by handling regional merges of multiple objects. They deal with another similar problem whereby a single object can split into multiple regions which yield separate measurements. The splitting may result from crossing occlusions or errors in background subtraction, and it can be generated despite the use of good background subtraction techniques [104]. Their methods can be largely divided into use and non-use of colour information.

A. Use of Colour

The system of Chen and colleagues [105] counts pedestrians, passing through a gate or a door, with a zenithal video camera, as shown in Figure 14a. Hue saturation intensity (HIS) colour histograms are used to distinguish one pedestrian from another in a multi-people heap because the hue component is intimately related to the way in which humans perceive colour. As the colour label can become ineffective in identification in the case of multiple pedestrians wearing same-coloured clothing, the two overlapping boxes in Figure 14b, which bound identical colour patterns in adjacent images, are judged as the same person. In order to analyse merging or splitting cases within the door area, area changes of moving regions are checked, on the basis of the fundamental cases of merging or splitting in Figure 14c. This study is not, however, concerned with a single person splitting into multiple regions owing to errors in background subtraction.



(a) (b)

(c)

Figure 14. (a) People Counting System, (b) its Tracking and (b) Basic Cases of Merge-Split [105]

Conversely, the approach of Medioni and colleagues [106] does not cope with multiple objects merging into one region, but only with a single object splitting into multiple regions. Their work involves detection and tracking of moving objects and analysis of their trajectories to recognise the behaviour of the moving objects. In order to extract the correct trajectory of each object, aperture problems, which can split a single object into multiple regions, are handled by measurement of the grey-level similarity between a moving region at one frame and a set of regions at the next frame in its neighbourhood. The size of this neighbourhood is estimated from the object motion amplitude, and the matches of moving objects between consecutive frames are represented by nodes and edges, as illustrated in Figure 15.

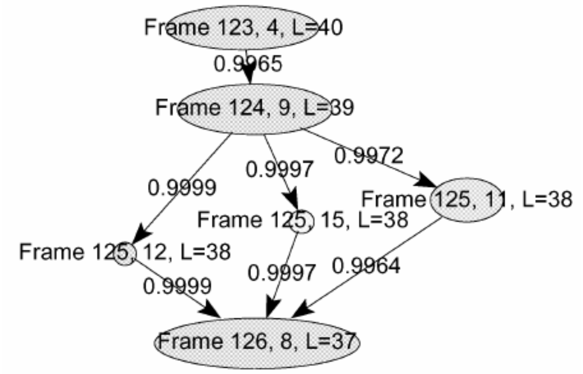
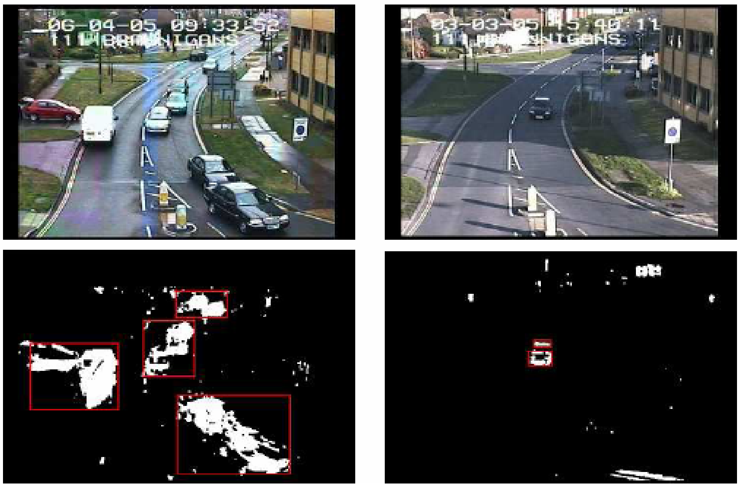


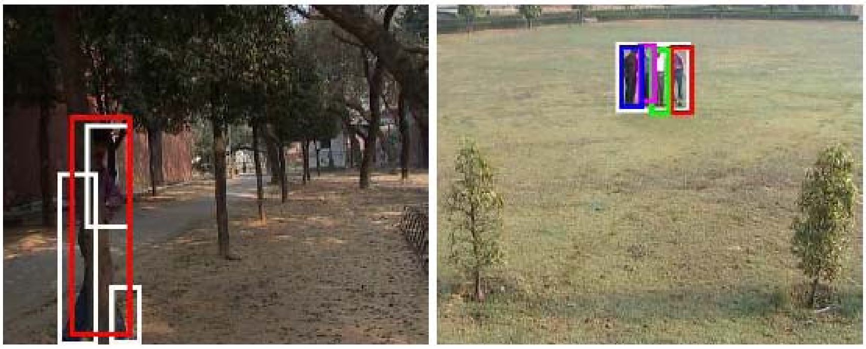
Figure 15. Detected Regions and Associated Graph [106]

The vehicle tracking system of Song and Nevatia [107] for street surveillance is based on an appearance model of a colour histogram to detect both multiple objects merging into one region (Figure 16a) and a single object splitting into multiple regions (Figure 16b). Apart from the colour histogram, each of the blobs, detected as moving vehicles, is modelled as a rectangle to predict its new position and check overlaps between predicted rectangles and observed rectangles for blob association over successive frames.



(a) (b)

Figure 16. (a) Merged and (b) Split Blobs [107]



(a) (b)

Figure 17. (a) Partial Occlusion and (b) Crowding [108]

Guha and colleagues [108] defined six qualitative occlusion primitives, based on the well-known cognitive assumption of persistence, under which objects continue to exist even when hidden from the view. The primitives are *isolated*, *partial occlusion*, *crowding*, *disappear*, *enter*, and *exit*, which respectively indicate blobs separated and fully visible, blobs separated but partially invisible (Figure 17a), blobs merged into one region (Figure 17b), blobs detected previously but no longer visible, new blobs with no relation to previous blobs, and blobs disappearing at the scene boundary. In order to recognise these occlusion primitives, each agent is characterised by its occupied pixel set, weighted colour distribution, and the trajectory of the minimum bounding rectangle of the pixel set. The agent is also associated with detected foreground blobs, based on the colour distribution and predicted agent position from the trajectory. All the occlusion primitive notations on each agent are recorded in the history for further recognition.

The method of McKenna and colleagues [104] tracks people through mutual occlusions when they form groups and separate from one another, as presented in Figure 18. In order to overcome the problem of a person splitting into multiple regions, the conditions of multiple regions to form a single person are defined to be in close proximity, to have overlapped projections onto the x-axis, and to have a total area larger than a threshold. In order to track people consistently when they enter and leave people groups, a colour model is built and adapted for each person being tracked. The tracker based on the colour information can fail in tracking of each person, however, when two people clothed in a very similar manner form a group and subsequently separate, for example.

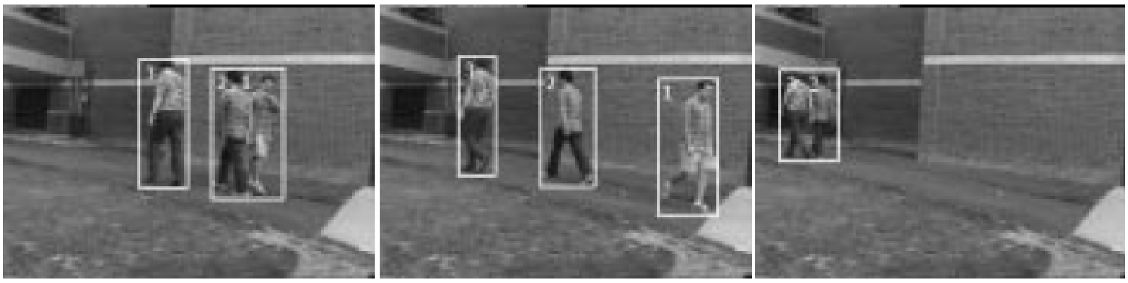


Figure 18. People in Groups [104]

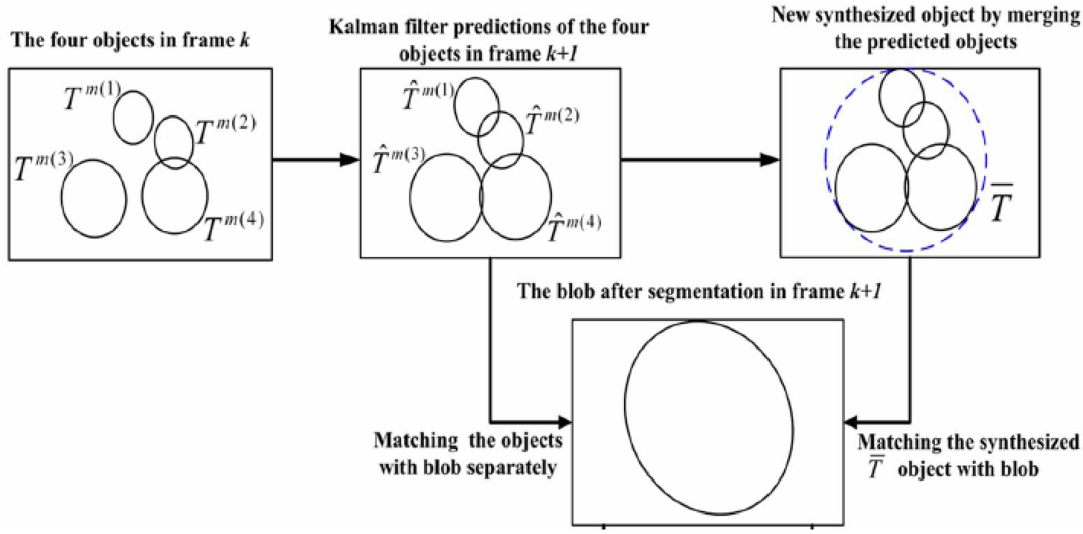

The four objects in frame *k*
$T^{m(1)}$
$T^{m(2)}$
$T^{m(3)}$
$T^{m(4)}$
Kalman filter predictions of the four objects in frame *k+1*
$\hat{T}^{m(1)}$
$\hat{T}^{m(2)}$
$\hat{T}^{m(3)}$
$\hat{T}^{m(4)}$
New synthesized object by merging the predicted objects
$\overline{T}$
The blob after segmentation in frame *k+1*
Matching the objects with blob separately
Matching the synthesized $\overline{T}$ object with blob

Figure 19. Process of Prediction and Matching [109]

B. Non-Use of Colour

Kumar and colleagues [109] used Kalman filter-based trackers to maintain the identity of multiple targets while tracking them in the presence of regional splits and merges. The trackers predict and estimate states of the target objects, and the predicted shape and position of the objects give rise to a new synthesised blob when the objects are predicted to merge. The blue ellipse of a broken line in Figure 19 is an example of the new synthesised blob. The real segmented blob is matched with the objects separately and also the synthesised blob as shown in Figure 19 by use of a geometric shape-matching algorithm. These association methods work well as long as position and motion of target objects are predictable.

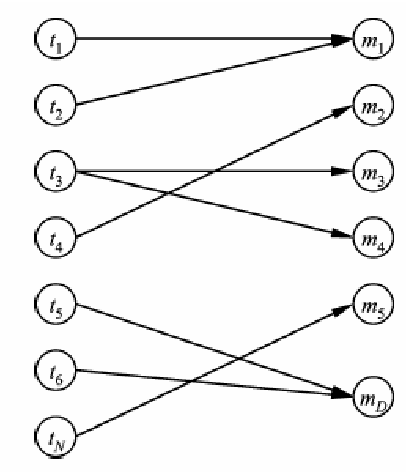


Figure 20. Target-Measurement Association [110]

Joo and Chellappa [110] proposed a multiple-hypothesis approach to tracking multiple objects by handling objects which enter or exit the view or regionally merge or split, as well as by detecting split fragments of a single object owing to limitations in background subtraction. For those kinds of objects, a single target (t_N) may need to be associated with multiple measurements (m_D), and multiple targets with a single measurement, as shown in Figure 20. The multiple-hypothesis tracking considers a set of feasible hypotheses, regarding joint associations between targets and their measurements. The centre coordinates, the bounding box size, and the velocity of each target are defined as its state at every frame, and the position is predicted and compared with the real measurements for the best match.

C. Summary of Studies in Handling of Regional Merges and Splits

The reviewed studies regarding handling of regional merges and splits are summarised in Table 3 to present the major cues and drawbacks.

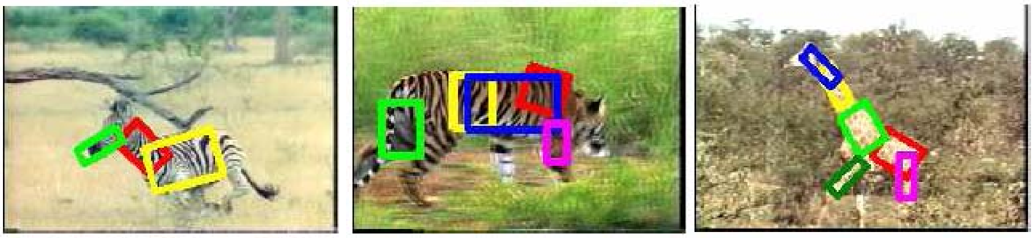
Table 3. Summary of Studies in Handling of Regional Merges and Splits

Classification	First Author (Year) of Related Studies	Cues for Handling of Regional Merges and Splits	Drawback
Use of Colour	Chen (2006)	HIS Colour Histogram + Box Overlap	The use of colour information can cause confusion in the case of a single object wearing multiple colours or multiple objects with similar colours in a group.
	Medioni (2001)	Grey-Level Similarity + Neighbourhood Size	
	Song (2007)	Colour Histogram + Predicted Position	
	Guha (2006)	Colour Distribution + Predicted Position	
	McKenna (2000)	Colour Model + x-Projection Overlap + Area Size	
Non-Use of Colour	Kumar (2006)	Prediction and Match of Shape and Position	The association of prediction and real measurements can work when position and motion of targets are predictable.
	Joo (2007)	Prediction and Match of Position	

Colour information is employed in many studies to identify each of the multiple objects which appear together in one image region, to detect regional fragments of a single object based on proximity, or to consistently track individuals despite regional merges and splits. As the sole use of colour information can incur confusion owing to different objects in similar colours or a single object in different colours, generally the position information of each object is added to limit the range of searching for the identical object in the next frame within the area where the object will possibly be.

The studies, which do not use any colour information, predict the shape or the position of each object and detect the closest match with the real measurements. The association can work only when the motion of each object is correctly estimated and its proper position in the next frame is predictable.

As the use of colour similarity can confuse tracking of individuals, the work in this thesis employs position information commonly used in existing methods of handling regional merges and splits, as described in Section 4.3.4A.



(a) (b) (c)

Figure 21. Pictorial Structures from Videos of (a) Zebra, (b) Tiger, and (c) Giraffe [111]

2.3.3 Classification of Animals

There are just a few studies on animal detection in images. The automatic system of Ramanan and colleagues [111] builds 2D articulated models which are pictorial structures as shown in Figure 21, from videos of different animals. The pictorial structures are augmented with a discriminative texture model, learned from a texture library, and the models are used to identify and track the animals. As the pictorial structure models are dependent on the aspect of the videos used for the model construction, the animal detection based on the models would not work well with videos capturing different aspects of the target animal. For example, if a model is learned from a video of a giraffe walking sideways, the system may not be able to find a giraffe walking towards a camera based on the model.

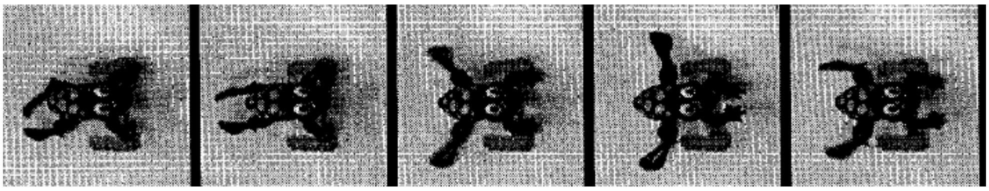


Figure 22. Periodic Activity of Frog [112]

The approach of Polana and Nelson [112] detects any repetitive activity using low-level, non-parametric representations in order to recognise locomotion of a human or animal. The approach is based on repetitive motion, and a moving actor is segmented and normalised spatially and temporally; the actor's activity can be recognised by matching against a spatio-temporal template of motion features. The sample images of a frog's periodic activity in Figure 22 are included in the reference database to build the activity template. Since this recognition focuses on specific activities of a human or an animal, the method would not be able to recognise a human or an animal that moves differently from the motion template.

Hayfron-Acquah and colleagues [113] focused on the symmetry properties of gaits in order to discriminate different animal movements and recognise human motion. Individual gait signature for different animals can be derived from image sequences of the animal movements by estimation of the symmetry of the movements, as presented in Figure 23. The zebra signature in Figure 23d is very close to the signature in Figure 23b, despite the appearance of the tail since the tail movement does not affect the resulting symmetry. The use of the symmetry signatures can separate different animals as well as distinguish the quadruped movement of general animals from human motions. Jiang and Daniell [114] also attempted to distinguish between two-legged human movements and four-legged animal movements by extracting motion features in spatio-time dimensions.

Those analyses of leg movements, however, would need side views to capture the movement of all the legs and have problems with occlusion on a leg.

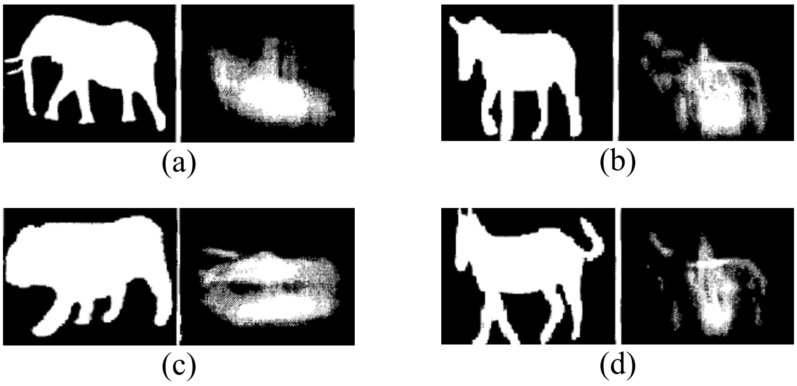

(a) (b) (c) (d)

Figure 23. Symmetry Signatures for (a) Elephant, (b) Zebra1, (c) Bulldog, and (d) Zebra2 [113]

This review is referred to in developing the novel method of classifying a human and a pet illustrated in Chapter 4.

2.3.4 Camera Calibration

In order to estimate information about a 3D original space from its projection images, the camera in use should be calibrated in advance to learn the relations between the original scene and its image. Camera calibration techniques can be roughly classified into two categories: object-based calibration and self-calibration. The object-based calibration is performed by observation of a calibration object, whose geometry in the 3D space is known with very good precision, and the calibration can be done very efficiently. The self-calibration, however, cannot always produce reliable results since a camera is calibrated by matching identities in several images, taken by the same camera, and there are many parameters to estimate [115].

Therefore, this section only reviews the studies on object-based calibration which have been published since 2000. They are categorised by the kinds of objects with known geometry used for camera calibration: background, planar, 3D, and moving objects. The background object is a geometric feature, which belongs to the background of the image, and the planar object and the 3D object are features placed on the scene for the purpose of camera calibration. With the moving object, its movement is tracked and used to calibrate a camera.

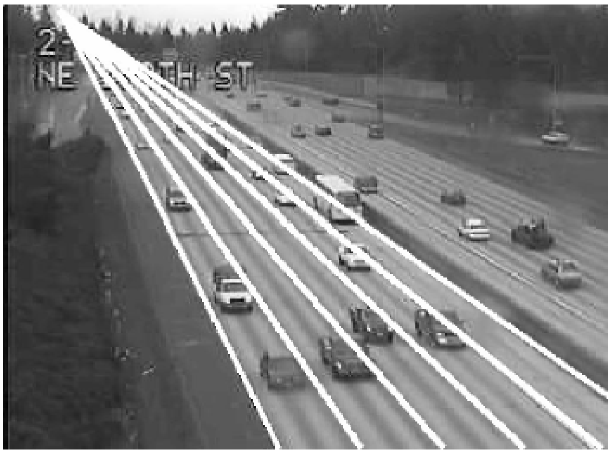


Figure 24. Lane Boundaries and their Vanishing Point [116]

A. Background Object

The algorithms of Schoepflin and Dailey [116] calibrate roadside traffic management cameras and track vehicles so as to sense traffic speeds. The camera position is estimated relative to the roadway, by use of the motion and edges of the vehicles. Given the camera position, the camera is calibrated by estimation of the lane boundaries and the vanishing point of the lines along the roadway in Figure 24. The scene is modelled as a set of parallel lines, viewed through a pinhole camera, which is a camera without a conventional glass lens and the scene does not get any effects of a lens.

The camera calibration algorithm of Yuan and colleagues [117] is for detection of video-based traffic information such as vehicle flux, speed, and possession ratio. The nine control points, made by the intersection of three horizontal lines with three vertical lines in Figure 25, are used for camera calibration, and the calibration model is based on a pinhole model and direct linear transformation[4] (DLT). The model is extended also to resolve nonlinear distortion problems owing to the use of a lens, which are the radial distortion, decentring distortion, and thin prism distortion.

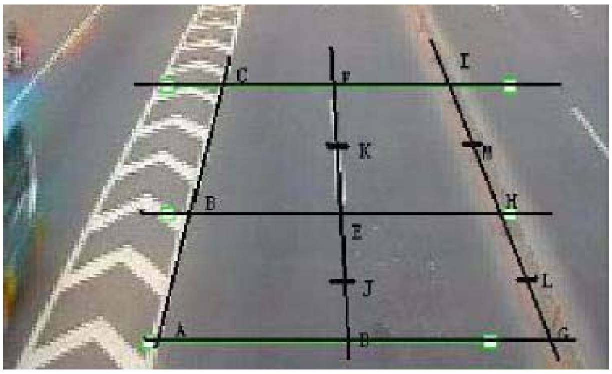


Figure 25. Control Points on Lanes [117]

Farin and colleagues [119] proposed a real-time camera calibration algorithm to obtain player and ball positions in the real-world coordinates for semantic analysis of sport sequences. The marker lines on the field for court sports like tennis are used to determine the calibration parameters. The correspondences are determined between the detected lines in the sport sequence (Figure 26a) and the lines in the court model (Figure 26b) and used to compute the homography between the image plane and the real world.

Although extra figures do not need to be prepared for camera calibration owing to making use of background image information, all the above methods are applicable to limited areas such as roadways or sport courts.

[4] Direct linear transformation is a reconstruction algorithm with a set of linear equations and a set of parameters, the most common approach to characterising the calibration, position, and orientation of a single camera [118].

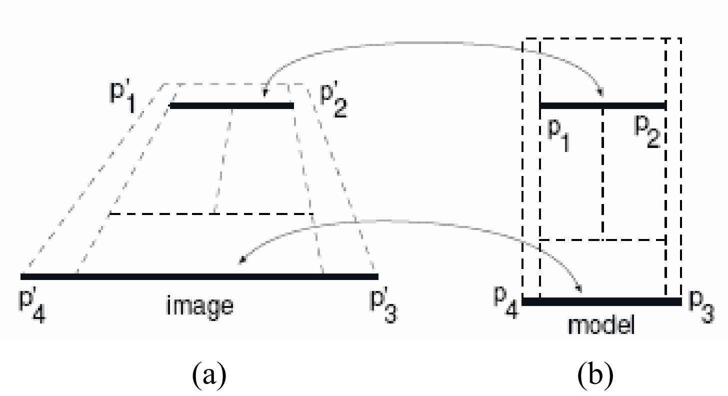


Figure 26. Court Marker Lines (a) in Projected Image and (b) in Model [119]

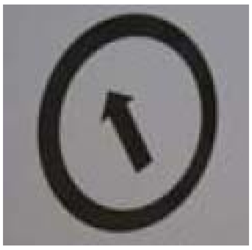


Figure 27. Circle and Right Angles [120]

B. Planar Object

For the environments without any geometric figure included, some planar objects can be placed on the scene and used for camera calibration such as a circle and right angles, concentric circles, and squares. Zhong and colleagues [120] used a planar pattern, which includes a circle and right angles from the orthogonal lines of an arrow head and tail, as shown in Figure 27. A dual conic constraint on the vanishing line of a plane can be derived from a right angle and a circle on the same plane, and a vanishing line can be uniquely identified from three independent right angles. On the basis of the vanishing line, imaged circular points are computed and used to calculate the coordinates of the principal point, focal lengths, and skew parameter of the pinhole camera model.

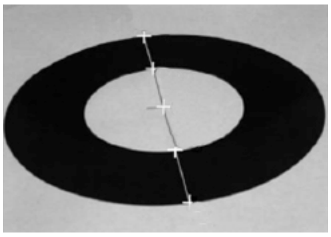


Figure 28. Concentric Circles [121]

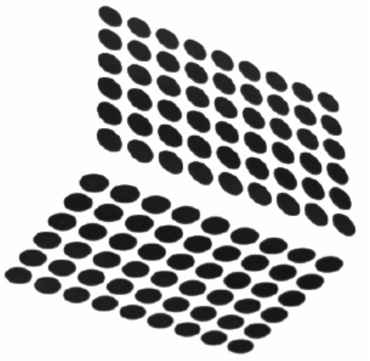


Figure 29. Two Perpendicular Planes with Circular Control Points [122]

Kim and colleagues [121] exploited a pair of concentric circles without knowing their centre and radii, as shown in Figure 28 for camera calibration. The centres of projected circles are generally treated as the projected centres of the circles, but it is actually improper under general perspective projection. The use of a pair of concentric circles instead of a circle can recover the projected centre very accurately. Moreover, this method can estimate imaged circular points without computation of a vanishing line and its intersection with a projected circle. Liu and Su [122] present a camera calibration method using circular control points, which just need to be corrected for the distortion, caused by the asymmetric perspective projection. The calibration pattern in use is two perpendicular planes, each with 24 circular control points, as shown in Figure 29, and the centres of the circles in the image plane are iteratively corrected from the distortion. The establishment of the relationship between 3D world coordinates and their corresponding 2D image coordinates follows DLT based on the pinhole camera model.

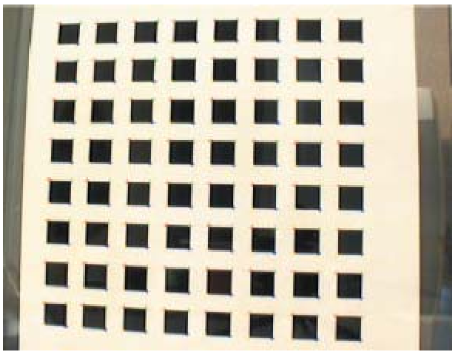


Figure 30. Square Pattern [115]

One of the most popular planar pattern units is a square [115, 123-126]. This may be because the four vertices of a square can be easily detected, and a square pattern can provide a large number of point features, whose actual positions can be easily defined from the square size and the regularity of arrangement. 256 points, for instance, can be detected from the pattern in Figure 30 and used for camera calibration.

Zhang [115] and Wang and colleagues [123] modelled a camera based on a pinhole, using the square pattern in Figure 30, and calculated the radial lens distortion, which desktop cameras usually have. The method of Baba and colleagues [124] takes into account both the geometric information of the feature points and the defocus information of edges of the square pattern in Figure 31. Zhang and colleagues [125] used a square pattern projected onto a planar surface placed on the scene, as illustrated in Figure 32, in order to recalibrate a camera based on the pinhole model with pre-defined intrinsic parameters. The calibration technique of Xin and Xiao-guang [126] is based on the pinhole camera model without considering the skew and uses a single regular quadrilateral, the coordinates of whose four points can be easily set up with the side length, as shown in Figure 33.

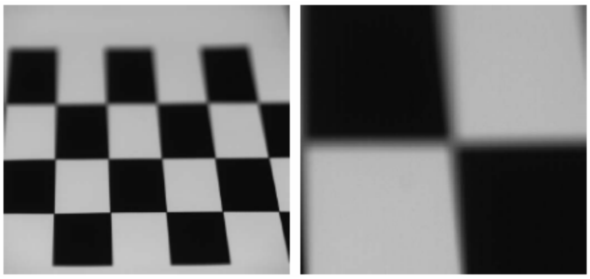


Figure 31. Blurred Edges of Squares [124]

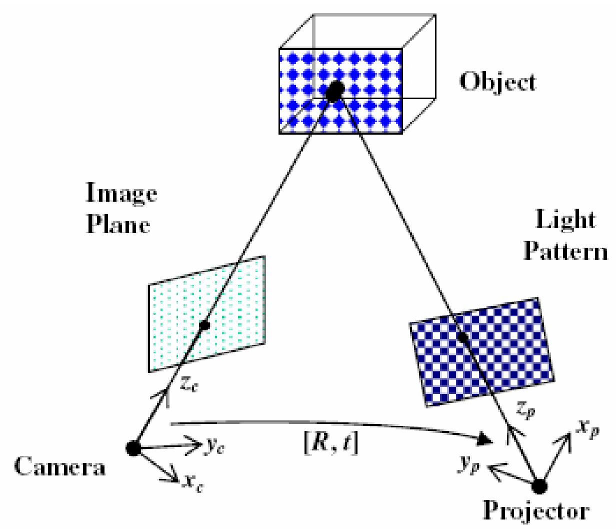


Figure 32. Projected Square Pattern [125]

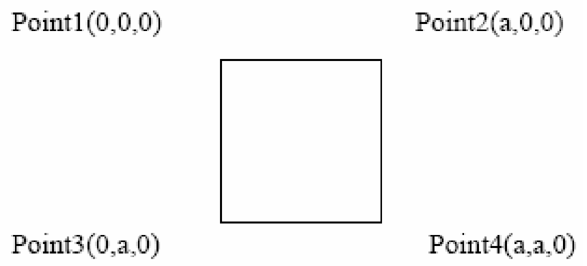


Figure 33. Regular Quadrilateral [126]

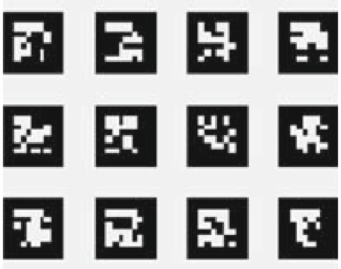
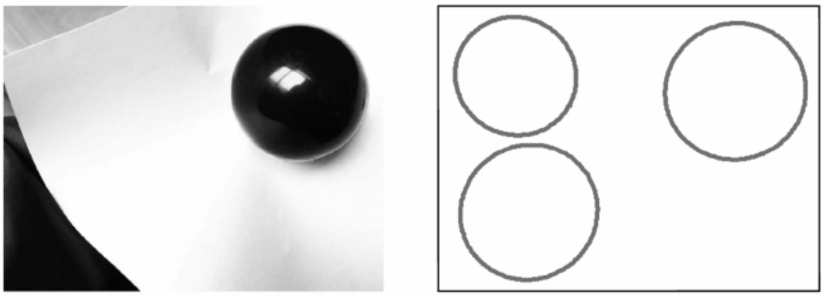


Figure 34. Self-Identifying Pattern [127]

The system of Fiala and Shu [127] allows a camera to be calibrated in a matter of minutes merely by being passed in front of a planar array of self-identifying markers. The marker system is bi-tonal and contains 2002 planar markers, each consisting of a square border and an interior region filled with a 6 x 6 grid of black or white cells as in Figure 34. The centre or the corners of each marker are used as correspondences between the world coordinates and their pinhole camera projection. The best calibration results are achieved using the marker centre instead of each corner owing to decreased sensitivity to lighting and focus.



(a) (b)

Figure 35. (a) Sphere Image and (b) Ellipses from 3 Images [128]

C. 3D Object

Ying and Zha [128] applied three images of a sphere and their ellipses in Figure 35 to camera calibration by geometrically interpreting the relation between sphere images and the image of the absolute conic (IAC), which plays a central role in camera calibration. The geometric interpretations are that each sphere image has double-contact with the IAC, and that the IAC is determined by use of some conic fitting method after six double-contact points from three sphere images are found. Their calibration method is derived from a pinhole camera model and geometric interpretations.

Avinash and Murali [129] used two vanishing points to estimate the focal length and the centre of focus. Vanishing points can easily be found from natural rectangular prisms such as cartons, boxes, and buildings owing to the perspective distortion of the prism's edges in an image, acquired on the principle of a pinhole camera. In their method, however, the vanishing points are calculated from points manually detected on a single image, and the orientation of the rectangular prism has to be known.

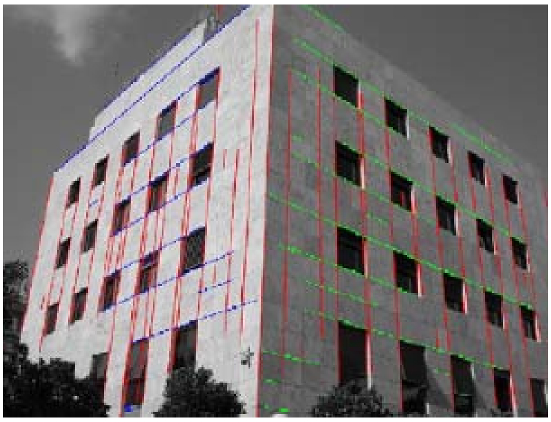


Figure 36. Line Segments in Vanishing Points [130]

Grammatikopoulos and colleagues [130] developed a camera calibration algorithm from single images, including three vanishing points of orthogonal space directions such as the building in Figure 36. Extraction of image line segments and their clustering into groups, corresponding to three dominant vanishing points, is performed. At the same time, the principal point location as well as the coefficients of radial symmetric lens distortion is estimated by a unified least-squares adjustment of all image points belonging to the lines, which intersect at the three dominant vanishing points of the scene.

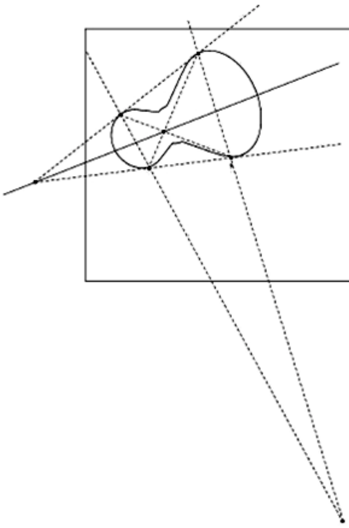


Figure 37. Vanishing Points from Revolution Surface [131]

The calibration method of Wong and colleagues [131] is based on the vanishing points of the symmetry properties exhibited in the silhouettes of revolution surfaces, as illustrated in Figure 37, which are commonly found in daily life such as bowls and vases. On the assumption of zero skew and unit aspect ratio in the pinhole camera model, the principal point of a camera would coincide with the ortho-centre of a triangle with vertices, given at three vanishing points from three mutually orthogonal directions. Such properties of vanishing points are used together with the symmetry properties associated with the silhouettes of revolution surfaces in order to derive a simple technique for camera calibration.

D. Moving Object

The algorithm of Zhao and Liu [132] recovers the intrinsic parameters as well as the extrinsic parameters of multiple cameras by capturing an 1D object's rotations around a fixed point. The calibration object has three collinear points as presented in Figure 38, and a series of its rotations around the point at the end is captured by each of the cameras. According to the geometric structure of the object, the projective depths are estimated on the basis of the pinhole camera model. For a camera calibration object, Chen and colleagues [133] employed a spatial triangle with known size, rotating freely at one of its vertices. Only two calibration images at a minimum are required to calculate the intrinsic camera parameters, based on the pinhole camera model, by making full use of the geometric information of the triangle.

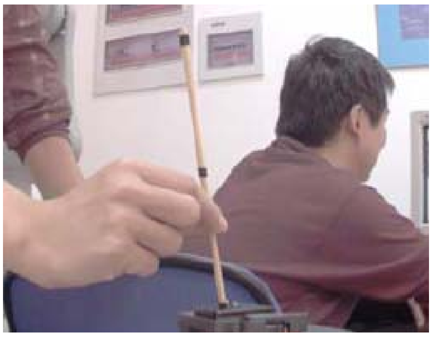
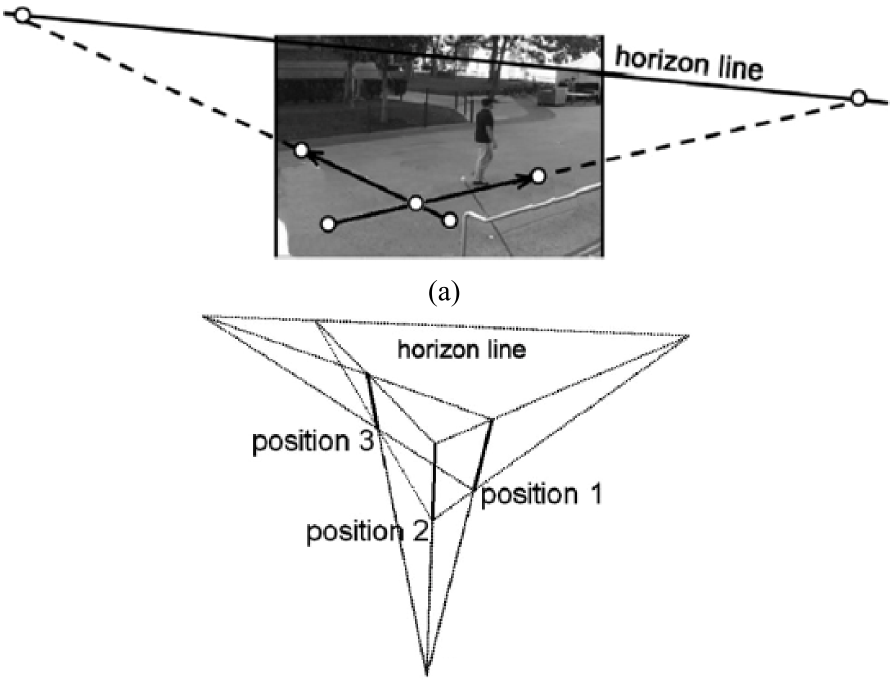


Figure 38. Collinear Points Rotating around Fixed Point [132]

The method of Lv and colleagues [134] estimates a camera's intrinsic and extrinsic parameters in the pinhole model from vertical line segments of the same height. The necessary lines are segmented by detection of the head and feet positions of a walking human in his/her leg-crossing phases as in Figure 39a. The three vanishing points and the horizon line are computed from vertical poles of the same height as presented in Figure 39b and used with the actual length of the human to estimate all the camera parameters.

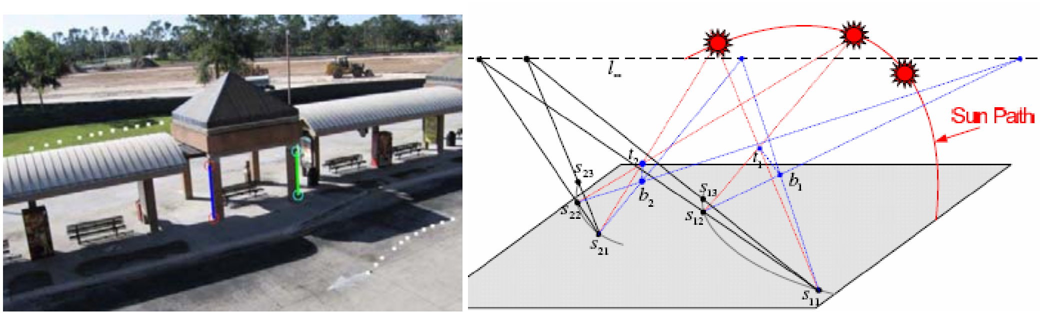
Lu and colleagues [135] used the cast shadows of two 3D points observed over time to recover the camera parameters. Assuming that the camera has a unit aspect ratio and zero skew, the horizon line in the image is estimated by use of the shadow trajectories of two stationary objects on the ground plane, as illustrated in Figure 40b. This is because the line segments defined by corresponding shadow points are parallel in the world and therefore intersect on the horizon line in the image plane.



(a)

(b)

Figure 39. (a) Walking Human and (b) Derived Vanishing Points [134]



(a)

(b)

Figure 40. (a) Shadow Trajectories and (b) Derived Horizon Line [135]

E. Summary of Studies in Camera Calibration

The studies reviewed here with regard to camera calibration are summarised in Table 4. They are categorised by the kinds of the object used for camera calibration, and the other parts of the table present the basic camera model, on which each study is based, and the drawback to the use of each kind of calibration object.

Most of the studies are based on the pinhole camera model, which specifies intrinsic camera parameters and spatial relationship between the world and its projected image without considering any effects of a lens on the images. Lens effects can be neglected according to the research subjects or purposes, but they need to be dealt with for accurate results. The lens effect taken into account the most is radial distortion.

The use of geometric figures, which belong to the background such as boundaries of lanes or marker lines of sport courts, requires no additional preparation of a calibration object and makes it possible to calibrate a camera at any time since the figures always appear on the scene unless occluded. The methods can, however, be applied to limited areas, and different models with geometric knowledge should be built in advance for different figures.

Planar figures or their patterns are used in many studies since they can be constructed very easily and accurately by printing. Square patterns are especially popular, possibly because four vertices of a square can be easily detected and their 3D coordinates can be simply identified with the length of one side of the square. Square patterns can also hardly be seen in the general background and thus are unlikely to be confused with other background objects.

Table 4. Summary of Studies in Camera Calibration

Calibration Object		First Author (Year) of Related Studies	Camera Model Basis	Drawback
Background Object		Schoepflin (2003)	Pinhole	It is applicable to limited areas, roadways or sport courts.
		Yuan (2006)	Pinhole + Distortion (Radial, Decentring, Thin Prism)	
		Farin (2005)	Pinhole	
Planar Object	Conic	Zhong (2006)	Pinhole	In the use of a single figure, a background object, with a similar appearance to but different properties from the figure, might be recognised and used for calibration.
		Kim (2005)	Pinhole	
		Liu (2008)	Pinhole	
	Square	Zhang (2000)	Pinhole + Radial Distortion	
		Wang (2006)	Pinhole + Radial Distortion	
		Baba (2006)	Pinhole + Defocus	
		Zhang (2007)	Pinhole	
		Xin (2005)	Pinhole	
		Fiala (2008)	Pinhole + Distortion (Radial, Thin Prism)	
3D Object		Ying (2006)	Pinhole	
		Avinash (2008)	Pinhole	
		Grammatikopoulos (2007)	Pinhole + Radial Distortion	
		Wong (2003)	Pinhole	
Moving Object		Zhao (2008)	Pinhole	The data are not very accurate.
		Chen (2008)	Pinhole	
		Lv (2006)	Pinhole	
		Lu (2006)	Pinhole	

All the studies exploiting 3D objects or moving objects for camera calibration try to detect lines which are actually parallel but generate vanishing points in the image or are same-sized in world coordinates but not in the image plane. Therefore, the 3D objects should be perfectly symmetrical in every aspect in the world coordinates, and the moving objects should move in a perfectly constant way in order to extract parallel or same-sized lines. Capturing the heights of a walking human or the trajectories of shadows generated by the sun in the distance, however, would not be able to provide very accurate data for the calibration.

This review helps to find a common method of camera calibration to be used for classification between a human and a pet, as explained in Section 5.3.2.

Chapter 3
Identification of Risk Factors of a Toddler’s Fall Injury

3.1 Introduction

One of the objectives of this research is to identify key factors leading to a toddler’s fall injuries. The factors should be related to a toddler’s behavioural or environmental changes, which need to be continuously watched in order to prevent fall injuries. In addition to the suggestions for preventing children’s fall injuries, reviewed in Section 2.2, more than two thousand records of falls, the injuries from which were severe enough to be treated in hospital, are analysed in Section 3.2. Based on the suggestions and the records, the risk factors of a toddler’s fall injury in the home environment are identified. Finally, the identified factors are assessed by experts on children’s home accidents by means of a questionnaire in Section 3.3.

3.2 Analysis of Fall Records

The Royal Society for the Prevention of Accidents (RoSPA) is a charity involved in the promotion of safety and the prevention of accidents in all areas of life [136], and provides data about accidents by email on request. The data are based on the Home Accident Surveillance System (HASS), which holds details of home accidents which caused injury serious enough to warrant a visit to hospital. The details of home accidents were gathered by interviewing a patient or a caregiver at Accident & Emergency (AE) units in a representative sample of up to eighteen hospitals across the United Kingdom from 2000 to 2002 [137]. The details of more than six thousand fall accidents of young children aged up to five in the home were obtained from RoSPA, and around two thousand of them are analysed in this section. This research focuses on accidents which happened to toddlers aged between one and three in lounge, study, living/dining/play area and inside stairs, where they generally experienced fall accidents. Two sample accident details are as follows:

- Sample 1: The patient at home, running around the room; tripped and fell, banging head on a skirting board and then on the carpeted floor;
- Sample 2: The patient fell down six or seven carpeted stairs, then hit her face on a cold metal radiator, breaking the valve; the accident was unwitnessed; the patient was coming down stairs unaided.

3.2.1 Accident Models

In order to analyse the massive number of accident details, an accident model is necessary for guidance. Harvey [138] reviewed four models for accident investigation which are applicable to accidents in general and widely used. The models are Heinrich's domino model, epidemiology, fault tree models, and multilinear events sequencing.

Heinrich proposed three elements that define the basic safety problem. The first element is the initial environment which consists of a safe or an unsafe state, and the second one is the decision space that involves choice between safe and risky acts. This is because the individual is assumed to diagnose the state of the environment as safe or unsafe and to choose a safe or a risky act. The third is the probabilistic nature of an accident given human error, a failure to detect an unsafe state [139]. Heinrich's domino model depicts an accident as a set of dominos, which tumble because of a unique initiating event. In this model, the dominos that fall represent the action failures, while the dominos that remain standing by default represent the normal events. This type of model is deterministic because the outcome is seen as a necessary consequence of one specific event, but it suffers from being oversimplified and easily confuses sequentiality with causality [140]. This model serves the legal and prevention purposes quite well, but it is potentially biased in the identification of causes and is most inadequate for descriptive research purposes [138].

The epidemiological type of accident model compares an accident to a spreading disease as the outcome of a combination of factors. The factors are associated with the host, agent, and environment, and accidents are stated to happen when a sufficient number of factors come together in space and time. Hence, epidemiological models provide a basis for discussing the complexity of accidents [140]. The epidemiological approach can potentially serve the description, research, and prevention purposes fairly well with a complete database identifying accident causes, but it does not concern itself with any of the legal purposes [138].

The general fault tree approach to accident investigation advocates a description of all the necessary and sufficient conditions for an accident within the work system in question [138]. A specific adaptation of the model, the Management Oversight and Risk Tree (MORT), investigates incidents, relying on a logic tree diagram. The decisive diagram serves to describe connections between an incident and individual features of the process safety management system, and it is to be applied to key episodes in the incident sequence of events [141]. MORT focuses on investigations largely dealing with management oversights and failures to attend closely to the accident event itself and therefore it is not adequate for the purposes of describing an accident, identifying causes, and conducting research [138].

Lastly, multi-linear event sequencing devotes close attention to the sequence of events leading up to the accident with special status given to the temporal relations between events [138]. A chronological array of the events helps to structure the search for the relevant factors and events involved in the accident, and it provides a method for testing the relevance of additional events or conditions encountered by the investigator [142]. The multi-linear events sequencing model can provide quality information for research and prevention purposes with an excellent description of the accident process, but it avoids causal and legal purposes [138].

3.2.2 Epidemiological Analysis

The research presented in this thesis intends to identify the causal factors of a fall accident, and the epidemiological approach has been chosen for the analysis of the fall records collected from RoSPA since it provides a complete database for identifying accident causes. Harvey also suggests that the epidemiological approach potentially can serve description, research, and prevention purposes fairly well [138]. As this research focuses on fall incidents which cause an injury to be treated in hospital, a conceptual framework for epidemiological analysis, developed by William Haddon, the so-called Haddon's Matrix, was selected to understand how injuries occur. Haddon is widely considered as the father of modern injury epidemiology [143].

In the epidemiology of injury, all the factors which interact with each other to account for the presence or absence of disease or injury, are categorised as *host*, *agent*, *vector*, and *environment*. *Host* is the person injured, the *agent* is the force or energy, *vector* is the person or thing that applies the force, transfers the energy, or prohibits its transfer, and *environment* is the situation or conditions under which the injury happens [144]. A similar model could be used for an act of interpersonal violence, in which a man or woman slaps his or her partner. In this case, *host* would be the person slapped, *agent* would be the mechanical force or energy, slapping, *vector* would be the person who does the slapping, and *environment* would include the domestic situation and the societal norms or values that make such behaviour acceptable [144].

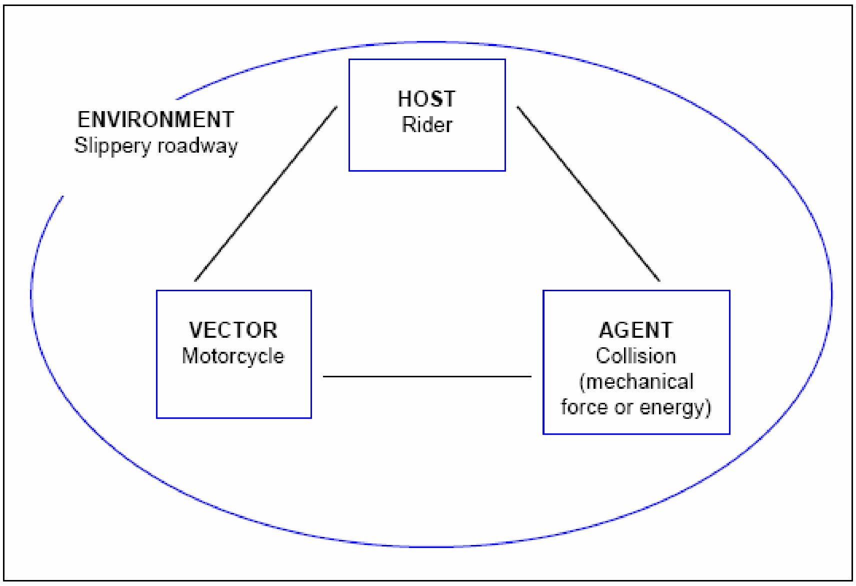


Figure 41. Epidemiological Model of Injury caused by Motorcycle Collision [144]

Use of a model of this type can help to identify all the factors involved in an injury. It also helps people to think about where they might intervene to prevent such injuries from happening in the future or to reduce the harm when they happen. For instance, in the motorcycle collision model in Figure 41, there may be things about the rider, the motorcycle, or the road that contributed to the crash. Perhaps there are things about motorcycle riders, motorcycles and/or road conditions that could be changed in order to prevent similar incidents in the future [144].

Table 5. Haddon's Matrix [144]

	Human (or Host)	Vector	Physical Environment	Socio-economic Environment
Pre-event	Is host pre-disposed or overexposed to risk?	Is vector hazardous?	Is environment hazardous? Does it have hazard-reduction features?	Does environment encourage or discourage risk-taking and hazard?
Event	Is host able to tolerate force or energy transfer?	Does vector provide protection?	Does environment contribute to injury during event?	Does environment contribute to injury during event?
Post-event	How severe is the trauma or harm?	Does vector contribute to the trauma?	Does environment add to the trauma after the event?	Does environment contribute to recovery?

Haddon's matrix has two dimensions, as presented in Table 5. The first is the three factors in the epidemiology of injury, introduced above, *human*, *vector*, and *environment*. The environment is often subdivided into *physical* and *socio-cultural*. The second dimension of the matrix is based on the fact that all the undesirable societal end-results of damaging interactions with environmental hazards are preceded by processes that naturally divide into three stages. The three stages are labelled as *pre-event*, *event*, and *post-event* [145]. The matrix can therefore be used to analyse any type of injury event and to identify interventions which might prevent such an event from happening again or which might reduce the harm by answering the questions in Table 5.

Table 6. Analysis of Motor Vehicle Collision using Haddon's Matrix [144]

	Human (or Host)	Vector	Physical Environment	Socio-economic Environment
Pre-event	Substance misuse, poor driving habits	Faulty brakes, bald tyres	Slippery road owing to rain	Social acceptance of high levels of alcohol use by males
Event	Not wearing seat belt	No airbag	Tree too close to the road	Ineffective enforcement of offences of driving under the influence of alcohol
Post-event	Elderly man, pre-existing medical condition		Slow emergency response, poor rehabilitation programme	Little help for reintegrating rehab patients into society

Table 6 illustrates the use of Haddon's Matrix to analyse a collision which happened when a male driver was returning home late one rainy night after attending a social event, where he had been drinking heavily. Neglecting to fasten his seat belt, he skidded and crashed into a tree at the edge of the road. There he remained until the driver of a passing vehicle stopped and took him to the nearest hospital. The injuries were made worse by improper handling of the injured man. At the time, however, taking him to hospital seemed a better alternative than waiting for an ambulance, given that the ambulance service was notoriously slow and unreliable [144].

As the *human*, *vector*, and *environment* factors in the *post-event* stage are about the trauma after an incident, they are excluded in the analysis of the fall records for this research about detection of the factors which may cause a toddler's fall injury. The *socio-economic environment* factor also is excluded owing to its irrelevance to the fall risk factors targeted in this research.

Table 7. Analysis of Toddler Fall Records in Lounge, Study, Living/Dining/Play Area

	Host	Vector	Environment
Pre-event	- Playing 284 - Running 209 - Toddling/Walking 79 - Holding object 44 - Pulled self up 19 - Climbed 19 - Dancing 15	- Playing with other 68	
Event	- Lost balance 26 - Fell onto body part 24 - Tripped over self 17 - Bit tongue/lip 12	- Tripped/Slipped 372 - Pushed/pulled 9 - Knocked over 7	- Banged/Landed on Table/Desk/Chair 244 TV/Speaker/Cabinet 80 Furniture 60 Floor 132 Wall/Door/Window/Steps 130 Fireplace/Radiator 118 Toy 58 Ornament/Pot/Jar/Cup/Tray 24 Stuff held 17 Person 2

Although many of the records of a toddler's fall in the home do not state every factor of Haddon's matrix, some outstanding information can be found in the massive number of stories and so can be identified as fall risk factors. Table 7 and Table 8 show the analysed results, numbering fall records with the same factor in Haddon's matrix. As can be seen from Table 7, many toddlers fell in lounge, study, living/dining/play area while moving around because they lost balance, tripped, or slipped. As a result, they banged on furniture (e.g. a table and a chair) or room structures (e.g. wall and fireplace), which may have been the direct cause of the severe injury, and then generally landed on the floor.

Table 8. Analysis of Toddler Fall Records in Inside Stairs

	Host	Vector	Environment
Pre-event	- Going up/down 288 - Playing 75 - Holding stuff 23 - Turned around 14 - Leaning onto gate 3	- Carried by person 40 - In baby walker/chair 4 - Person pushed 5	- Stair-gate not closed properly 41
Event	- Lost balance 60 - Bit lip/tongue 5	- Tripped/Slipped 122 - Dropped by person 36 - Person falling on 3	- Fell/Rolled/Banged on stairs 966 - Banged/Landed on Floor 129 Wall/Window/Door 42 Radiator 33 Stair gate 31 Banister/balustrade 17 Furniture 14 Box/vase/jug/clock/toy 9 Pram/buggy/bike 7

In inside stairs, many toddlers experienced a fall and got injured while going up or down the stairs since they lost balance, tripped, or slipped, as can be seen in the large numbers in Table 8. Some of them banged on room structures or furniture and landed or were found on the floor. The records also reveal that many parents did not witness the accident to their child, and some of them had not noticed the accident until they realised the child was not using a body part for days.

3.2.3 Risk Factor Identification

Based on the suggestions of child safety organisations to prevent falls, which were reviewed in Section 2.2, and the above analysis of fall records in the home environment, three fall risk factors are identified below for detection by vision-based analysis.

- Tripping or slipping hazards on the floor such as a toy or a spill
- A toddler, moving close to furniture or room structures
- A toddler, climbing furniture or stairs

Both the suggestions and the fall records point to tripping or slipping hazards as one of the main causes of a toddler's fall injury. Most of the organisations suggest removing spills or tripping hazards as shown in Table 1, and many of the falls in the records happened after being tripped or slipped as presented in Table 7 and Table 8. The fall records also indicate that a strike on a hard surface of furniture or room structures as a result of a fall is the main cause of the injury treated in hospital. Accordingly, toddlers, who are developing motor skills and can easily lose balance and fall down without any exterior influence, should not be near furniture or room structures without a caregiver next to them. Climbing is also a fall risk factor for toddlers because children can climb once they can crawl [49], and a fall from a higher level would cause a bigger impact and a worse injury. The child safety organisations advise discouraging climbing as presented in Table 1, and according to the records many toddlers fell while going up or down the stairs as shown in Table 8.

3.3 Evaluation of the Fall Risk Factors

The identified fall risk factors need to be confirmed if they are major factors to be continuously watched to prevent a toddler's fall injuries. Hence, the factors were evaluated by experts working in organisations for child safety through a questionnaire. Questionnaires are a way of getting information from people at a low cost in terms of time and money. They also allow the respondents to complete the form when they have time or to think or go and check on something if they need to [146].

3.3.1 Questionnaire

Questionnaires also have drawbacks. The response rate is typically low unless the questionnaire is seen as interesting and worth completing despite the time and effort expended to finish and return it. The respondents can also misunderstand the questions and the misunderstandings cannot be corrected [146]. In order to overcome those drawbacks, the questionnaire for the factor evaluation starts with explanation about the background and aim of this research, as shown in Appendix A. Such explanation may attract the attention of the respondents, who are specialised in child safety and advise on prevention of children's accidents, which is much related to the aim of this research. The explanation also presents the organisations whose resources were referred to for the identification of fall risk factors. This is because it was planned to send the questionnaire to employees of the organisations and the respondents were expected to be reasonably motivated by the questionnaire.

After the explanation about the research, people are asked to confirm each factor's significance in terms of continuous supervision to prevent a toddler's fall injuries in a home environment where enough safety products have been installed to satisfy any safety standards. This assumption about the home environment is to exclude any risk factor which can be simply prevented by use of a safety product. The factors are rephrased as the actions, suggested to perform for fall injury prevention, as presented in Appendix A. For instance, the first factor, *tripping or slipping hazards on the floor*, is changed to *keep floors clear of toys and other clutter which might trip toddlers when they walk around*. This is to avoid misunderstandings by the respondents because they would be people working in organisations which promote actions to prevent accidents. They are also asked to write down other situational or behavioural changes to be continuously watched if the identified factors are not thought to be enough.

The questionnaire was sent off by email to organisations related to child safety such as CAPT and RoSPA in the United Kingdom, MUARC of Australia, National Safety Council (NSC) of the United States of America, country members of Kidsafe, and ECSA of Europe.

3.3.2 Questionnaire Result

Fourteen responses were collected and three of them are displayed in Appendix B. The details of the overall results are shown in Table 9. Three respondents did not confirm the second fall risk factor, *a toddler moving around near furniture or room structures*. This was because sharp edges were not seen as a sudden change to the environment, and one respondent corrected this to *ensure that children play away from sharp edges or glass panels*. Spills were added as one of the other factors by the questionnaire respondents since only tripping hazards were indicated as the first fall risk factor in the questionnaire. The first factor in this thesis, however, also includes slipping hazards, as described in Section 3.2.

Table 9. Results of Questionnaire to Confirm Fall Risk Factors

Confirmation of Three Factors (Number of Confirmation/ Total Reponses)	Keep floors clear of toys and other clutter which might trip toddlers when they walk around	14/14
	Ensure there are no sharp or hard edges near them that could cause injuries when they fall	11/14
	Discourage children from climbing on furniture	13/14
Additional Factors	• Ensure spills are mopped up • Maintenance of stairs generally, repairing damaged carpets • Stairs should always be well-lit • Do not leave children sitting on chairs or tables unattended • Ensure footwear if worn is appropriate and fitted on the foot correctly • Avoid allowing toddler to drink whilst walking	

Some of the factors added by the respondents are related to a toddler's behavioural or environmental changes, but they are very tardy changes or can be prevented by not performing some actions. Hence, they are excluded from this research, which intends to detect fall risk factors related to sudden changes of a toddler left alone under conditions initially thought to be safe. For instance, stairs and carpets need to be repaired every so often to keep them safe for children. A toddler should not be left on chairs or tables unattended, put shoes on correctly, and be allowed to drink only in front of the caregiver.

Therefore, the three identified risk factors of a toddler's fall injury in the home environment were confirmed to be major factors requiring continuous supervision for prevention of the fall injuries.

3.4 Chapter Summary

In order to achieve the first objective of this research, *the identification of key risk factors of a toddler's fall injuries in a home environment*, this chapter analysed around two thousand fall records. The records are about fall incidents of toddlers, aged between one and three, in a home environment, who sought treatment at a hospital from 2000 to 2002 in the United Kingdom. An epidemiological injury framework, Haddon's matrix, was used to analyse the massive number of records.

As the research presented in this thesis aims to aid a caregiver's supervision to prevent a toddler's fall injuries, the risk factors were filtered to be related to behavioural or environmental changes which require continuous supervision, by excluding the factors preventable by safety products. The final fall risk factors were identified in the light of the suggestions and the fall records as follows:

- Tripping or slipping hazards on the floor such as a toy or a spill;
- A toddler moving close to furniture or room structures;
- A toddler climbing furniture or stairs.

The identified factors were evaluated through a questionnaire from experts in children's home accidents, and it was confirmed that the factors are major fall risk factors, preventable by continuous supervision of a caregiver. There are previous attempts to identify fall risk factors, but they are generally for elderly people [147-

153] or patients [154-158]. Studies on young children's fall risks investigated rates of fall injuries by 3-month intervals [159] or the relationship between height of fall and long bone fracture [160]. Meanwhile, the fall risk factors identified here can be detected and eliminated by supervision and thus are different from previously identified factors.

The following chapters describe vision-based analysis methods, developed to recognise the fall risk factors, and to meet the aim of this research, which is to aid a caregiver's supervision of a toddler.

Chapter 4
Classification of Human and Clutter

4.1 Introduction

In order to detect the identified fall risk factors based on image analysis, it is essential to classify segmented foreground objects into human and non-human. This is because the first fall risk factor, *a tripping or slipping hazard on the floor*, is about the existence of clutter on the floor, and the other factors, *toddlers moving near* or *climbing furniture*, are related to human behaviour when the toddler is the only human on the scene.

Section 4.2 presents the whole flow of the methods developed in this research to detect the risk factors, including a novel method for classification of a human and clutter. Section 4.3 delineates the details of the whole methods, and Section 4.4 presents implementation and evaluation of the methods. The first stage of the methods involves an easy interface to let the user manually select the floor area after installing a single fixed camera so as to focus on the floor area to detect the first factor, *a tripping or slipping hazard*. Second, as the differentiation of human and clutter is based upon the irregular motions inside a human region owing to the different motions of body parts, the diversity of the internal motion vectors of a human contour is calculated and tested to find the threshold to separate human and non-human. The rest of the methodology section describes the detection of clutter on the floor, which is the first fall risk factor, *a tripping or slipping hazard*, and a tracking method for detecting the other factors, *a toddler moving near* or *climbing furniture or stairs*.

4.2 Method Overview

This section briefly explains how the vision-based methods work to detect the fall risk factors identified in Section 3.2, including the novel method to classify foreground objects into a human and clutter using a dynamic motion cue. A brief flowchart of the methods is presented in Figure 42.

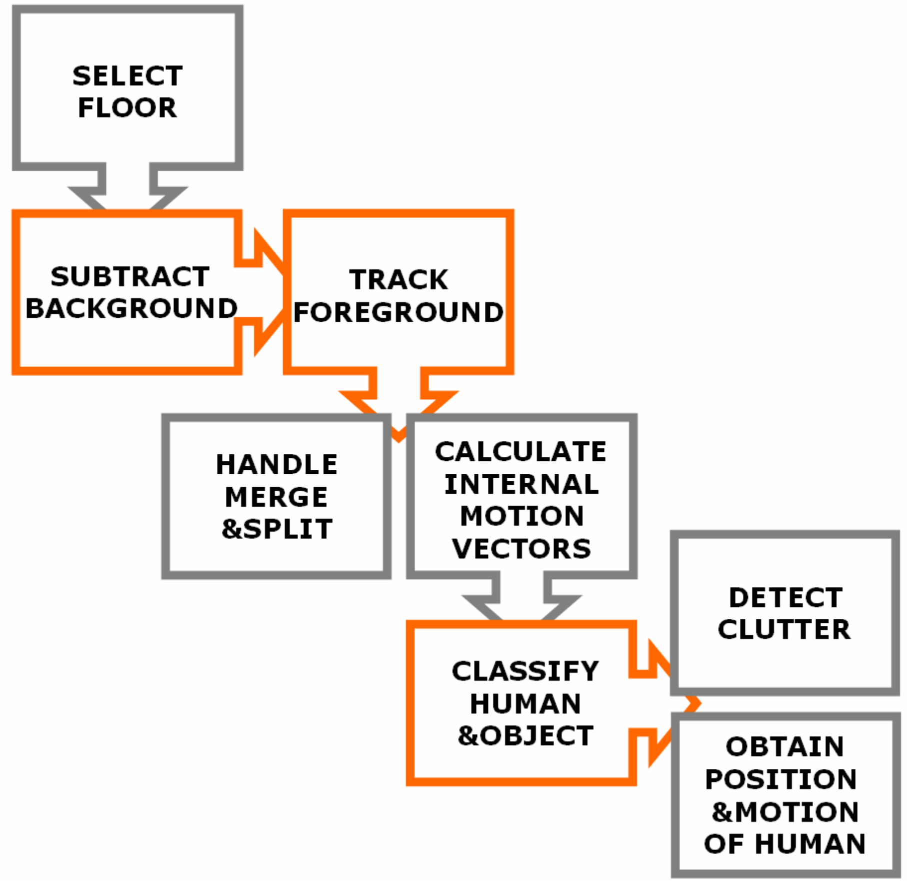


Figure 42. Method Flowchart for Classification of Human and Clutter

The major tasks of this vision-based analysis are background subtraction, foreground tracking, and foreground classification into a human and clutter, as orange-lined in Figure 42. These are for detection of any clutter, which may be a tripping or slipping hazard, and obtaining a human's motion and position information, which can be used to check a toddler moving close to or climbing furniture or stairs.

Before real-time image analysis begins, the floor area is manually selected to limit the area for detecting clutter on the floor. The floor area is also used to estimate the relative position of a toddler to furniture or stairs, on the assumption that non-floor area is filled with furniture and room structures. The foreground regions in real-time images are segmented by a simple background subtraction method and tracked by connection of the closest region centres between consecutive frames. As a single object can correspond to multiple regions and multiple objects can correspond to one region owing to occlusions or errors in the background subtraction, those regional merges and splits are handled during the tracking. Although the main purpose of the foreground tracking is to estimate the motion and position of a toddler, the tracking is preceded by the human classification owing to the use of a motion-related cue in the classification. In Section 4.3 on methodology, the tracking is described after the human classification.

Once all foreground regions are segmented, a toddler needs to be separated from clutter, which may be toys that he or she plays with. This classification uses different moving characteristics of a human and clutter in an indoor home environment. While a toddler has irregular internal motion vectors owing to the different motions of body parts when the whole body is mobile, indoor clutter may have relatively constant motion vectors. Hence, a toddler is detected by calculation of the similarity of the motion vectors in each foreground region. The algorithms are detailed in the next section.

4.3 Methodology

This section describes the details of the methods developed to detect the fall risk factors. As an initial setting, the floor area is manually selected and the method is detailed in Section 4.3.1. During real-time image analysis, the foreground objects are segmented and classified into a human (Section 4.3.2) and clutter (Section 4.3.3), and the human is tracked over time (Section 4.3.4).

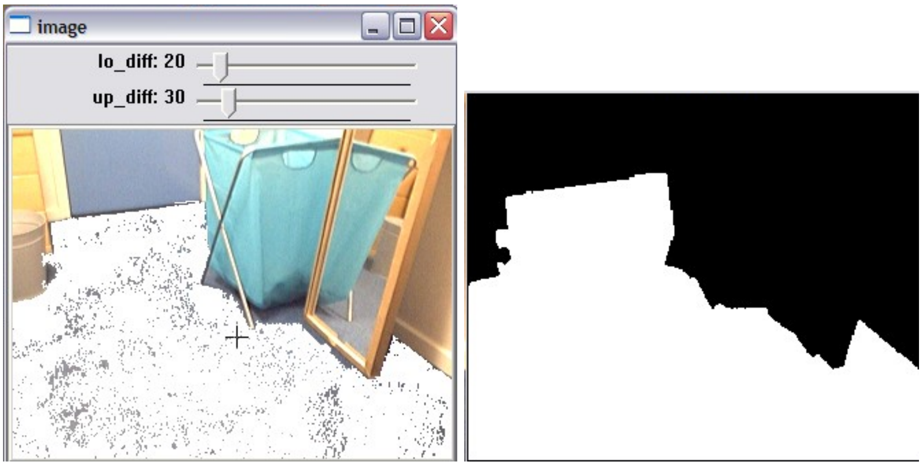
4.3.1 Floor Selection

In order to limit the detection range of clutter, which may become tripping or slipping hazards, to the floor area, the floor should be identified in the captured images. As the camera is fixed, the floor area will hardly be changed, and its single detection of the floor area will be enough. Therefore, the floor area is to be manually selected at a single when the camera is set up.

After a background image is captured when there is no foreground object in the environment, the FloodFill method helps the user select the floor region in the image. The FloodFill method fills neighbouring pixels, whose values are close to the pixel clicked by the user. The pixels will belong to the repainted domain if their value *v* meets the following condition:

$$v_0 - \delta_{lw} \le v \le v_0 + \delta_{up} \, . \qquad (1)$$

v_0 is the value of one of the pixels in the repainted domain, which begins with the selected pixels [161]. $\delta_{lw,}$ the maximal lower difference, and δ_{up}, the maximal upper difference between the pixels, can be defined by the user with the sliding bar controls in Figure 43a. In this way, the user can select the floor area with several clicks. As the selected area gathers lots of tiny chinks, when the user submits the floor-selected image, a mask image is returned with filled contours of the selected area, as presented in Figure 43b.



lo_diff: 20

up_diff: 30

(a) (b)

Figure 43. Floor Selection

4.3.2 Human Classification

A. Background Subtraction

In order to segment foreground objects before classifying them into a human and clutter, a background subtraction method is used. Background subtraction basically finds the difference between the current image and the background image. As reviewed in Section 2.3.1G, background subtraction is the most commonly-used method to segment motion for human detection. In this work, it was decided to use a background subtraction method, not only because it is popular but also because stationary clutter, a tripping hazard, cannot be found by other methods such as successive image differentiation or colour detection owing to its stillness or lack of constant colour.

In the background subtraction method adopted here, a simple background model ($bgMean_{(x,y)}$) is built up when the floor area is clear, by accumulation of several frames (N) and calculation of the mean value of each summed pixel ($bgSum_{(x,y)}$) to get their mean brightness.

$$\begin{aligned} bgMean_{(x,y)} &= bgSum_{(x,y)}/N, \\ bgDiff_{(x,y)} &= abs(bgMean_{(x,y)} - Cur_{(x,y)}). \end{aligned} \tag{2}$$

Then the absolute differences (*bgDiff*$_{(x,y)}$) between the background model and the current image (*Cur*$_{(x,y)}$) are then calculated by pixels, as shown in Equation (2).

In order to eliminate noise in the results of the background subtraction, differences smaller than a threshold value are returned to zero, and a binary image is created by returning the others to 255, as detailed in Equation (3).

$$bgDiff_{(x,y)} = \begin{cases} 255, & if\ diff_{(x,y)} > threshold \\ 0, & otherwise. \end{cases} \tag{3}$$

Whenever this binary image becomes null, the background model is updated to cope with slight changes of sunlight that are ignored by thresholding. For the dramatic lighting changes such as turning on/off a lamp, the background model is also updated when the differences are similar all over the image, as shown in Figure 44, on the assumption that there is no spot light but a ceiling fixture that lights the whole room.

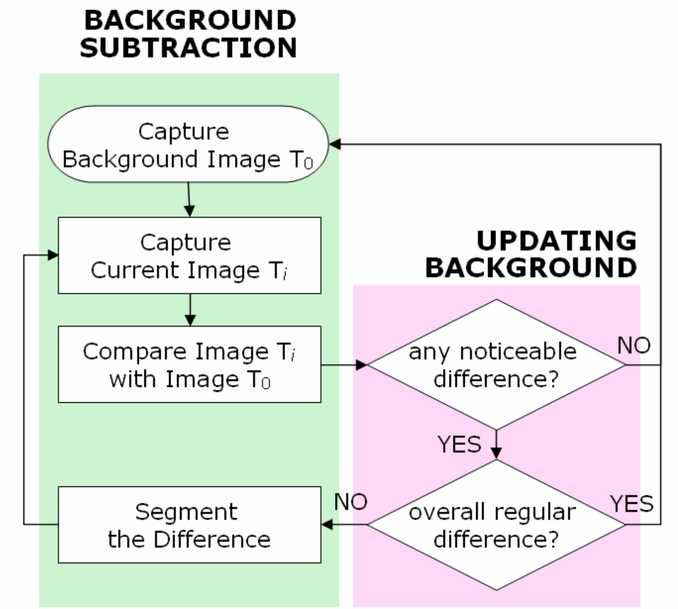

BACKGROUND SUBTRACTION
Capture Background Image T0
Capture Current Image Ti
Compare Image Ti with Image T0
Segment the Difference
UPDATING BACKGROUND
any noticeable difference?
NO
YES
overall regular difference?
NO
YES

Figure 44. Background Subtraction Flowchart

B. Similarity of Motion Vectors

The studies reviewed in Section 2.3.1 use diverse cues to differentiate between a human and a non-human object in the images. Most of the cues, however, are related to human appearance such as geometric knowledge, skin colour, and silhouette and are therefore not very reliable with respect to occlusions. In this work, a dynamic motion cue is used to classify a human body. The cue is based upon the irregular motions inside a human region owing to the different motions of body parts.

In order to capture the different internal motions, some features that are good to track are detected within each ROI, which is a bounding box of each noticeable background-subtracted region. Such features are actually the corner points that have relatively big eigenvalues in the pixels and are at a satisfactory distance from one another [161]. The detected features are tracked by calculation of the optical flow between every two successive frames for each feature, by means of the method proposed by Lucas and Kanade [162]. This is because the optical flow method is sensitive to small movements even in the case of low contrast owing to its simultaneous consideration of spatial and temporal changes [163]. Barron and colleagues [164] also found that Lucas and Kanade's method was the most reliable among other optical flow methods.

$$I(x_1, y_1, t_1) = I(x_2, y_2, t_2)\,. \tag{4}$$

The method is used to calculate the displacement of a pixel in two consecutive frames, assuming that the brightness of the pixel belonging to a moving object remains fixed in the consecutive frames. This assumption is mathematically translated into Equation (4), when I, x_i, y_i, and t_i for i=1,2, denote brightness, 2D spatial coordinates, and time of a pixel in the first and second image frames respectively.

$$\begin{aligned} x_2 &= x_1 + \delta x, \\ y_2 &= y_1 + \delta y, \\ t_2 &= t_1 + \delta t. \end{aligned} \tag{5}$$

The relations of x_2, y_2, and t_2 to the correspondent parameters of the previous frame are defined in Equation (5), where δx and δy are the spatial differences and δt is the time difference between two successive frames.

$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + \ldots. \tag{6}$$

$$\frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + \ldots = 0\,. \tag{7}$$

Then $I(x_2, y_2, t_2)$ can be expanded to Equation (6) by a Taylor series, and it generates Equation (7) to satisfy Equation (4).

When u and v are the speeds of the pixel moving in the x and y directions respectively, Equation (7) becomes Equation (8), which is called the optical flow constraint equation. This is how to track each feature over frames.

$$\begin{aligned} &\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v = -\frac{\partial I}{\partial t}, \\ &u = \frac{\delta x}{\delta t}, v = \frac{\delta y}{\delta t}. \end{aligned} \tag{8}$$

If any features detected by the optical flow calculation get out of any ROIs found on the same frame, the features are discarded to focus on the ROIs. Whenever there are fewer than five features left within a ROI, the feature detection is executed anew in the ROI to avoid capturing very few motions in one region.

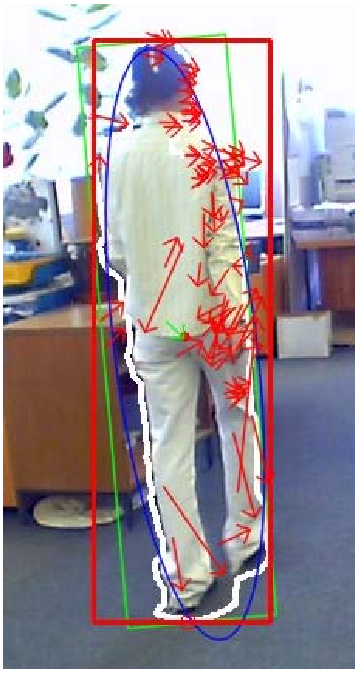


Figure 45. Internal Motion Vectors

As a result, the relation between one feature's coordinates and its new position, detected on the next frame, is presented as an arrow, indicating a motion vector, and one ROI gets multiple motion vectors, as shown in Figure 45. Therefore, using the dot product of any two vectors, $a \cdot b = a_x \times b_x + a_y \times b_y = \| a \| \cdot \| b \| \cos\theta$, the similarity of the motion vectors in one ROI is calculated over two adjacent frames.

$$avg\_\cos\theta_c = \\ [\sum_{i=0}^{n-1} \sum_{j=i+1}^{n} \{(a_{cx}^{i} \times a_{cx}^{j}) + (a_{cy}^{i} \times a_{cy}^{j})\} / \| a_c^i \| \cdot \| a_c^j \|] / \sum_{k=1}^{n} k. \tag{9}$$

When all the vectors in the c^{th} ROI are defined as $a_c^0, a_c^1, ..., a_c^n$, the average of *cosθ* between the vectors, *avg_cosθ_c*, can be calculated with Equation (9). As the vectors are same-directional when *θ* is zero, the closer to one *avg_cosθ$_c$* is, the closer the similarity of the vectors is. The threshold value to classify human and clutter is defined after tests in Section C.

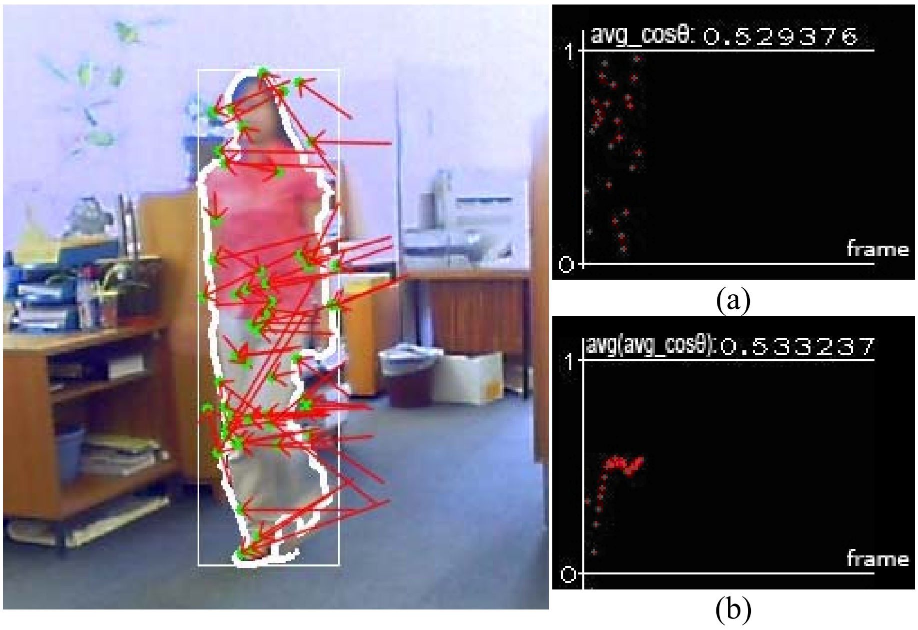
C. Test to Set Threshold

Several tests have been carried out to detect the threshold value to classify a human and clutter with an adult, a ball, and a radio-controlled model car, which represent human, rolling, and straight motions respectively. It is advised by Hamleys, one of the largest toy shops in the world, that other toys which move more dynamically such as dancing robots or realistic animal toys, are for children over three years old, who are not toddlers any more [165]; they were not used in these tests.

The averaged value of *cosθ* between every two vectors within one ROI (*avg_cosθ_c*) was fairly dynamic over the frames for a walking human (Figure 46a) and was constantly close to one for a rolling ball (Figure 47a) and a radio-controlled model car (Figure 48a).

$$avg(avg\_\cos\theta_c^t) = (\sum_{i=1}^{i=t} avg\_\cos\theta_c^i)/t\,. \tag{10}$$

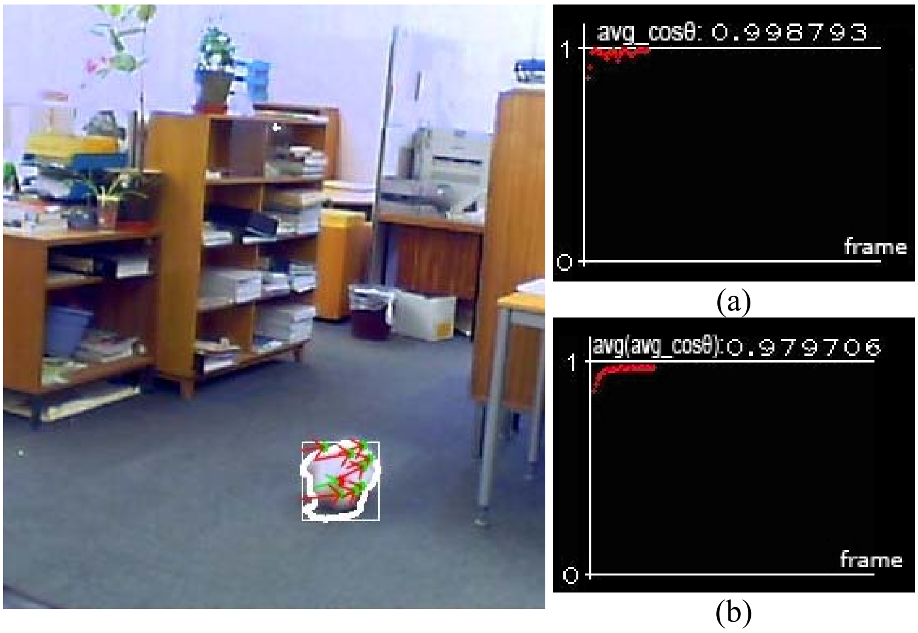
As sometimes the $avg\_cos\theta_c$ value became very close to one for a human motion and somewhat lower than one for clutter motion, all the $avg\_cos\theta_c$ values from the past frames($avg\_cos\theta_c^i$ when i = 1,2,…,t) were also averaged at every frame, as given in Equation (10), where t is the current frame number. On the basis of several tests, it was found that the $avg(avg\_cos\theta_c^t)$ value stays under 0.75 for a walking human (Figure 46b) and over 0.9 for a rolling ball (Figure 47b) and a moving model car (Figure 48b). Hence, the threshold for classifying human and non-human was defined as 0.8.



(a)

(b)

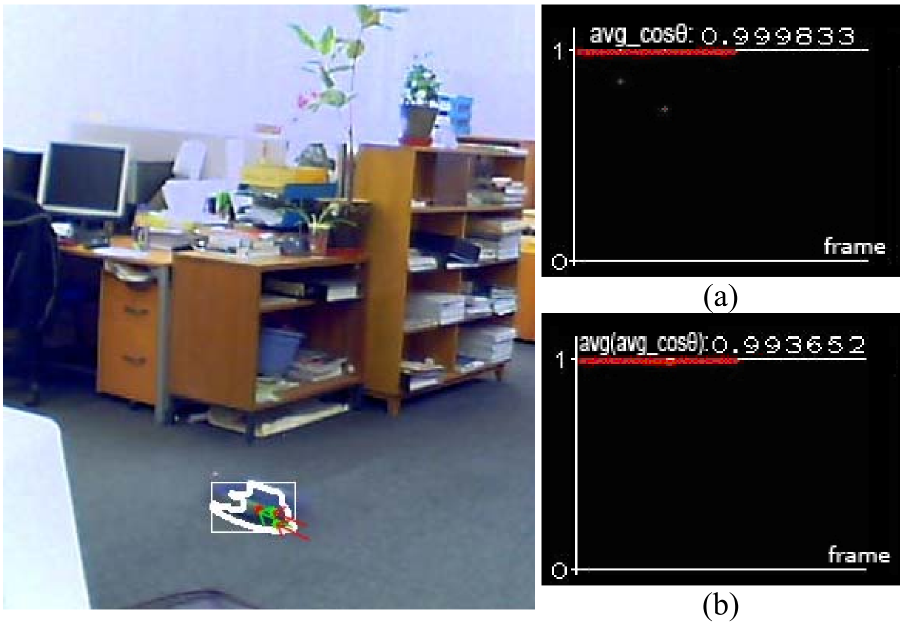
Figure 46. Internal Motion Vectors of Walking Human



(a)

(b)

Figure 47. Internal Motion Vectors of Rolling Ball



(a)

(b)

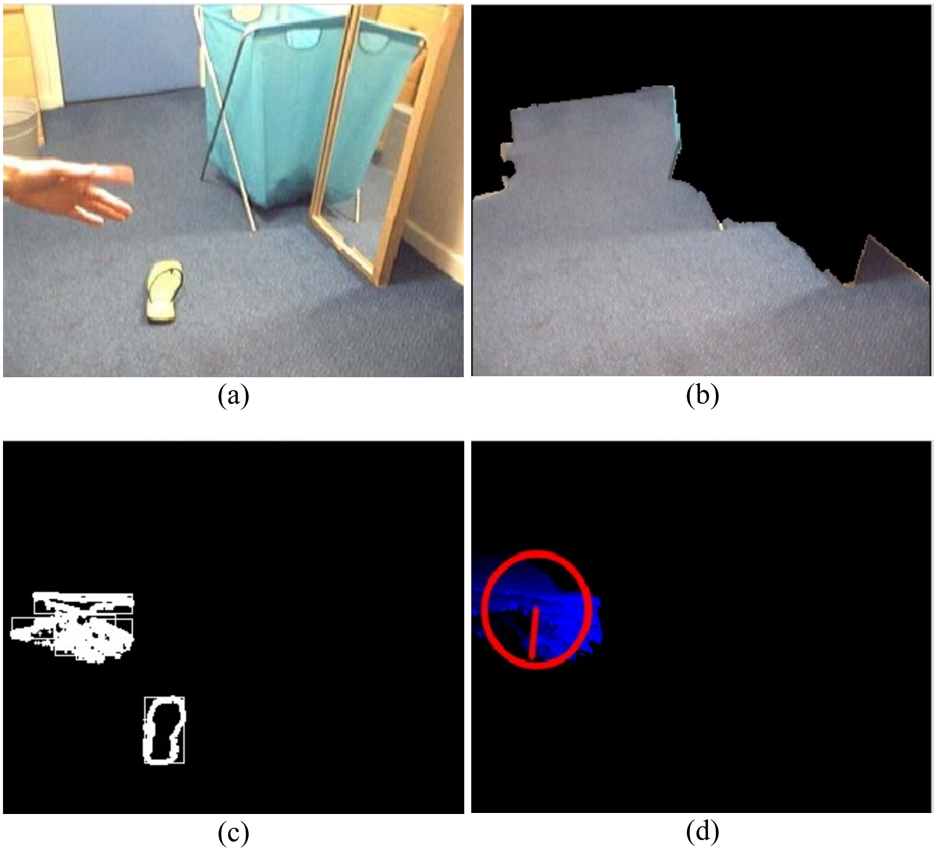
Figure 48. Internal Motion Vectors of Radio-Controlled Model Car

4.3.3 Detection of Clutter on the Floor

Since clutter could become a tripping or slipping hazard on the floor, anything classified as clutter using the method in Section 4.3.2 gets filtered to be at a standstill on the floor in order for clutter to be detected.

$$curDiff_{(x,y)} = abs(Cur^{t}_{(x,y)} - Cur^{t-1}_{(x,y)})\,. \tag{11}$$

For this, any noticeable motion is detected at first by comparison of every two consecutive images, $Cur^{t}_{(x,y)}$ and $Cur^{t-1}_{(x,y)}$ by pixels, as given in Equation (11) where t is the current frame number. Then every ROI, bounding a background-subtracted region, within the floor area, which is selected with the method described in Section 4.3.1, is checked for motion inside. Any ROI without motion is considered as still clutter, being a tripping or slipping hazard on the floor. Figure 49 shows an instance of differentiating a waving hand and a standstill slipper on the floor. While a slipper and a hand both appear on the background-subtracted image within the floor area (Figure 49c), the difference between successive frames shows only the moving hand (Figure 49d). For cases of clutter without tripping or slipping hazards such as toys with which a sitting toddler plays, the movement of the subject toddler is checked at the same time.


(a) (b)

(c) (d)

Figure 49. Detection of Clutter on Floor

4.3.4 Toddler Tracking for Status Recognition

In order to detect the second fall risk factor, *a toddler moving around near furniture or room structures* and the third one, *a toddler climbing furniture or stairs*, all the foreground objects need to be tracked individually. This is because tracking not only provides information about movement but also links identical foreground objects over the frames, so that information captured from previous frames can be saved for the identical foreground objects. The human classification is based on dynamic motion vectors from past frames, as detailed in Section 4.3.2C, and the past information should be saved separately for each foreground object to classify multiple foreground objects at the same time.

First, regions which are background-subtracted from each current image are focused individually to detect the contour of each foreground region and the centre of mass (x_c,y_c), as given in Equation (12). In it, $I(x,y)$ is the intensity value of pixel in the position (x,y) of the image, where each contour is drawn [161], and the red dot in Figure 50 is a resulting centre of mass of the human contour.

$$\begin{aligned} x_c &= \sum_x \sum_y xI(x,y) / \sum_x \sum_y I(x,y) \\ y_c &= \sum_x \sum_y yI(x,y) / \sum_x \sum_y I(x,y) \end{aligned} \tag{12}$$

The coordinates of the centre of mass found on each region's contour from one frame are saved in order to be connected to the centre of mass of its corresponding region's contour on the next frame. The distances between a centre of mass from a frame and all the centres from the previous frame are calculated, and the centre is connected to the closest one from the previous frame. This connection is separately conducted on every contour centre detected on each frame. The speed and direction of each contour are calculated for motion information, by means of the coordinates of two connected centres over two consecutive frames, and the green arrow in Figure 50 displays the connection.

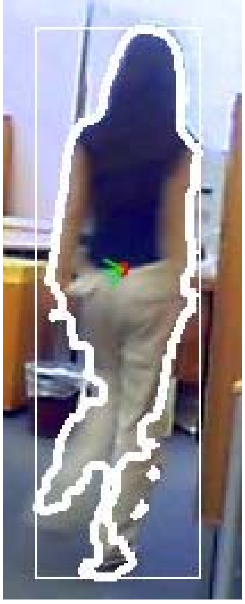


Figure 50. Tracking of Centre of Mass

Whereas the centre of mass of a background-subtracted region is used to obtain the motion information, the vertically lowest point of a toddler’s region contour becomes the focus to check the position of each human. As toddlers barely can jump, the vertically lowest point of the contour is considered to be where the toddler stands on the floor. As this image analysis needs to check if a toddler moves near or climbs furniture, the lowest point is checked for every frame to see if it is close to the boundary of the floor area detected when the camera is installed, or if it gets out of the floor area. This is based on the consideration that the non-floor area is filled with furniture or room structures. In Figure 51, the human region is bounded by a red box, which indicates it is classified as human, while the ball region is not. The shortest distance of the lowest point of the human contour from the blue boundary of the floor area is also calculated and displayed in red numbers underneath, as shown in Figure 51.

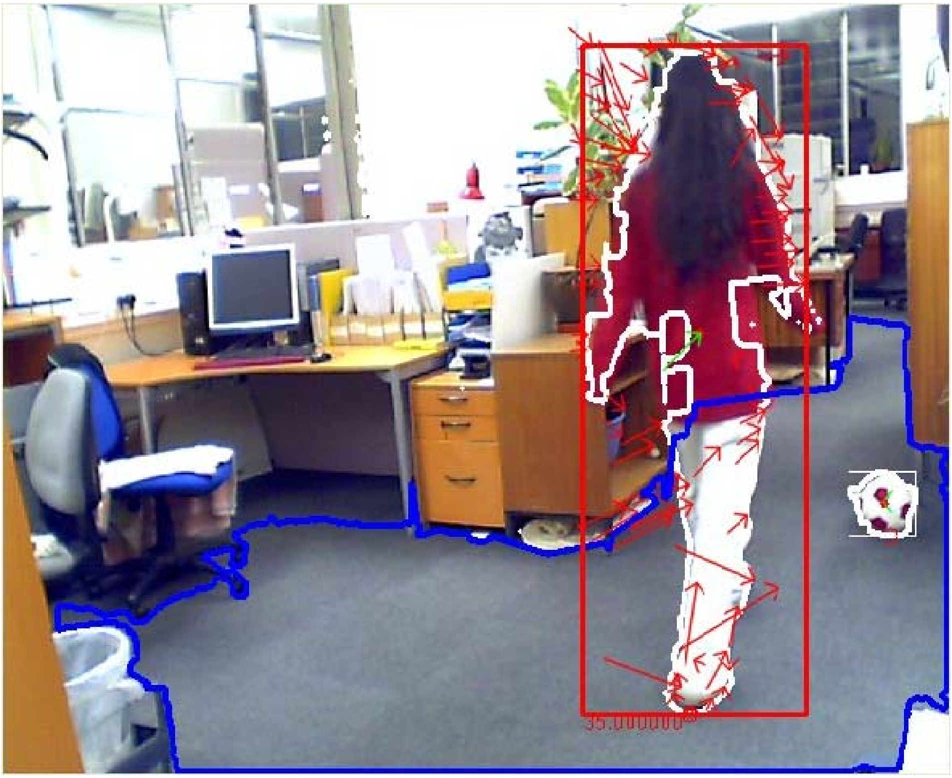


Figure 51. Position Relative to Floor Area

The motion and position information estimated by the above methods is based on the image plane and therefore could be different from the actual information in the 3D world space. The motion information of a toddler, however, is simply used to exclude a toy with which a sitting toddler plays, when clutter as a tripping or slipping hazard is detected, as mentioned in Section 4.3.3. The position information is also necessary to be relative to the boundary of the floor area, and the normal child's room, the target environment in this research, would not be big enough to have large differences between the image plane and the world space in the relative position information. Accordingly, the motion and position information obtained from the image plane is used for the recognition of a toddler's status.

A. Handling of Merges and Splits

In the camera view, multiple objects can appear to be one single connected component, and a single object can split into multiple regions owing to occlusions or errors in background subtraction. These merging or splitting phenomena can confuse the tracking of each foreground object, which is to match identical objects over consecutive frames. Therefore, regional merges and splits should be recognised and managed. According to Section 2.3.2, existing studies generally use the colour and position information of each target to handle regional merges and splits. The use of colour information, however, would fail to manage merges and splits of regions corresponding to a single object in various colours, or multiple objects in similar colours since it separates individuals according to the colour similarity. The work in this research does not need to identify a toddler and each piece of clutter but to distinguish a toddler from clutter in an indoor home environment, which is fairly restricted. Therefore, this research seeks a simple position-based method to handle visual merges and splits in images.

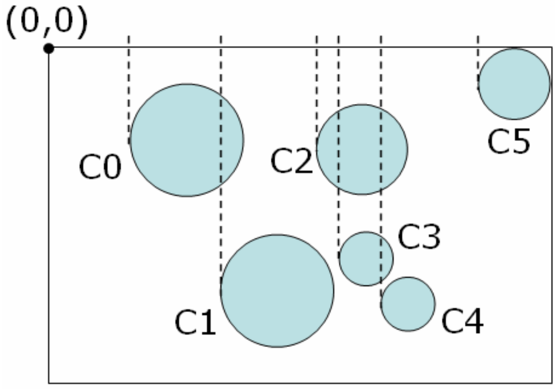
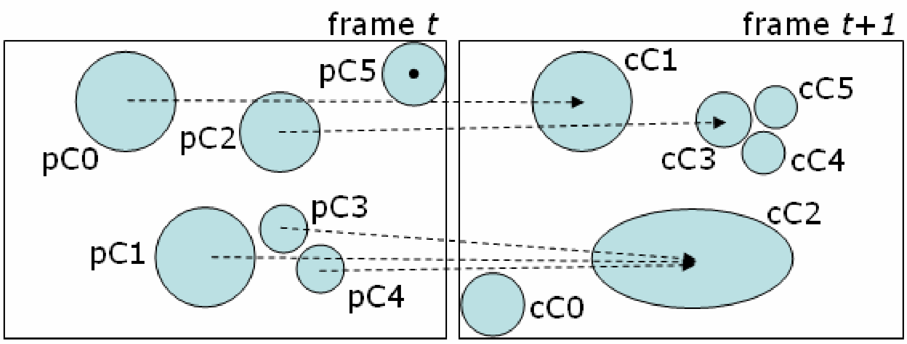
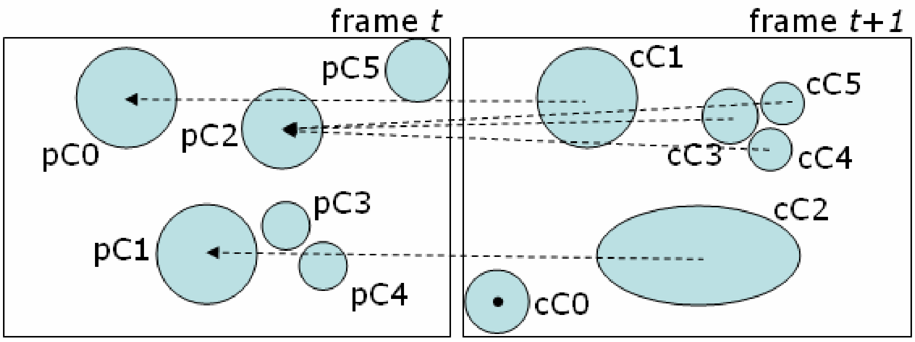


Figure 52. Contour Indexing

Initially, every foreground region on each frame is ordered by the smallest x-coordinate of its contour and numbered by the order. Figure 52 shows an example of the numbering. All the information obtained from a region such as the coordinates of its centre and bounding box, is also tagged with the region's number and kept over every two successive frames to be used in connecting identical foreground objects over the two frames. As the numbering of foreground regions is conducted anew on every frame, an object can get a different number on the next frame owing to regional merges and splits as well as simple position changes. In order to connect correct regions for an identical object over two consecutive frames despite regional merges and splits, the closest centre detection is carried out twice, from the previous frame to the current frame and vice versa.



(a)



(b)

Figure 53. Double Detections of Closest Centres over Two Frames

For instance, when regions merge, split, and move in and out at the same time, as shown in Figure 53, each contour centre on frame *t* is connected to the closest among the contour centres detected on frame *t+1* (Figure 53a), and the reverse connection from frame *t+1* to frame *t* is conducted (Figure 53b). In order to prevent any wrong connection owing to an appearance or a disappearance, which does not have any identical region to be connected on the previous frame or the current frame, the distance between the closest centres over two frames is limited within the half length of the diagonal line, connecting two opposite points of the bounding box of each region. This is because the moving speed of a toddler is assumed to be slow enough to catch up within the limitation at the rate of 30 frames per second.

These two different connections are compared in order to check where a merge, a split, an appearance, a disappearance, or a one-to-one connection happens. The one-to-one connection, which means tracking of a region without any merge or split, is confirmed when the two connections are both singular. For example, when all the regions in frame *t* are indexed with pC0, pC1, …, pC*n* and the regions in frame *t+1* are indexed with cC0, cC1, …, cC*n*, only pC0 and cC1 have single connections in both directions, as shown in Figure 53a and Figure 53b. In this case, all the recorded information for one foreground object tagged with zero gets indexed with one. The information includes the averaged *cosθ* value of internal motion vectors of a region, described in 4.3.2B, in order to keep and use all the past values for human classification.

When one region in the current frame has multiple connections in the closest centre detection from the previous frame, like cC2 in Figure 53a, it is viewed as a merge, and a region in the other way around like pC2 in Figure 53b is viewed as a split. These merges and splits need to be further examined to differentiate between occlusion of multiple objects and separated blobs of one object. As a split should happen before a merge in a single object and a merge should come before a split in multiple objects, the differentiation is based upon a history of merges and splits for each region. This is based on the assumption that each foreground object corresponds to a single region without any merge or split when appearing on the scene.

When a single object splits into multiple regions, all the split regions inherit the past information tagged to the region before the split. When they merge afterwards, the information of any region before the merge is transferred to the merged region. When multiple objects merge into one connected region, all the past information of each region before the merge is saved individually with the region's size, and the merged region starts with null information unless the multiple objects include a toddler. If a toddler is included there, the toddler region's information is kept on the merged region because a toddler carrying toys, for instance, needs to keep being tracked as a toddler for fall risk detection. Then when any of the objects gets separated from the merge, among the pieces of information saved before the merge, a correct piece is returned to the split object by comparison of its regional size with the saved regional sizes. The comparison of region sizes allows a 10 per cent error margin.

A region with no connection in the closest centre detection from the previous frame, like pC5 in Figure 53a, is regarded as an object's disappearance, and the region's information is removed. A region with no connection in the opposite way, like cC0 in Figure 53b, is regarded as an appearance and begins a new data collection.

4.4 Implementation

For the visual analysis to detect the fall risk factors, a single Logitech Quickcam Pro5000 was used to capture real-time images in a fixed position. The image size is 640 x 480 pixels, the frame rate is 30 frames per second, and the developed prototype software has dialogue-based interfaces to set up and control the system. The image analysis methods were programmed in Visual Studio C++.NET using

OpenCV[5], an open source computer vision library developed by Intel Corporation. The whole workflow of the real-time image analysis for the fall risk detection is illustrated in Figure 54.

When there is no foreground object in the environment where a fixed camera is installed and a subject toddler will play alone, a background image is captured by pressing button *Capture Initial Background* on the window in Figure 55, before selection of the floor area. Then the captured image appears on screen *Initial Background*.

A window of the background image like the one in Figure 43a pops up when button *Detect the Floor* is pressed, and the window lets the user select the floor region in the initial background image, as described in Section 4.3.1. The floor selection works well, even when there is more than one separate region corresponding to the floor in the background image. This is because the contour of each region is detected and filled respectively. As the floor is detected only once at the beginning, if any structure in the room moves during the clutter detection, the floor mask image should be updated manually, and it can be done with the easy interface described in Section 4.3.1.

When button *Build BG Model* is pressed, a background model is built to perform background subtraction, as described in Section 4.3.2A. Then the whole real-time image analysis can begin when button *Start Supervision*, which converts to *Stop Supervision* while real-time images are being processed and can be pressed to finish.

[5] The installation files of OpenCV can be downloaded from http://sourceforge.net/projects/opencvlibrary [accessed: 30 Jan. 2009].

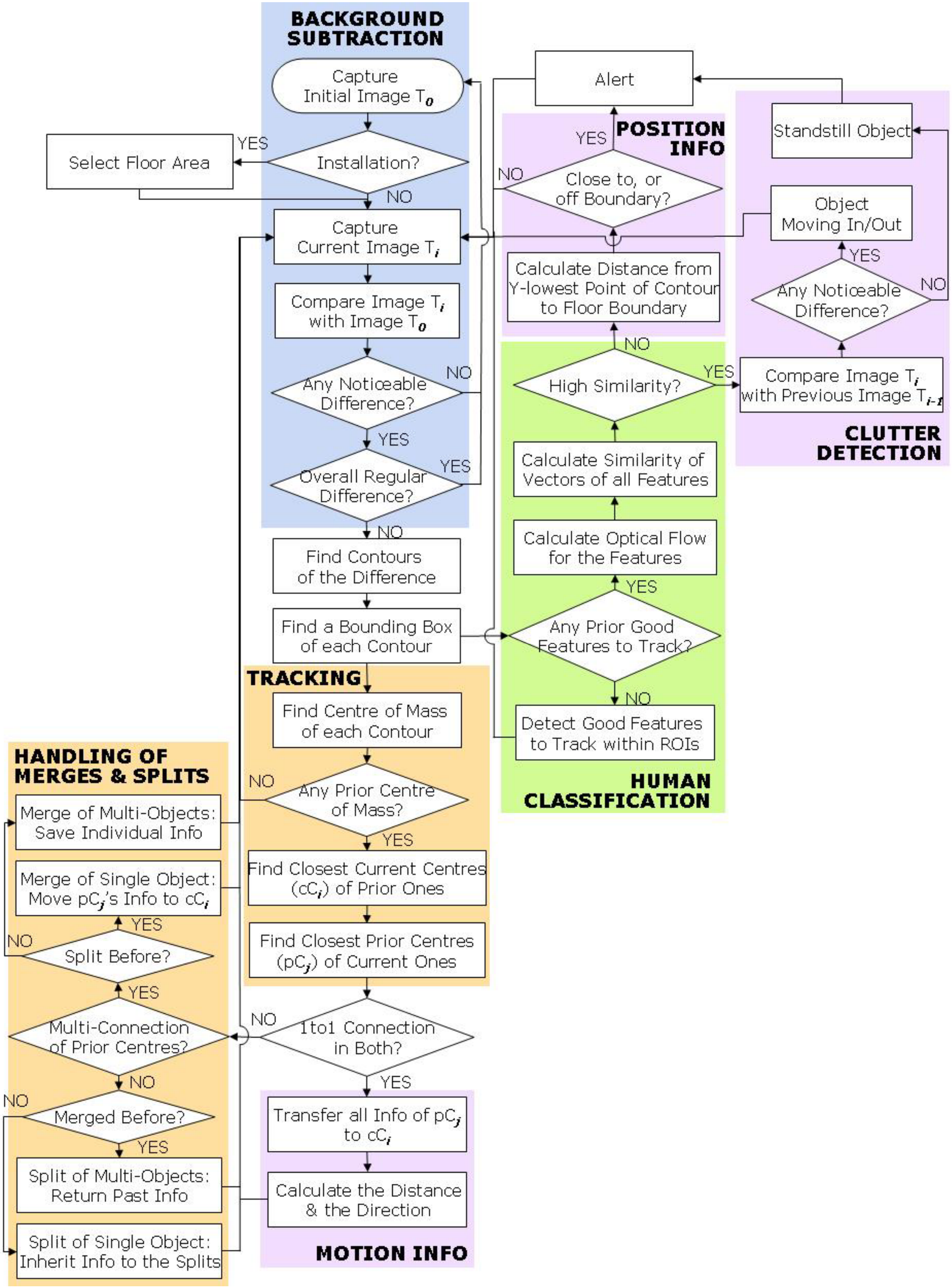

BACKGROUND SUBTRACTION
Capture Initial Image T_0
Installation?
YES
Select Floor Area
NO
Capture Current Image T_i
Compare Image T_i with Image T_0
Any Noticeable Difference?
NO
YES
Overall Regular Difference?
YES
NO
Find Contours of the Difference
Find a Bounding Box of each Contour
TRACKING
Find Centre of Mass of each Contour
Any Prior Centre of Mass?
NO
YES
Find Closest Current Centres (cC_i) of Prior Ones
Find Closest Prior Centres (pC_j) of Current Ones
1to1 Connection in Both?
NO
YES
Transfer all Info of pC_j to cC_i
Calculate the Distance & the Direction
MOTION INFO
HANDLING OF MERGES & SPLITS
Merge of Multi-Objects: Save Individual Info
Merge of Single Object: Move pC_j's Info to cC_i
YES
NO
Split Before?
YES
Multi-Connection of Prior Centres?
NO
NO
Merged Before?
YES
Split of Multi-Objects: Return Past Info
Split of Single Object: Inherit Info to the Splits
Alert
POSITION INFO
YES
NO
Close to, or off Boundary?
Calculate Distance from Y-lowest Point of Contour to Floor Boundary
NO
High Similarity?
YES
Calculate Similarity of Vectors of all Features
Calculate Optical Flow for the Features
YES
Any Prior Good Features to Track?
NO
Detect Good Features to Track within ROIs
HUMAN CLASSIFICATION
Standstill Object
Object Moving In/Out
YES
Any Noticeable Difference?
NO
Compare Image T_i with Previous Image T_{i-1}
CLUTTER DETECTION

Figure 54. System Workflow for Classification of Human and Clutter

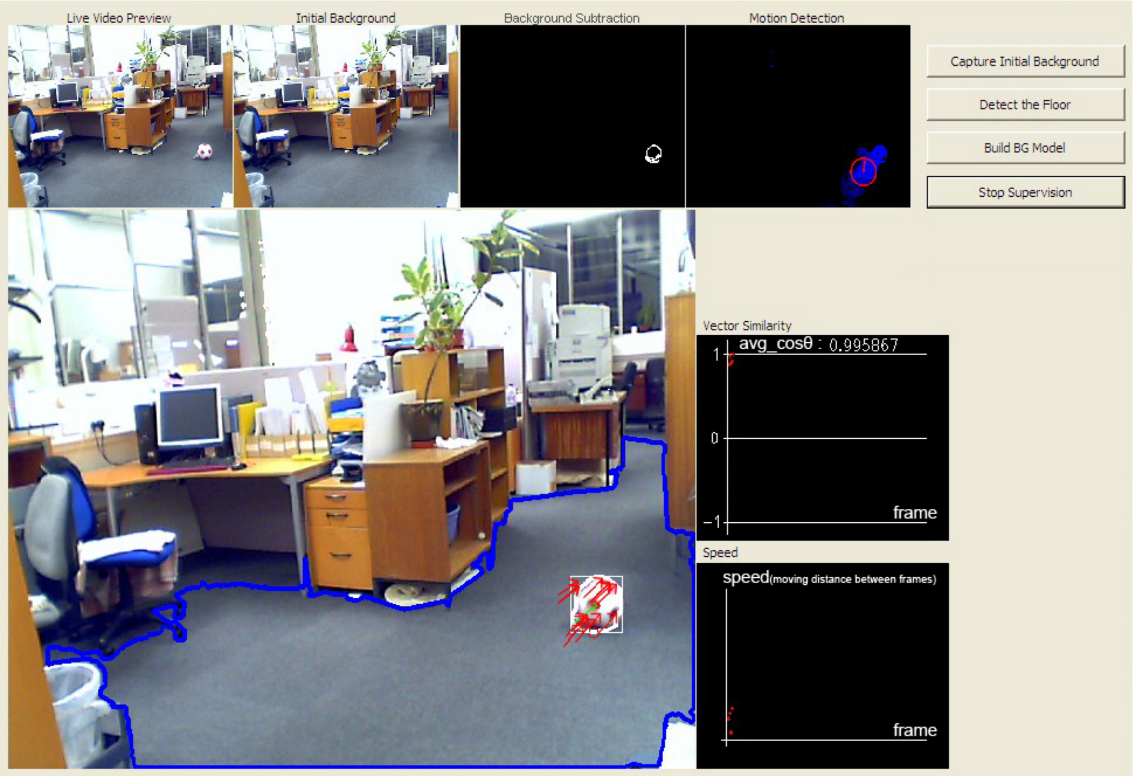
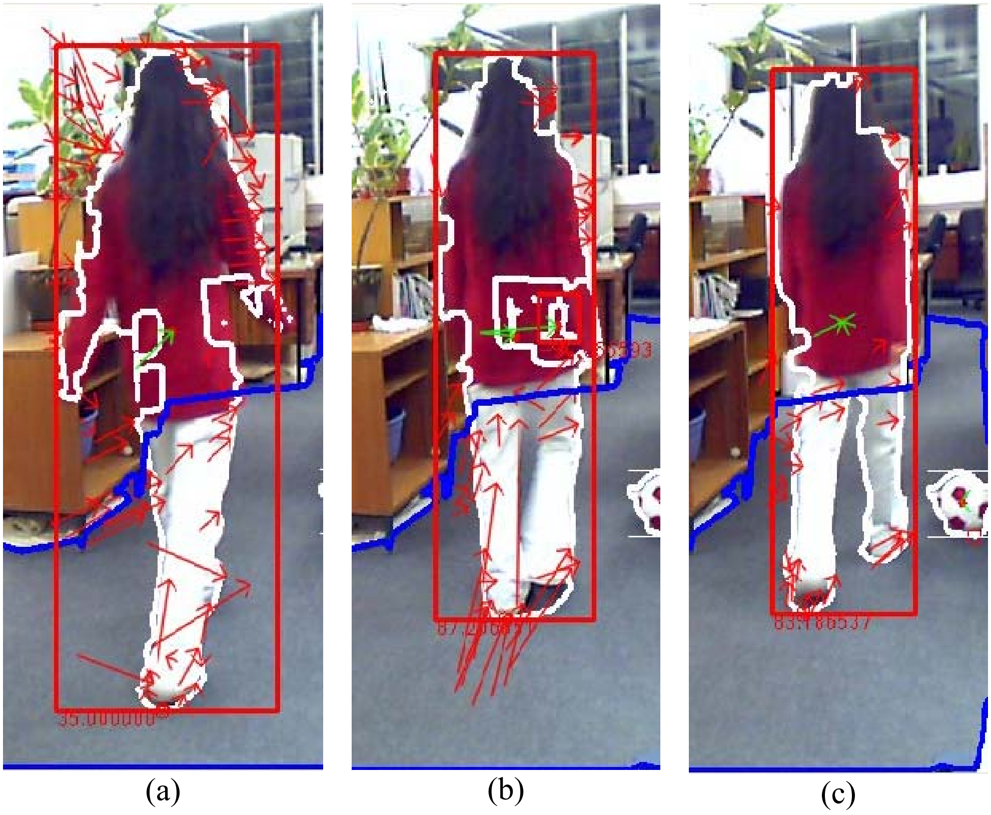


Figure 55. Dialogue-Based Window for Image Analysis

The background subtraction method works adequately with 640 x 480 images from the QuickCam Pro 5000. Logitech’s other web cameras of lower or higher performance such as the QuickCam Pro 4000 or the Ultra Vision, are more prone to noise owing to low resolution or visible compression artefacts. As the background subtraction method compares pixel intensity values, if the colour and texture of any foreground objects are very similar to those of the background, they may not get segmented completely. Therefore it is assumed that there is no foreground object with the same colour and texture as the background. Another limitation is that the background model can be automatically updated only when no foreground object is present on the scene.

The methods of human classification and handling of regional merges and splits, illustrated in Section 4.3.2 and Section 4.3.4, work adequately. Figure 56 and

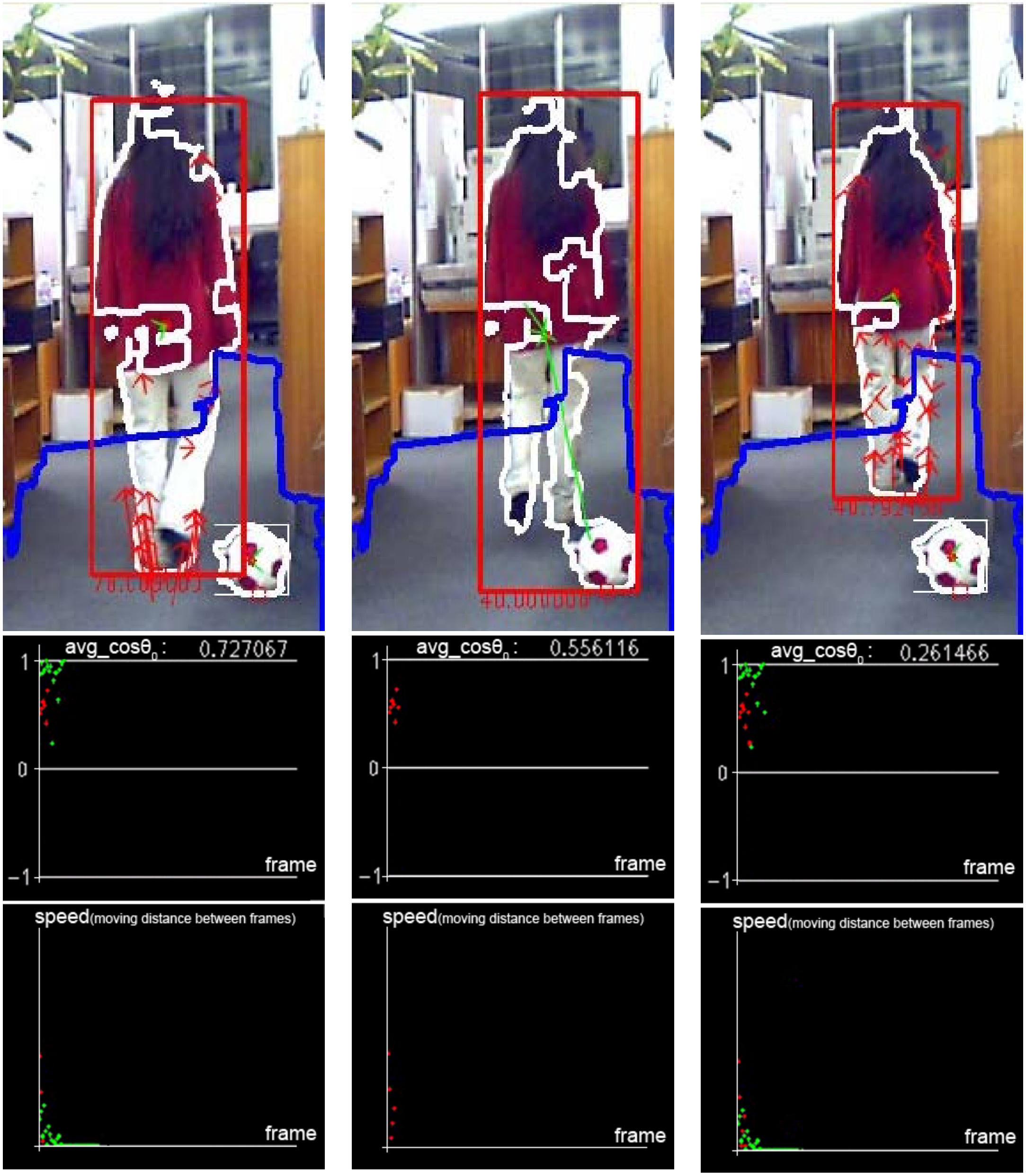
Figure 57 present successful recognition of human classification and regional merges and splits. In Figure 56, a walking person is classified as a human by the dynamic internal motion vectors and bounded by a red box while the ball is not. The person's region splits (Figure 56b) and the two split regions (one inside the other) are both bounded by a red box because the person region's data are inherited. The split regions merge immediately (Figure 56c), and the two green arrows heading the merged region's centre represent the merge.



(a) (b) (c)

Figure 56. Split and Merge of Single Object

In Figure 57, a person is passing by a ball, and their past information is separately recorded and displayed in red for the person and in green for the ball in the graphs of Figure 57a. The past information is *avg_cosθ_c* of Equation (9) and the speed, moving distances between frames. When the regions merge, only the person region’s data are kept in the merged region, as shown in Figure 57b, and when they split, the ball region’s data are returned to the ball region, as presented in the graphs of Figure 57c.


avg_cosθ0 : 0.727067
1
0
-1
frame
speed(moving distance between frames)
frame
avg_cosθ0 : 0.556116
1
0
-1
frame
speed(moving distance between frames)
frame
avg_cosθ0 : 0.261466
1
0
-1
frame
speed(moving distance between frames)
frame

(a) (b) (c)

Figure 57. Merge and Split of Multiple Objects

The system occasionally has problems, however, with differentiating merges and splits of a single object from the ones of multiple objects based on each region's size and history of merges and splits. A problematic instance is that a person's region splits while the person occludes a ball, and the size of the split region from the person is fairly similar to the ball region size. This split region would be regarded to correspond to the ball owing to the person's merge history and the similar region size, and the past information of the ball will be transferred to the split region.

4.4.1 Evaluation by Case Studies

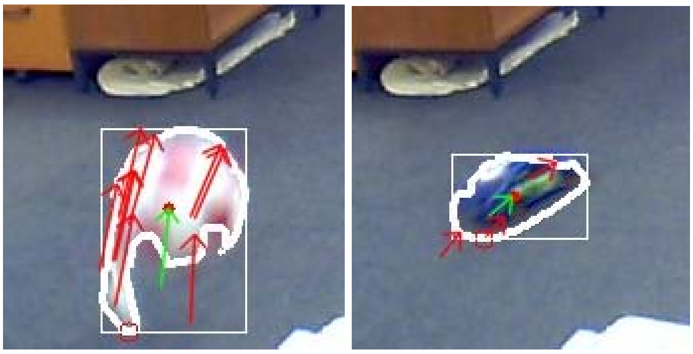
In order to objectively evaluate the performance of the novel methods proposed in this thesis for detection of the fall risk factors, a framework has been set up based on the following principles:

- The movements of the subjects in this research are recorded on 640 x 480 image resolution at 30 frames per second, and several sets of sequences, whose foreground objects are adequately segmented by the simple background subtraction method used here, are manually selected;
- A few kinds of toys moving differently are selected and are filmed solely or together when they move and then stop for assessing the methods of detecting a tripping hazard;
- A toddler is filmed with his or her toy appearing on the scene for evaluation of the human detection method;
- Each foreground region is labelled at every frame with its contour numbers in the previous and current frames for evaluation of the tracking method;
- Connections of foreground regions between every two frames for tracking are labelled with no merge/split, split of a single object, separation of multiple objects, reunion of split regions of a single object, or occlusion of multiple objects, in order to assess the method for handling regional

merges and splits, and the labels are recorded in the group of ‘merge/split’ at every frame;

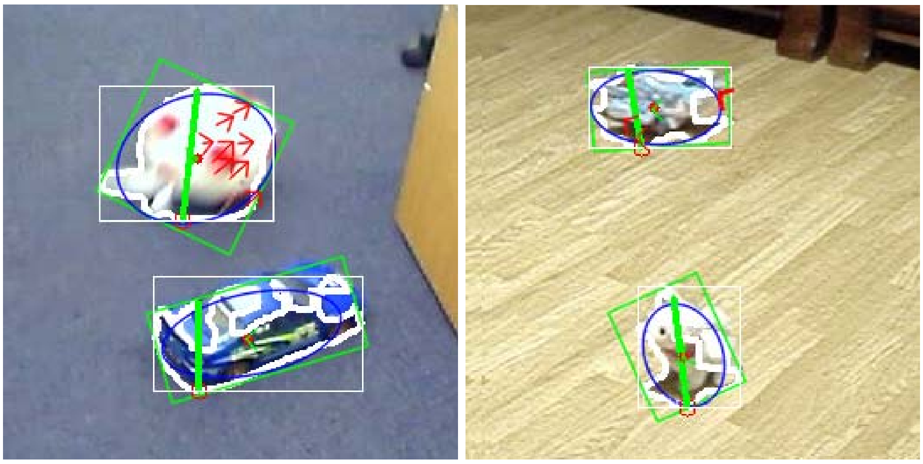
- Each foreground object is labelled with its classification, *a toddler*, *moving clutter*, *stationary clutter*, or *noise*, and the labels are recorded in the group of ‘classification’ at every frame;
- Every foreground object classified into a toddler is labelled with *safe on the floor*, *approaching furniture*, or *off the floor*, depending on its position against the floor area, and the labels are recorded in the group of ‘toddler status’ at every frame;
- All the recorded labels are checked with the sequences to judge whether they are correct or not within their groups ‘merge/split’, ‘classification’, and ‘toddler status*’*;
- The number of correct labels is divided by the total number of frames for the groups of ‘merge/split’ and ‘classification’, and by the number of frames, whose foreground region is classified into a toddler, for the group of ‘toddler status’.

First of all, the sequences to be used for the evaluation were filmed in advance and manually filtered for two reasons. One of the reasons is that the background subtraction method employed in this research is very simple and its errors directly influence the subsequent processes. The other reason is that a toddler feels comfortable and safe in his or her own home environment and only a digital camera was brought when the environment was visited to lessen the feeling of intrusion and reduce time for installation. This was also for the convenience of the toddler’s carer, who has to be present at all times during filming, according to the ethical approval given to this evaluation (Appendix C). As each frame of the input sequences was processed at the same speed of the real-time image capture (30 frames per second), it can still be said to be a real-time image analysis.



(a) (b)

Figure 58. Evaluation with Individual (a) Ball and (b) Car



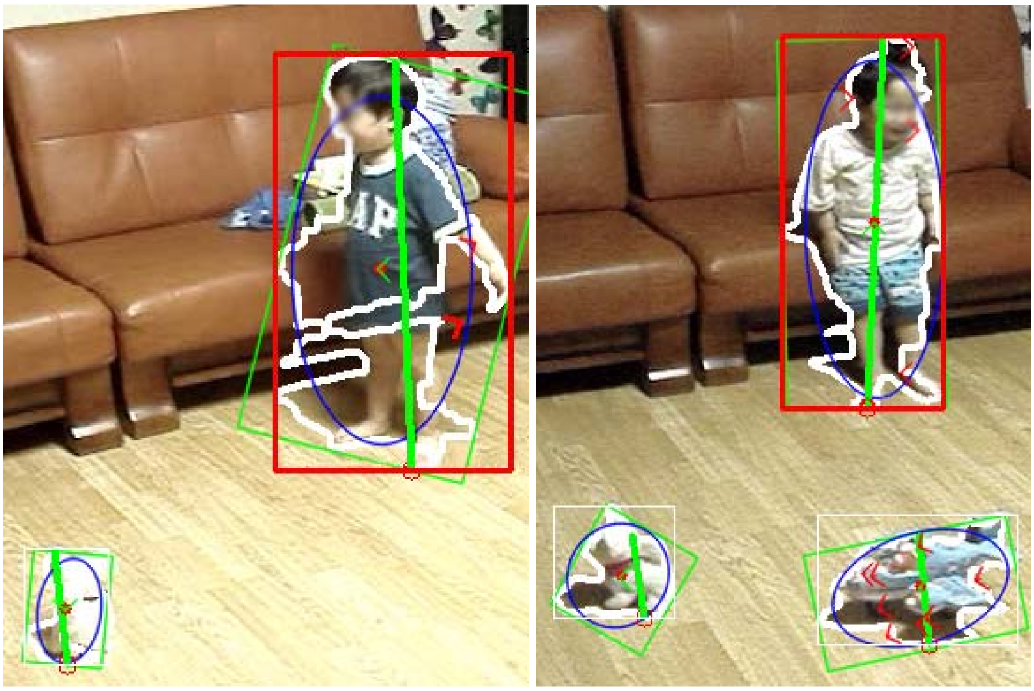
(a) (b)

Figure 59. Evaluation with (a) Ball, Car, (b) Plane, and Dog Toys

In order to evaluate the clutter detection method, not only rolling or straight-moving toys but also a simple walking toy dog, safe with a toddler, was employed. The toys were filmed solely, together, or with a toddler to assess the methods of classifying a human and clutter and handling regional merges and splits while tracking. The toys' footage was recorded when they moved and stopped in order to assess the method of differentiating stationary clutter as a tripping or slipping hazard from moving clutter. The details of each case study are as follows:

- Case 1: A ball, bouncing, rolling, and stopping (Figure 58a);

- Case 2: A radio-controlled model car, moving straight back and forth and stopping (Figure 58b);
- Case 3: A ball, rolling and stopping and a radio-controlled model car, moving straight and stopping (Figure 59a);
- Case 4: A toy plane, moving forward and a toy dog, walking back and forth (Figure 59b);
- Case 5: A 2-year-old toddler, walking and stopping and a toy dog, walking back and forth (Figure 60a);
- Case 6: A 2-year-old dancing toddler, a straight-moving toy plane, and a walking toy dog (Figure 60b)



(a) (b)

Figure 60. Evaluation with Toddler and Toys

After a set of sequences was collected and filtered to obtain foreground objects adequately segmented according to the above principles, the methods proposed in this thesis were applied to process the sequences, and each foreground object in the images was labelled at every frame with what the methods had found. One foreground object at one frame obtained labels in five groups: ‘previous contour number’, ‘current contour number’, ‘merge/split’, ‘classification’, and ‘toddler status’, as presented in Table 10.

Table 10. Labels on each Foreground Object at every Frame for Evaluation

Frame Number	Previous Contour Number	Current Contour Number	Merge/Split		Classification		Toddler Status	
			- No Merge/Split - Split of a Single Object - Separation of Multiple Objects - Reunion of Split Regions of a Single Object - Occlusion of Multiple Objects	Correct? (Yes/No)	- Toddler - Moving Clutter - Stationary Clutter - Noise	Correct? (Yes/No)	- Safe on the Floor - Approaching Furniture - Off the Floor	Correct? (Yes/No)

As described in Section 4.3.4A and illustrated in Figure 52, every foreground region is ordered by the smallest x-coordinate of its contour and numbered from zero. The number is named ‘contour number’ here, and the contour numbers of each foreground object at the previous and current frames were recorded at every frame to check if identical objects were matched over the frames. As the tracking method also deals with regional merges and splits, the state of each foreground region recognised with respect to merges and splits was recorded. The resulting classification of each foreground region was also recorded, and only for the regions classified as a toddler, ‘toddler status’ is detected and added to the labels.

After processing all the sequences, the labels recorded for each foreground region at every frame were manually checked to see whether they were correctly recognised or not within the label groups. As groups ‘previous contour number’, ‘current contour number’, and ‘merge/split’ are all related to the tracking method, the checking was executed at once together for the labels in those groups, as shown in Table 10. For instance, the evaluation results with the sequence where a ball rolled and then stopped are presented in Table 11.

Table 11. Labels on Ball for Method Evaluation

Frame No.	Previous Contour No.	Current Contour No.	Merge /Split	Correct?	Classification	Correct?	Toddler Status	Correct?
1~ 49	0	0	“No Merge/Split”	Y	"Moving Clutter"	Y		
50	0	0	“No Merge/Split”	Y	“Stationary Clutter”	N		
51~ 52	0	0	“No Merge/Split”	Y	"Moving Clutter"	Y		
53	0	0	“No Merge/Split”	Y	“Stationary Clutter”	Y		
54~ 103	0	0	“No Merge/Split”	Y	“Stationary Clutter”	Y		

The number of the labels judged to be correct was divided by the number of the total frames where the foreground object was present, for the labels in groups ‘merge/split’ and ‘classification’. For the labels in ‘toddler status’, the number of correct labels was divided by the number of the frames which included a foreground object classified as a toddler. The results are displayed in Table 12 as the success rates of the methods proposed in this research.

The method of handling regional merges and splits performed fairly well in both sets of the sequences of *clutter* and *a toddler with clutter*. This is because the better the foreground objects are segmented, the fewer problems occur in regional merges and splits, and the sequences for this evaluation have been filtered so that they have moderately segmented foreground regions. Clutter was classified correctly at high rates into *moving* and *stationary clutter*, and a toddler appearing with clutter was also favourably detected. Since the ’toddler status’ was simply based on the distance from the floor boundary, it performed well as long as toddlers were correctly detected.

Table 12. Results of Method Evaluation

<table>
<tr><th colspan="2" rowspan="2">Foreground Object</th><th colspan="2" rowspan="2">Total Frame No.</th><th colspan="7">Number (Percentage) of Frames Correctly Recognised</th></tr>
<tr><th colspan="2">Merges & Splits</th><th colspan="3">Classification</th><th colspan="2">Toddler Status</th></tr>
<tr><th>Case Group</th><th>Each Case</th><th>Each</th><th>Group</th><th>Each</th><th>Group</th><th colspan="2">Each</th><th>Group</th><th>Each</th><th>Group</th></tr>
<tr><td rowspan="4">Moving/ Static Clutter</td><td>Ball</td><td>103</td><td rowspan="4">687</td><td>103
(100%)</td><td rowspan="4">99.42%</td><td colspan="2">102
(99.02%)</td><td rowspan="4">98.40%</td><td rowspan="4"></td><td rowspan="4"></td></tr>
<tr><td>Toy Car</td><td>51</td><td>51
(100%)</td><td colspan="2">50
(98.04%)</td></tr>
<tr><td>Ball +
Toy Car</td><td>135</td><td>133
(98.52%)</td><td colspan="2">132
(97.78%)</td></tr>
<tr><td>Toy Dog +
Plane</td><td>399</td><td>397
(99.50%)</td><td colspan="2">392
(98.25%)</td></tr>
<tr><td rowspan="4">Toddler +
Moving/ Static
Clutter</td><td rowspan="2">Toddler +
Toy Dog</td><td rowspan="2">226</td><td rowspan="4">556</td><td rowspan="2">226
(100%)</td><td rowspan="4">100%</td><td>Clutter</td><td>Toddler</td><td>Clutter</td><td rowspan="2">145
(97.97%)</td><td rowspan="4">98.57%</td></tr>
<tr><td>226
(100%)</td><td>148
(65.49%)</td><td>99.28%</td></tr>
<tr><td rowspan="2">Toddler +
Toy Dog +
Plane</td><td rowspan="2">330</td><td rowspan="2">330
(100%)</td><td>Clutter</td><td>Toddler</td><td>Toddler</td><td rowspan="2">270
(98.9%)</td></tr>
<tr><td>326
(98.79%)</td><td>273
(82.73%)</td><td>75.72%</td></tr>
</table>

As there is no existing study on classification of a human and indoor clutter, it is difficult to compare the performance of the methods proposed in this thesis with the one of different methods.

4.5 Chapter Summary

This chapter delineated the details of the image analysis methods for detection of clutter on the floor as a tripping or slipping hazard, and a toddler moving around or climbing furniture or stairs. The major tasks of this image analysis are background subtraction, foreground tracking, and foreground classification into a human and clutter. A novel method using dynamic motion cues has been developed and tested for the classification of a human and clutter.

Real-time images are captured by a fixed webcam, and foreground objects are segmented by comparing the images with a simple background model. The differences are thresholded to get rid of noise and neglect slight changes of the sunlight. Each of the foreground objects is tracked over time by connecting it with the one at the nearest distance in the next frame. For this purpose, the centre of mass of each region is detected at every frame, and the distances between each centre in one frame and all the centres detected in the next frame are calculated. The motion information (speed and direction) of each object can be obtained from the relation of the connected centres.

Meanwhile, multiple objects can appear to be a connected component, and a single object can split into multiple regions owing to occlusions or errors in background subtraction that will confuse the tracking of each foreground object. In order to connect identical objects between consecutive frames, despite the regional merges and splits, the closest region centres between two successive frames are detected twice from the previous to the next frames and vice versa. In addition, each region's size and its history of merges and splits are used to distinguish between multiple objects and a single object in merges and splits.

On the other hand, all the foreground objects are classified into a human and clutter since the necessary information from a toddler is different from that from clutter in detecting the fall risk factors. The classification is based on dynamic human motion owing to the different body part motions when the whole body is mobile. Conversely, typical clutter in an indoor home environment may have relatively constant motion vectors. Therefore, the internal motion vectors are captured in each foreground region, and their similarity is calculated.

The fall risk factors related to a toddler's behaviour are activated if he or she is moving close to or climbing furniture or room structures. Hence, the floor region is manually selected once when the camera is installed and used to determine if any toddler region approaches the boundary of the floor area or off the area, with the assumption that the non-floor area is filled with furniture and room structures.

All the methods introduced in this chapter were objectively evaluated by case studies of toys solely, together, or with a toddler. The evaluation shows that stationary clutter being a tripping or slipping hazard was detected at high rates and a toddler was also moderately classified. The recognition of toddler status regarding the fall risk factors also performed well as long as toddlers are correctly detected. The use of dynamic internal motions, however, would not good enough for human classification when a pet appears with a toddler on the scene. This is because a pet moves its body parts diversely like a human. Therefore, the next chapter introduces additional motion cues considering pets so as to strengthen the human classification.

Chapter 5
Classification of Human and Pet

5.1 Introduction

Many families with young children may also have a dog or a cat. In case both a toddler and a pet are seen in images, the human classification using the different internal motions described in Section 4.3.2 would not work well since a pet also moves its body parts differently. Thus, other dynamic cues regarding pets have been added to reinforce the human classification.

The existing studies on animal detection reviewed in Section 2.3.3 use articulated models or motion templates focusing on leg movement to discriminate different animals, distinguish between animals and humans, or further recognise specific activities of animals. The work in this thesis does not need to identify different animals or recognise their activities but needs to differentiate a human from pets to detect a toddler’s fall risk factors. Thus, it would be excessive to build and apply motion templates for different animals as the existing studies do. The use of cues related to gaits may also cause problems when the legs are occluded or out of the camera view. Therefore, the cues used in this research have been discovered from observation of pets’ movement and are deemed to be adequate for the human classification in indoor home environments.

The first cue is dynamic posture changes of pets, especially within a limited space like an indoor home environment, compared with a toddler generally being upright while walking. This is detailed in Section 5.3.1, following the method overview in Section 5.2. The other cue is the actual height information illustrated in Section 5.3.2 since the size of a toddler is usually different from that of a pet. These two cues are applied together to the human classification in Section 5.3.3.

Lastly, Section 5.4 presents the implementation and the evaluation by case studies of the methods.

5.2 Method Overview

One of the cues for classification between a toddler and pets is the diversity in the posture changes of a dog or a cat within a limited space such as an indoor home environment, and the other is actual height information in 3D world coordinates. The method flowchart for collecting those two pieces of information and using them to classify a toddler and a pet is shown in Figure 61. The grey-labelled elements of the flowchart were already introduced in Section 4.2, which gave an overview of the methods of classifying a toddler and clutter. As many of the elements are also involved in the method of classifying a toddler and a pet, they are presented in the flowchart here to show the relations but are not explained in this section.

In order to obtain the actual height information of each foreground object, the camera should be calibrated first to learn the relations between the projected image plane and the real-world space. Individual reference datasets for a toddler and a pet are then built in terms of their actual heights and dynamic posture changes by analysis of separate sample videos of their movements. These datasets are referred to later for estimating the possibilities of each foreground object to be a toddler and a pet respectively. The camera calibration and the dataset construction need to be done before analysis of real-time images.

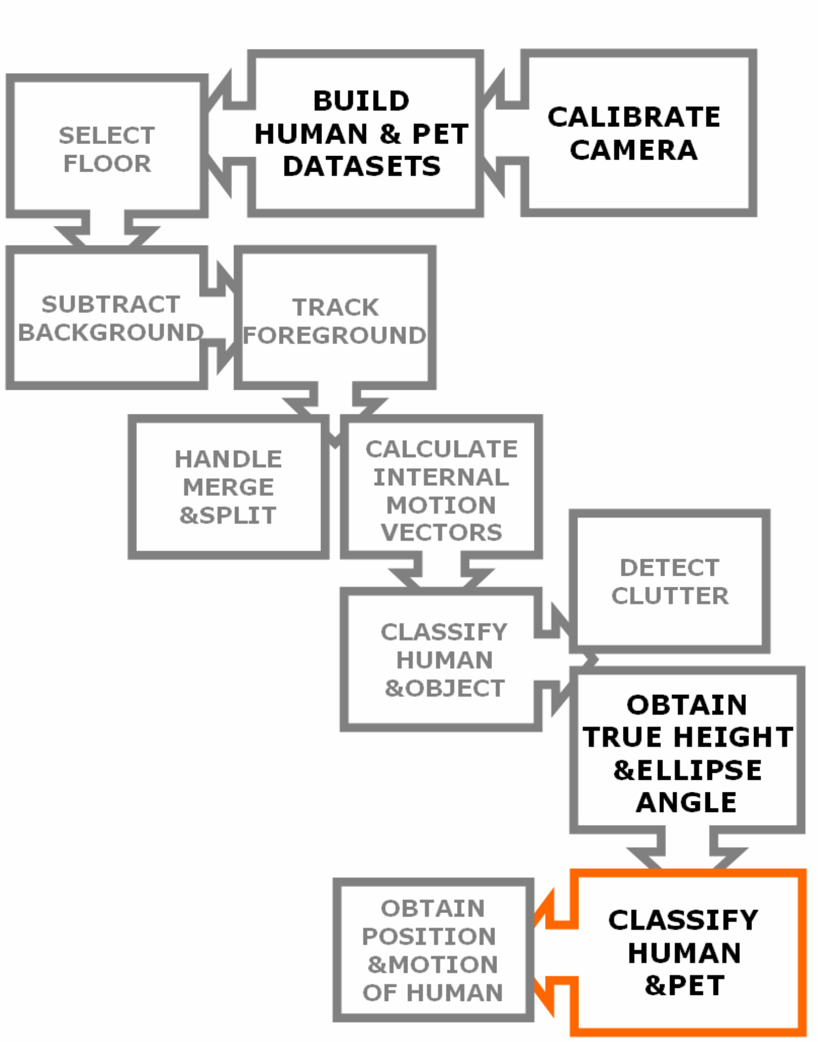


Figure 61. Method Flowchart for Classification of Human and Pet

During the analysis of real-time images, the foreground objects with dynamic internal motion vectors are focused on for the classification of a toddler and a pet. Their actual heights and fitting ellipse angles are calculated and saved in each frame, and the data from all the past frames are used to calculate the certainty of each foreground object to be a toddler or a pet. Then the foreground objects are classified in terms of the certainty.

5.3 Methodology

This section only illustrates the additional methods of human classification considering pets. The use of two dynamic motion cues regarding pets is individually delineated, and the details of their combination follow.

5.3.1 Angle Changes of Fitted Ellipse

It was observed that a dog or a cat changes its posture very dynamically when it moves around because they tend to turn around frequently in a limited space like the indoor home environment. A cat especially stretches the body or the tail widely when moving. On the other hand, the pose of toddlers does not change as much since they generally move the limbs and keep the trunk almost perpendicular to the ground when they toddle around. Therefore, it has been decided to use the diversity of a pet's posture changes to classify a toddler and a pet. As the difference we need to capture between a toddler and a pet in terms of posture changes is dependent on keeping upright or not, it was decided to analyse the angle changes of the ellipse fitted to a foreground region rather than its contour changes. This is also because the contour analysis may cause another problem because of uncertain segmentation from some errors in background subtraction.

In order to obtain an ellipse fitted to the contour of a foreground object, the method of Fitzgibbon and colleagues [166], which is robust to noise and an excellent trade-off between speed and accuracy for ellipse fitting, was applied. A general conic is presented by the implicit second order polynomial, as follows:

$$F(x,y) = ax^2 + bxy + cy^2 + dx + ey + f = 0 . \qquad (13)$$

$F(u,v)$ is called the algebraic distance of a point (u,v) to the conic $F(x,y) = 0$. The fitting of a general conic is approached by minimising the sum of squared algebraic distances of the curve to the N data points (x_i, y_i) as given with Equation (14).

$$D_A = \sum_{i=1}^{N} F(x_{\mathrm{i}}, y_i)^2 = \sum_{i=1}^{N} (ax_i^2 + bx_i y_i + cy_i^2 + dx_i + ey_i + f)^2 \ . \tag{14}$$

$$4ac - b^2 = 1. \tag{15}$$

$$ang = \frac{1}{2} arctg\left(\frac{b}{a-c}\right). \tag{16}$$

Furthermore, Fitzgibbon and colleagues constrain the parameters of Equation (13) to satisfy Equation (15) in order to achieve ellipse-specific fitting. The rotation angle (*ang*) of the resulting ellipse can be derived with Equation (16) [167].

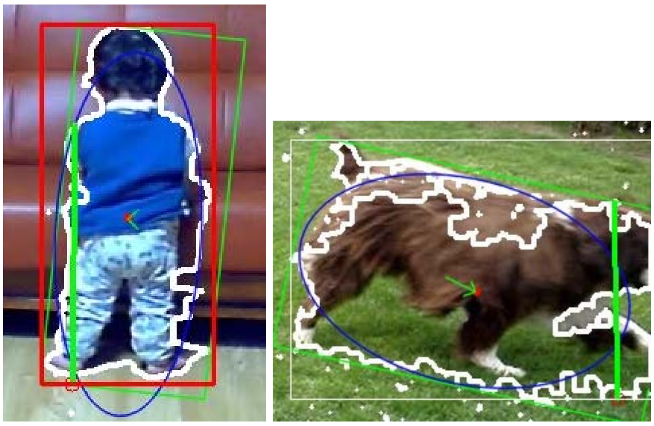


Figure 62. Ellipses Fitted to Toddler and Dog

The results of the ellipse fitting method are the blue ellipses in Figure 62. The angle of the resulting ellipse is returned to be the one between horizontal axis of the image and the ellipse axis with the shorter length as illustrated in Figure 63. Owing to the symmetry of ellipses, the range of ellipse angles is from 0 to 180 degree, and the angle of almost upright ellipses becomes 0, slightly less than 180, or slightly more than 0 degree. Hence, the ellipse angles are subtracted 90 and turned to be positive as given with Equation (17) so as to make the angles of almost upright ellipses have similar degrees and derive relative consistency of toddlers' upright postures.

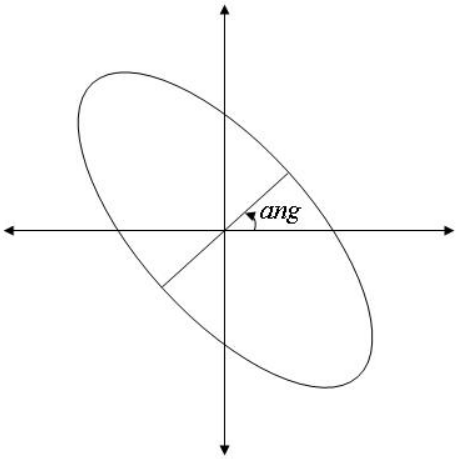


Figure 63. Ellipse Angle

$$ang = abs(ang - 90)\,. \tag{17}$$

In order to capture the diversity, all the angle changes of the ellipse fitted to each contour are recorded over time only when the whole body is mobile. The mean average ($angMean_c^t$) and the standard deviation ($angSD_c^t$) of the angle history (ang_c^i) for a foreground object (c^{th} ROI) at frame t, are calculated with Equation (18).

$$\begin{aligned} angMean_c^t &= \frac{1}{N}\sum_{i=1}^{t} ang_c^i, \\ angSD_c^t &= \sqrt{\frac{1}{t}\sum_{i=1}^{t}(ang_c^i - angMean_c^t)^2}\,. \end{aligned} \tag{18}$$

5.3.2 Actual Height

The body size of a toddler or a pet is meant to be various depending on their growth level or race/species and would be an obvious feature for differentiating between them. Hence, a body size factor, height, has been chosen as a cue for differentiating a toddler, whose primary growth direction is vertical, from a pet, whose direction is horizontal. Camera calibration is required to obtain the actual height information of each foreground object, and a single initial calibration is enough as the camera is fixed. This section delineates the details of the camera calibration method applied in this research, and the calculation of each object's actual height at every frame, based on the calibration parameters.

A. Camera Calibration

According to Section 2.3.4, which analysed existing studies related to camera calibration, a pinhole camera model is commonly used to calibrate a camera, and radial distortion is the popular lens effect to be handled. Therefore, a basic pinhole camera model is applied here to calibrate the camera used in this research, and the radial distortion problem is also handled.

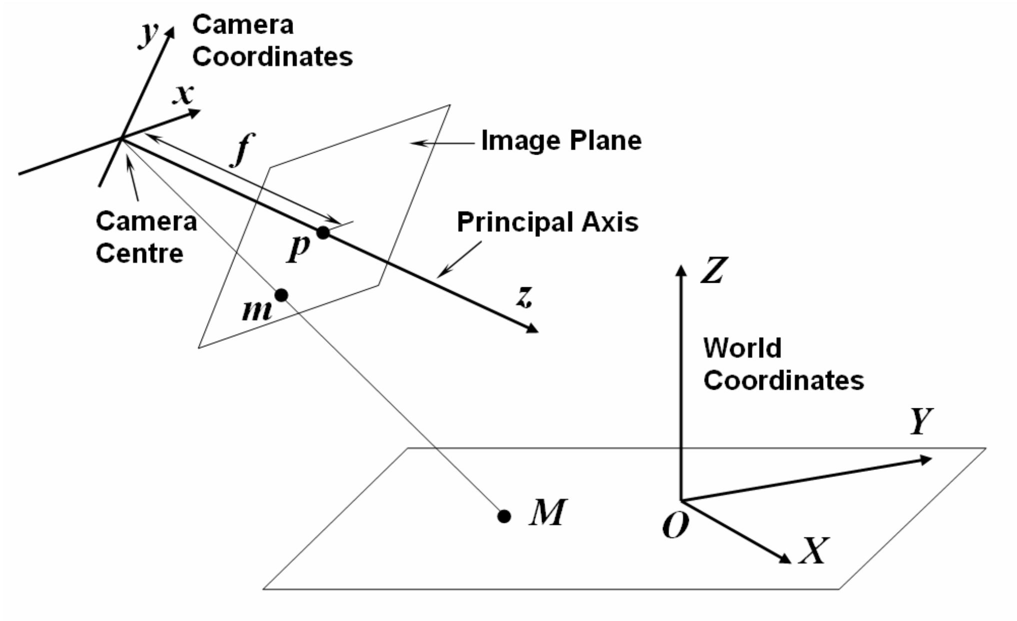


Figure 64. Relationship of Image Plane and World Space

The pinhole camera model describes the mathematical relationship between the coordinates of a 3D point and its projection onto the image plane of an ideal pinhole camera which has no lens but a pinhole to let very little light through

[168]. Figure 64 shows the relationship between a point, *M*, in the 3D world space and its projection, *m*, onto the image plane. The details of the mathematical relationship follow:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = [R|t]M, \textit{where } R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, t = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}. \tag{19}$$

$$x' = x/z, y' = y/z . \tag{20}$$

$$\begin{aligned} & x'' = x'(1+k), y'' = y'(1+k), \\ & \textit{where } k = k_1(x'^2 + y'^2) + k_2(x'^2 + y'^2). \end{aligned} \tag{21}$$

First of all, the camera coordinates of the point *M* are calculated as (*x*, *y*, *z*) in Equation (19), using the rotation matrix, *R* and the translation vector, *t*, both of which relate the world coordinate system to the camera coordinate system. The x- and y-coordinates of (*x*, *y*, *z*) are normalised as (*x'*, *y'*) in Equation (20) by the z-coordinate, which is the principal axis of the camera coordinates and then are distorted as (*x*", *y*") in Equation (21), where k_1 and k_2 are the two radial distortion coefficients, considering only the first two terms of radial distortion [115].

$$u = f_x \times x'' + c_x, v = f_y \times y'' + c_y . \tag{22}$$

Finally, the distorted coordinates (*x*", *y*") are transformed into (*u*, *v*), the image plane coordinates of the point *m*, as given in Equation (22), where f_x and f_y are focal lengths and c_x and c_y are the coordinates of a principal point, all of which are expressed in pixel-related units. Although some studies reviewed in Section 2.3.4 include a skew parameter in camera calibration, it is not considered here since the skew parameter is zero for most normal cameras [169].

Camera calibration is the process of obtaining all those parameters and coefficients which can define the relationship between a point in the world space and its projection, using some features of an image whose world coordinates are known. The chessboard in Figure 65 is used in this work since a square pattern is popular in camera calibration owing to the easy detection of the point features, as shown in Section 2.3.4E. Moreover, a planar pattern laid on the floor is good enough to discover the planar world coordinates of the floor and the real position of each foreground object on the floor. As a toddler can hardly jump and a pet usually walks around in the indoor home environment, it is assumed that they always walk when their whole body is mobile. Therefore, each object's actual height can be calculated from the position on the floor, assuming every foreground object is always perpendicular to the ground when walking around.

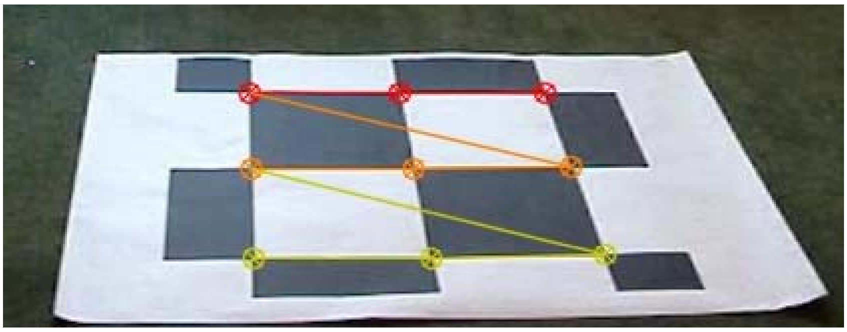


Figure 65. Part of Chessboard

The chessboard from Figure 65 contains four-by-four squares, each of which measures 20 cm x 20 cm. The chessboard is spread on the floor while capturing images for the camera calibration, and the z-coordinates of the nine points detected on the chessboard are all fixed as zero. The x- and y-coordinates can be defined by the square size and are used to calculate the parameters and the coefficients. After all the parameters and the coefficients have been obtained, the image plane coordinates of the chessboard points are calculated from their world coordinates, the parameters, and the coefficients by following Equations (19), (20), (21), and (22). This is for checking whether the camera has been correctly calibrated, and the calculated coordinates are compared with the real image plane coordinates. When the average of the differences is below one, the parameters and the coefficients are confirmed to be correct and applied to calculate the actual height of each foreground object.

B. Calculation of Actual Height

First, the lowest point of the contour of each foreground object is defined as where the object stands on the floor. The world coordinates of the point are calculated by fixing the z-coordinates to be zero and reversing Equations (19), (20), (21), and (22) with the camera parameters and the distortion coefficients obtained during the camera calibration.

Then the world z-coordinate of the highest point of the object contour is calculated to have the same x- and y-coordinates as the ones of the lowest point and saved as the object's actual height. As the calculated heights may be wrong when toddlers stay seated, the heights are calculated when the whole body moves in order to capture the heights when they walk. All the heights of each object calculated over the time are also recorded and used because toddlers can look shorter when they bend the body or dogs can appear taller owing to their posture or errors in foreground segmentation.

5.3.3 Information Fusion of Multiple Cues

The three kinds of data, the mean average and the standard deviation of the angle changes of contour-fitting ellipse and the history of actual heights, need to be simultaneously applied to classify each foreground region into a toddler or a pet. A toddler and pet show fairly different aspects from each other in terms of the three kinds of data, and each foreground object can be classified by finding which aspect the object's data are close to. Hence, a theory is necessary here to quantify how close to each aspect the data of a foreground object are, and entropy is employed. Entropy is generally known as the degree to which a given quantity of thermal energy is available for doing useful work [170], but Claude Shannon redefined it as uncertainty of a system, represented by the logarithm of possible combinations of states in that system [171]. As Shannon's entropy measures uncertainty of a system based on a set of probabilities [172], the classification could be customised to a specific toddler and pet when a short footage of the toddler and the pet is used to identify the probabilities.

There has been some work on fusion of multiple sorts of data or information based on entropy. Zhou and Leung [173] fused multi-sensor data using entropy in order to find the optimum weights that minimise the uncertainty of the fused information, based on the empirical distribution of the sensor data. Kern-Isberner and Rödder [174] combined pieces of probabilistic information stemming from different sources using the principle of maximum entropy, which processes information most faithfully. Hsu and Chang [175] applied maximum entropy to fuse diverse features from multiple levels and modalities effectively, including visual, audio, and text so as to segment and classify news video stories.

$$H = -\sum_{i=1}^{n} p_i \log p_i . \tag{23}$$

$$H_\alpha = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} p_i^\alpha \right). \tag{24}$$

$$H_2 = -\log \sum_{i=1}^{n} p_i^2 . \tag{25}$$

Shannon's entropy (*H*) is estimated with Equation (23), when $p_1, p_2, ..., p_n$ are given as a set of probabilities. It cannot, however, be calculated when any of the probabilities is zero because log0 does not exist. Alfred Rényi [176] generalised Shannon's entropy as presented in Equation (24), where order *α* is no less than zero and p_i are a set of input probabilities. In the situation where α approaches one, H_α converges to Equation (23), Shannon's entropy. When *α* = 2, Rényi's entropy becomes Equation (25), which can also be calculated when any of the probabilities is zero and thus is adopted for the data fusion in this research.

$$C^h = \sum_{i=1}^{n} (p_i^h)^2, C^p = \sum_{i=1}^{n} (p_i^p)^2 \,. \tag{26}$$

Equation (26), where –log is omitted from Equation (25) and p_i^h and p_i^p are probabilities of a toddler and a pet respectively, is used to classify each foreground object into a toddler and a pet in this research. Basically log *x* is proportional to *x*, and in this work the entropies will be measured for a toddler and a pet separately and compared with each other, which means this research only concerns the relative values of the entropies. Therefore, –log can be omitted from Rényi's entropy and the omission of minus(-) changes the entropy (uncertainty) to certainty. Therefore, C^h and C^p of Equation (26) become the certainties of being a toddler and a pet respectively, and the bigger value decides the classification.

In order to determine the probabilities p_i for the certainty calculation, one brief video of a subject toddler and one of a subject pet are applied to set up reference probabilities by range for the history of actual heights and the mean average and standard deviation of the angle changes of fitted ellipse. The ranges are from zero to 100 at ten intervals for the actual heights and from zero to 90 for the data related to the ellipse angle changes, as shaded with grey in Table 13 and Table 14. This is because neither a toddler nor a pet would look taller than 100 centimetres and ellipse angles are always between zero and 90 degrees, as explained in Section 5.3.1.

The videos used to make the reference probabilities are subject to containing the whole body movement of a toddler and a pet for more than ten seconds at the rate of 30 frames per second. In other words, the three kinds of data are acquired from at least 300 frames and arranged by the range, and the number of the frames belonging to each range is counted and divided by the total frame number. Then the results are multiplied by 100 to define in percentage terms the possibilities of having the toddler and the probabilities of having the pet on the image, as presented in Table 13 and Table 14.

Table 13. Sample Probabilities of Toddler

Range of Height(cm)/ Angle(degree)	Probabilities		
	History of Actual Height	Mean of Ellipse Angle Changes	Standard Deviation of Ellipse Angle Changes
0 ~ 10	0	0	41.9354838709677440
10 ~ 20	0	0	58.0645161290322560
20 ~ 30	0	0	0
30 ~ 40	0	0	0
40 ~ 50	0	0	0
50 ~ 60	75	0	0
60 ~ 70	25	0	0
70 ~ 80	0	85.4838709677419360	0
80 ~ 90	0	14.5161290322580640	0
90 ~ 100	0		

Table 14. Sample Probabilities of Dog

Range of Height(cm)/ Angle(degree)	Probabilities		
	History of Actual Heights	Mean of Ellipse Angle Changes	Standard Deviation of Ellipse Angle Changes
0 ~ 10	0	25.3968253968253950	25.3968253968253950
10 ~ 20	45.4545454545454530	5.2910052910052912	1.5873015873015872
20 ~ 30	54.5454545454545470	54.4973544973545000	44.9735449735449750
30 ~ 40	0	14.8148148148148150	28.0423280423280410
40 ~ 50	0	0	0
50 ~ 60	0	0	0
60 ~ 70	0	0	0
70 ~ 80	0	0	0
80 ~ 90	0	0	0
90 ~ 100	0		

During the real-time image analysis, the actual heights and the fitted ellipse angles are collected from all the past frames as well as the current frame, each piece of data obtains its own possibility, referring to the reference probabilities. The possibilities are detected separately for a toddler and a pet and used to calculate C^h and C^p of Equation (26), and the bigger value decides the category of the foreground region.

5.4 Implementation

A single Logitech Quickcam Pro5000 was used in a fixed position to capture real-time images of 640 x 480 pixels at 30 frames per second, and the algorithms were written in C++ using OpenCV for method implementation. First, the posture changes of a toddler, a dog, and a cat were separately estimated by calculating angles of the ellipse fitted to their contour over time. Their actual heights were also obtained after camera calibration, and then the combination of those two kinds of information were implemented. Figure 66 presents the whole workflow of the human detection.

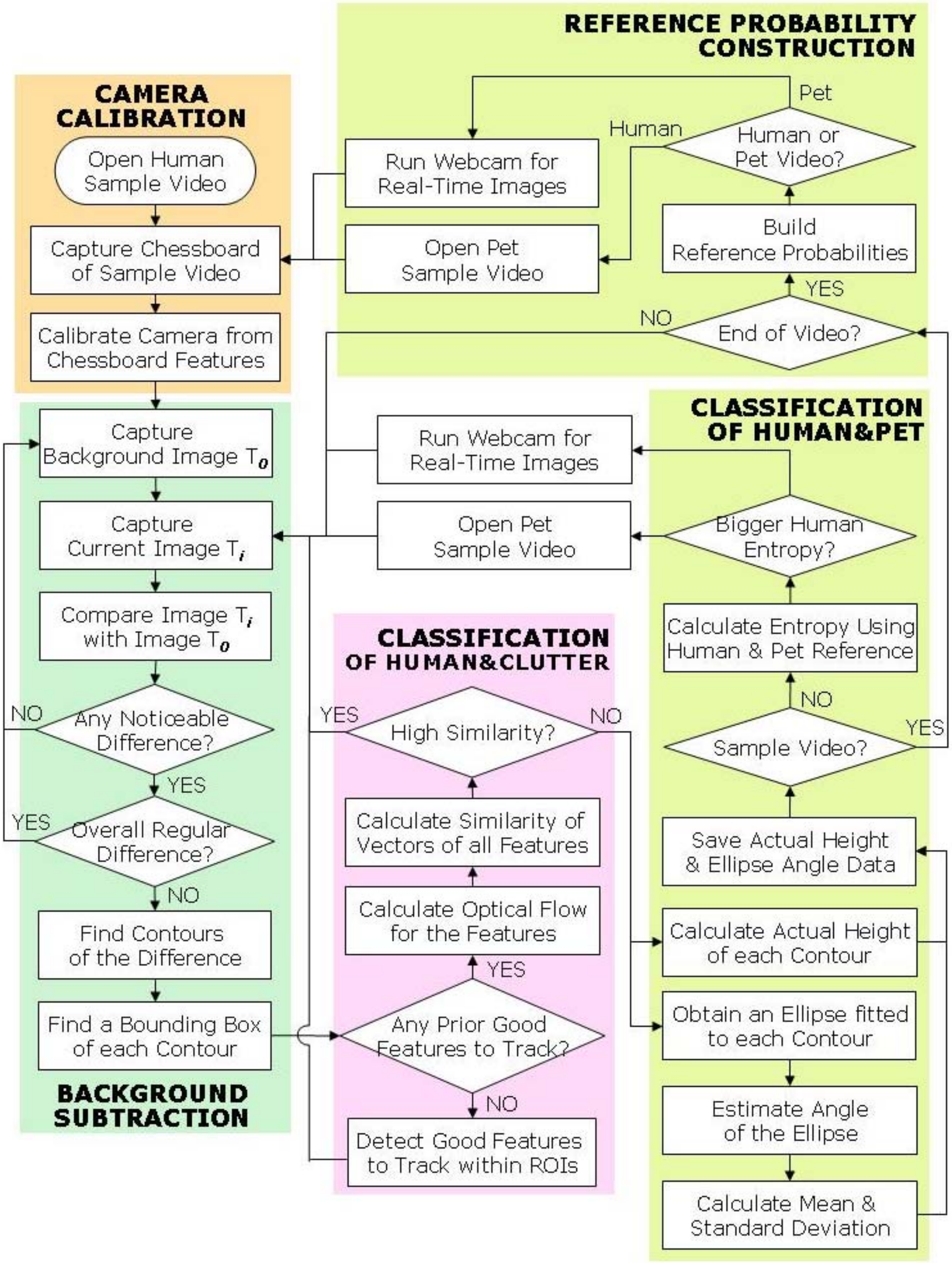
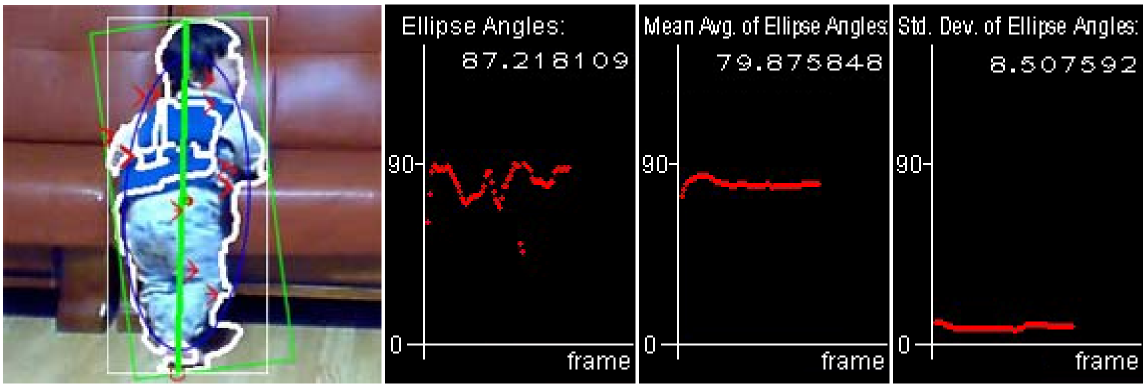

REFERENCE PROBABILITY CONSTRUCTION
Pet
Human
Human or Pet Video?
Run Webcam for Real-Time Images
Open Pet Sample Video
Build Reference Probabilities
YES
NO
End of Video?
CAMERA CALIBRATION
Open Human Sample Video
Capture Chessboard of Sample Video
Calibrate Camera from Chessboard Features
Capture Background Image T_0
Capture Current Image T_i
Compare Image T_i with Image T_0
NO
Any Noticeable Difference?
YES
YES
Overall Regular Difference?
NO
Find Contours of the Difference
Find a Bounding Box of each Contour
BACKGROUND SUBTRACTION
CLASSIFICATION OF HUMAN&CLUTTER
YES
High Similarity?
NO
Calculate Similarity of Vectors of all Features
Calculate Optical Flow for the Features
YES
Any Prior Good Features to Track?
NO
Detect Good Features to Track within ROIs
CLASSIFICATION OF HUMAN&PET
Run Webcam for Real-Time Images
Open Pet Sample Video
Bigger Human Entropy?
Calculate Entropy Using Human & Pet Reference
NO
Sample Video?
YES
Save Actual Height & Ellipse Angle Data
Calculate Actual Height of each Contour
Obtain an Ellipse fitted to each Contour
Estimate Angle of the Ellipse
Calculate Mean & Standard Deviation

Figure 66. System Workflow for Classification of Human and Pet

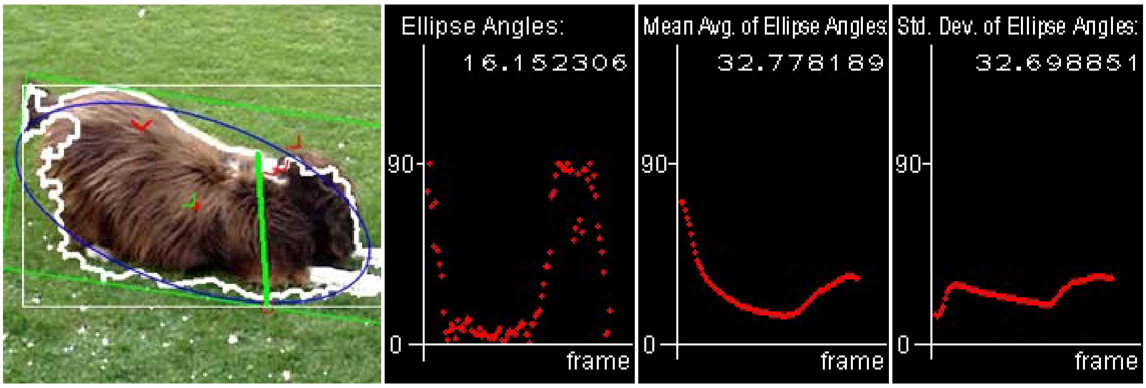
The ellipse fitted to a foreground contour appeared fairly differently over time between a toddler and pets. Figure 67a, Figure 68a, and Figure 69a show the ellipse angles over time of a toddler, a dog, and a cat respectively, and the angles are somewhat constant for the toddler and relatively dynamic for the pets.

Whereas the mean average of ellipse angles for the walking toddler in Figure 67 stays near 90 degrees (Figure 67b), the mean averages of ellipse angles for the wandering dog in Figure 68 and the moving cat in Figure 69 are relatively dynamic (Figure 68b and Figure 69b). Moreover, the standard deviations of ellipse angles for the dog and the cat (Figure 68c and Figure 69c) are noticeably larger than the one for the toddler (Figure 67c).



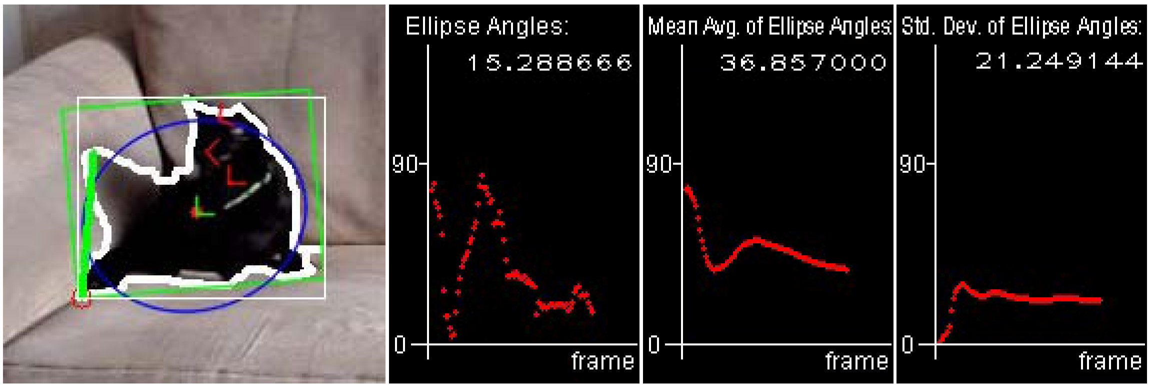
(a) (b) (c)

Figure 67. Angle Changes of Ellipse Fitted to Toddler



(a) (b) (c)

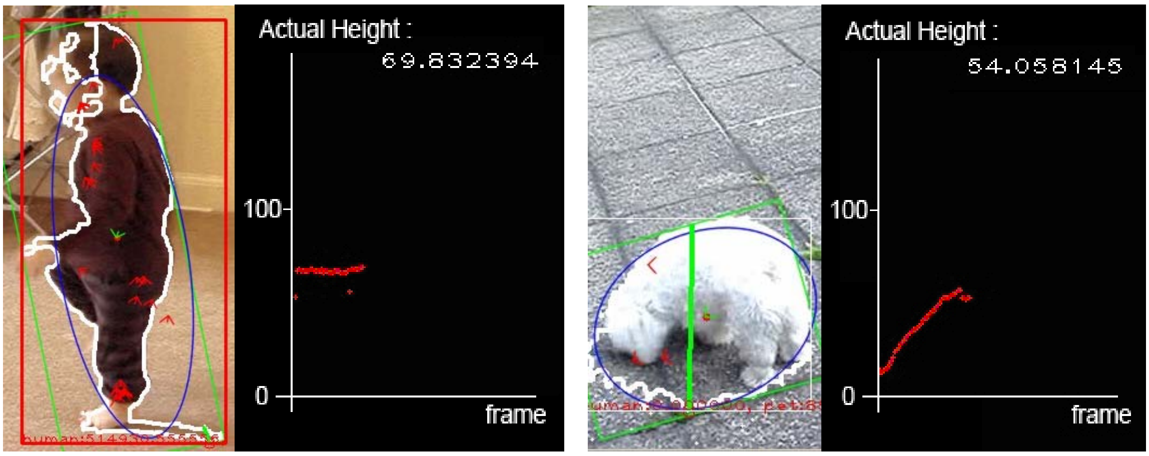
Figure 68. Angle Changes of Ellipse Fitted to Dog



(a) (b) (c)

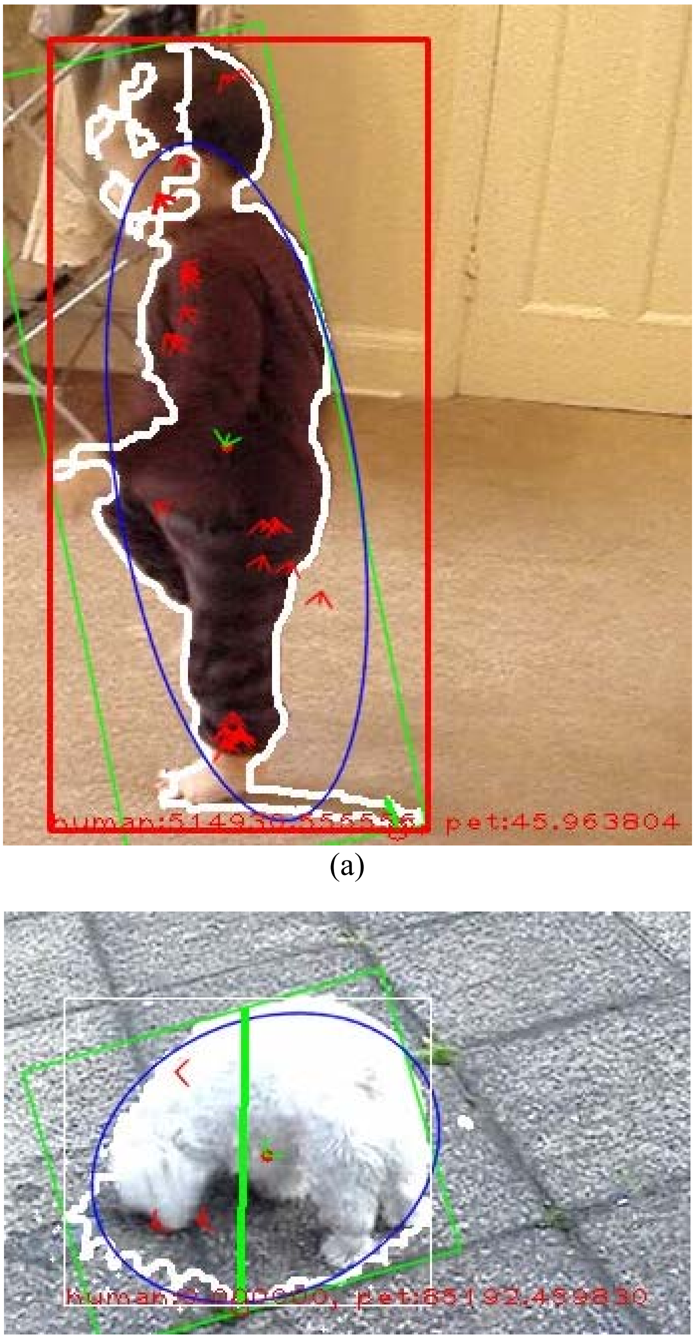
Figure 69. Angle Changes of Ellipse Fitted to Cat

The actual heights were estimated over time for a walking toddler and a wandering dog, and Figure 70 presents the graphs of heights. While the toddler's heights constantly appear to be around 70 centimetres, the dog's heights vary over time. This is because in images a pet frequently look taller or shorter than its real height owing to the dynamic postures, and thus the history of actual heights is also useful information for classification between a human and a pet.



(a) (b)

Figure 70. Height Records of (a) Walking Toddler and (b) Dog


human:514930.356536 pet:45.963804
human:0.000000, pet:85192.489830

(a)

(b)

Figure 71. Classification of (a) Toddler and (b) Dog

The fitted ellipse angle changes and the actual height history were fused for human classification, and some results can be seen in Figure 71. The numbers on the bottom of the bounding boxes are the calculated C^h and C^p of Equation (26), which are 514930.555556 and 45.963804 for the toddler in Figure 71a and 0.000000 and 85192.459830 for the dog in Figure 71b. C^h is bigger than C^p for the toddler, and it is the other way around for the dog. Therefore, they were correctly classified by the fusing of their ellipse angle changes and actual heights after clutter was discarded according to the dynamic internal motion vectors.

5.4.1 Evaluation by Case Studies

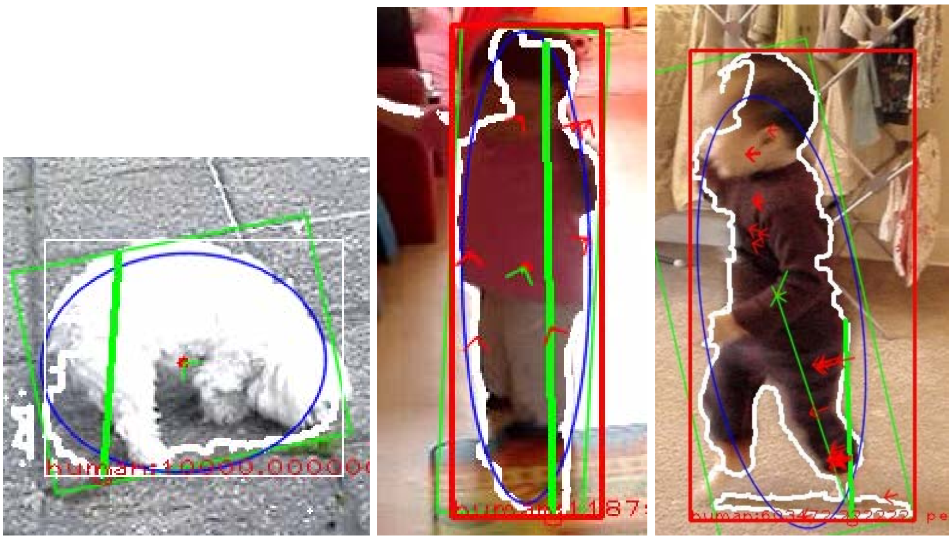
The principles for evaluation of the methods delineated in this chapter are almost the same as the principles set up in Section 4.4.1 and are as follows:

- The movements of the subjects in this research are recorded on 640 x 480 image resolution at 30 frames per second, and several sets of sequences, whose foreground objects are adequately segmented by the simple background subtraction method used here, are manually selected;
- Whenever the camera is installed, a chessboard pattern is laid on the floor appearing on the camera scene and filmed for five seconds. Then the pattern is removed, and a toddler and a pet are filmed separately or together;
- Each foreground region is labelled at every frame with its contour numbers of the previous and current frames for evaluation of the tracking method;
- Connections of foreground regions between every two frames for tracking are labelled with no merge/split, split of a single object, separation of multiple objects, reunion of split regions of a single object, or occlusion of multiple objects, in order to assess the method for handling regional merges and splits, and the labels are recorded in the group of ‘merge/split’ at every frame;

- Each foreground object is labelled with its classification, *a toddler*, *a pet*, *moving clutter*, *stationary clutter*, or *noise*, and the labels are recorded in the group of 'classification' at every frame;
- Every foreground object classified as a toddler is labelled with *safe on the floor*, *approaching furniture*, or *off the floor*, depending on its position against the floor area, and the labels are recorded in the group 'toddler status' at every frame;
- All the recorded labels are checked with the sequences to judge whether they are correct or not within their groups 'merge/split', 'classification', and 'toddler status';
- The number of correct labels is divided by the total number of frames for the groups of 'merge/split' and 'classification', and by the number of frames whose foreground region is classified as a toddler for the group 'toddler status'.

As most of the above principles are the same as the principles in Section 4.4.1, only the differences are described here. The first, small, part of the sequences for this evaluation included the chessboard pattern in Figure 65 laid on the floor for camera calibration. Then footages of a toddler and a pet were recorded individually or together, and used to assess the methods related to classification of a toddler and a pet. The individual footages were also used to set up the reference probabilities for the entropy-based calculations described in Section 5.3.3. The details of each case study are as follows:

- Case 1: A dog, walking or running and stopping for food (Figure 72a);
- Case 2: An 1-year-old toddler, dancing in front of a CD player (Figure 72b);
- Case 3: An 1-year-old toddler, running around his room (Figure 72c);
- Case 4: An 1-year-old toddler, walking around and a cat, sneaking and stopping (Figure 73).



(a) (b) (c)

Figure 72. Evaluation with (a) Dog and (b)(c)Toddlers

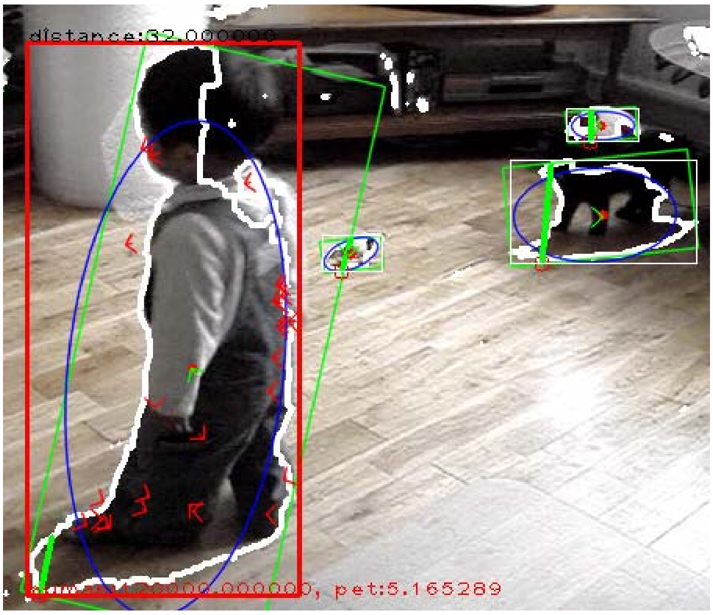


Figure 73. Evaluation with Toddler and Cat

Each foreground region was labelled with what the methods had found out about the region, like the evaluation explained in Section 4.4.1, and *a pet* was added to the group 'classification' since pets were taken into account in the human detection method in this chapter. One instance of the labels can be seen in Table 15, which is the result of the sequence where a toddler and a cat appeared together on the scene.

Table 15. Labels on Toddler and Pet for Method Evaluation

Frame No.	Previous Contour No.	Current Contour No.	Merge /Split	Correct?	Classification	Correct?	Toddler Status	Correct?
1	0	0	"No Merge/Split"	Y	"Moving Clutter"	N		
2	0	1	"No Merge/Split"	Y	"Moving Clutter"	N		
3	1	1	"Split of a single Object"	Y	"Moving Clutter"	N		
	1	2	"Split of a single Object"		"Moving Clutter"			
	1	3	"Split of a single Object"		"Moving Clutter"			
	1	4	"Split of a single Object"		"Stationary Clutter"			
4	4	1	"Reunion of Splits of a single Object"	Y	"Moving Clutter"	N		
	3	3	"No Merge/Split"		"Moving Clutter"			
5	1	1	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
6	2	0	"Reunion of Splits of a single Object"	Y	“Toddler”	Y	"Safe on the Floor"	Y
7	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
8	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	1	1	"No Merge/Split"		"Moving Clutter"	Y		
9	0	0	"Split of a single Object"	Y	“Toddler”	Y	"Safe on the Floor"	
	0	1	"Split of a single Object"		“Toddler”		"Off the Floor"	N
	1	2	"No Merge/Split"		"Moving Clutter"	Y		
10	0	1	"Reunion of Splits of a single Object"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	2	2	"No Merge/Split"		"Moving Clutter"	Y		
11	2	2	"No Merge/Split"	Y	"Moving Clutter"			

Table 15. Labels on Toddler and Pet for Method Evaluation (Continued)

12	1	1	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	2	2	"No Merge/Split"		"Moving Clutter"	Y		
13	1	0	"No Merge/Split"	Y	"Moving Clutter"	N		
	2	1	"No Merge/Split"		"Moving Clutter"	Y		
14~ 15	0	0	"No Merge/Split"	Y	"Moving Clutter"	N		
	1	1	"No Merge/Split"		"Moving Clutter"	Y		
16	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	1	1	"No Merge/Split"		"Stationary Clutter"	Y		
17	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	1	1	"No Merge/Split"		"Moving Clutter"	Y		
18	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	1	1	"No Merge/Split"		"Moving Clutter"	Y		
19	0	0	"No Merge/Split"	Y	"Moving Clutter"	N		
	1	1	"No Merge/Split"		"Moving Clutter"	Y		
20	0	0	"Split of a single Object"	Y	"Moving Clutter"	N		
	0	1	"Split of a single Object"		Clutter			
	1	2	"No Merge/Split"		"Moving Clutter"	Y		
21	0	0	"No Merge/Split"	Y	"Moving Clutter"	N		
	1	1	"No Merge/Split"		Clutter			
	2	2	"No Merge/Split"		"Moving Clutter"	Y		
22	1	0	"Reunion of Splits of a single Object"	Y	"Moving Clutter"	N		
	2	1	"No Merge/Split"		"Moving Clutter"	Y		
23~ 27	0	0	"No Merge/Split"	Y	"Moving Clutter"	N		
	1	1	"No Merge/Split"		"Moving Clutter"	Y		
28	0	0	"Split of a single Object"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	0	1	"Split of a single Object"		"Moving Clutter"	Y		
	1	2	"No Merge/Split"		"Moving Clutter"			
29	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	1	1	"No Merge/Split"		"Moving Clutter"	Y		
	2	2	"No Merge/Split"		"Moving Clutter"			

Table 15. Labels on Toddler and Pet for Method Evaluation (Continued)

30~ 39	1	0	"Reunion of Splits of a single Object"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	2	1	"No Merge/Split"		"Moving Clutter"	Y		
40	1	0	"Reunion of Splits of a single Object"	N	“Toddler”	Y	"Safe on the Floor"	Y
41~ 44	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
45	0	0	"No Merge/Split"	Y	"Moving Clutter"	N		
46~ 47	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
48~ 54	0	0	"No Merge/Split"	Y	"Moving Clutter"	N		
55~ 58	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
59	0	0	"Split of a single Object"	Y	“Toddler”	Y	"Safe on the Floor"	
	0	1	"Split of a single Object"	Y	“Toddler”	Y	"Off the Floor"	N
60	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	
	1	1	"No Merge/Split"		“Toddler”		"Off the Floor"	N
61	1	0	"Reunion of Splits of a single Object"	Y	“Toddler”	Y	"Safe on the Floor"	Y
62~ 64	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
65	0	0	"Split of a single Object"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	0	1	"Split of a single Object"		“Toddler”		"Safe on the Floor"	
66~ 67	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	1	1	"No Merge/Split"		“Toddler”		"Safe on the Floor"	
68~ 70	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
71	0	0	"Split of a single Object"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	0	1	"Split of a single Object"		“Toddler”		"Safe on the Floor"	
72	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Safe on the Floor"	Y
	1	1	"No Merge/Split"		“Toddler”		"Safe on the Floor"	

Table 15. Labels on Toddler and Pet for Method Evaluation (Continued)

73	0	0	"No Merge/Split"	Y	“Toddler”	Y	"Approaching the Furniture"	N
	1	1	"No Merge/Split"		“Toddler”		"Safe on the Floor"	

The labels were checked to see if they were correct, and the success rates were calculated and displayed in Table 16. The method for handling regional merges and splits worked well, as it did during the previous evaluation in Section 4.4.1. The toddler detection method performed adequately here as well, but the pets were classified as clutter in more than half of the total frames. This is acceptable, however, because a pet can become a tripping hazard like clutter and this research generally focuses on a toddler. The ‘toddler status’ was less correctly recognised than in the previous evaluation. This is because the toddler region sometimes split and the ‘toddler status’ became *approaching furniture* or *off the floor*, indicating fall-related risk status, when the vertically lowest point of any of the split regions was found close to or off the floor boundary.

Table 16. Results of Method Evaluation

<table>
<tr><th rowspan="2">Foreground Object on the Scene</th><th rowspan="2">Total Frame No.</th><th colspan="4">Number (Percentage) of Frames Correctly Recognised</th></tr>
<tr><th>Merges/Splits</th><th colspan="2">Classification</th><th>Toddler Status</th></tr>
<tr><td>Dog</td><td>836</td><td>828
(99.04%)</td><td colspan="2">370(as Pet) +
455(as Clutter)
(98.68%)</td><td></td></tr>
<tr><td>Toddler</td><td>376</td><td>364
(96.81%)</td><td colspan="2">343
(91.22%)</td><td>284
(82.8%)</td></tr>
<tr><td rowspan="2">Cat + Toddler</td><td rowspan="2">570</td><td rowspan="2">560
(98.25%)</td><td>Pet</td><td>Toddler</td><td rowspan="2">393
(89.93%)</td></tr>
<tr><td>31(as Pet) +
505(as Clutter)
(94.04%)</td><td>437
(76.67%)</td></tr>
</table>

5.5 Chapter Summary

This chapter illustrated a novel approach to reinforcing the method of detecting humans using the dynamic internal motion vectors introduced in Chapter 4 by considering pets, whose body parts also move diversely. The method considering pets in classification is applied, after filtering of clutter based on diverse internal motion vectors, and also uses dynamic cues, which are the history of actual heights and angle changes of the ellipse fitted to a foreground region.

The approach starts with analysing a toddler sample video to set up the reference probabilities of being the toddler. After the camera is calibrated with a chessboard pattern placed on the scene of the video, the foreground regions are segmented by background subtraction, and clutter regions are discarded by thresholding in internal motion vectors at each frame of the video. Then the actual height and the angle of ellipse fitted to each foreground contour are estimated, and the actual height is saved with the mean average and the standard deviation of all the past ellipse angles. This whole process except the camera calibration is repeated at every frame of the video, and then all the saved data are averaged by range for construction of the reference possibilities for the toddler. A pet sample video is put through the same process to build the reference probabilities of being the pet.

After construction of a toddler set and a pet set of reference probabilities, the system begins to analyse the real-time images captured by a webcam. Each real-time image also goes through the same process to discard clutter and collect data regarding the actual height and the ellipse angle of every foreground object. In the case of the ellipse angle, the mean average and the standard deviation of the data from all the past frames are calculated. Each piece of data is linked to a quantified possibility of being a toddler and a possibility of being a pet from the reference probabilities. Those two possibilities are individually applied to calculate certainties of being a toddler and a pet (C^h and C^p of Equation (26)) based on Rényi's entropy, and the two values are compared with each other to classify the region into the category with the bigger value. This classification works well even when a toddler occasionally looks small like a pet owing to his or her bent body because the entire data from the current and past frames are employed in the certainty calculations.

The methods in this chapter were evaluated by case studies of a toddler and a pet solely or together. The toddler was adequately detected, but the pet was classified as clutter in many frames. This, however, can be allowed since a pet can become a tripping hazard like clutter. Hence, the methods introduced in this chapter are satisfactory to enhance the human classification part of the methods for detection of fall risk factors delineated in Chapter 4. The evaluation, however, shows that errors in background subtraction affect the highest level of image analysis in this work, recognition of the toddler status regarding fall risks and thus are a major issue.

Chapter 6
Conclusions

6.1 Introduction

The work presented in this thesis aims to investigate and develop image analysis methods to recognise risk factors of a toddler's fall injuries in an indoor home environment in real time. The research objectives are identification of key factors that cause a toddler's fall injuries and investigation and development of vision-based analysis methods to separate a toddler from other foreground objects and to collect information related to them for recognising the factors. This research has developed and tested two novel approaches to detecting humans in images using dynamic motion cues.

The research's contributions to knowledge are listed in Section 6.2 and its limitations in Section 6.3. Future research to overcome the limitations and other issues is introduced in Section 6.4.

6.2 Contributions to Knowledge

First of all, three risk factors of a toddler's fall injury have been identified for continuous observation in injury prevention. This identification was based on both analyses of official suggestions from organisations for child safety and a large number of toddler fall records collected from hospitals. An injury epidemiological framework was used to analyse the fall records.

The major contribution of this research is novel computational methods of detecting a human in images. As shown in Section 2.3.1G, most of the previous vision-based studies for human detection use cues related to human appearance such as skin colour or body geometry. The use of those cues, however, would have problems matching the cues with a human body under partial occlusion. Although a complex human model can be used to overcome the occlusion problems, it is generally expensive in computation terms. Therefore, this research employed dynamic motion cues to differentiate a human from indoor clutter and a pet. The cues are also different from the motion cues already used in some existing studies on human detection, and the method turned out to be effective as long as every foreground object was correctly segmented.

One of the typical problems of tracking is regional merge and split owing to occlusions or errors in foreground segmentation. This research proposed a novel approach to detecting the merging or splitting regions by finding the nearest regions between two successive frames twice, from the first frame to the second frame and vice versa. This simple method worked well for detecting regional merges and splits.

6.3 Limitations of Research

As can be seen in Section 4.4.1 and Section 5.4.1, the toddler videos used for method evaluation by case studies only contain a toddler playing safely within the floor area. This is because it is not acceptable to put or leave the toddler in danger to assess performance of the methods for fall risk detection.

Most of the technical limitations of the methods in this research are related to foreground segmentation in images. Since standstill clutter on the floor as well as a toddler has to be detected to recognise a tripping or slipping hazard and a toddler’s potential behaviours for fall injuries, this research could not apply any segmentation methods based on motion or skin colour but applied background subtraction methods. As methods of comparing current images with a background model, background subtraction methods should update the model whenever any change occurs in the background owing to illumination variance, for instance. The current method in this research, however, can automatically update the background model only when the scene is clear of foreground objects, and this may become a general problem when the system is run for a long time.

The background subtraction method employed in this research relies on intensity difference that causes problems such as inability to distinguish between objects and backgrounds with similar intensities. A single object whose intensities are partially similar to its background can correspond to multiple regions segmented by the background subtraction method. Multiple objects also can appear to be a connected component by occlusion. Such regional merges and splits were handled well by the novel method in this thesis but sometimes it failed to recognise whether those regions correspond to a single object or multiple objects.

An application-type limitation is that this research never attempted to distinguish between a toddler and an adult caregiver in images on the assumption that a toddler is left alone on the scene when the system runs. It aims, however, to aid the supervision of a caregiver who stays in the same house but cannot always watch a toddler carefully enough to prevent fall injuries. Hence, it would be better if the system developed in this research can differentiate a toddler from a caregiver, in case the caregiver appears on the scene but does not give full attention to the toddler.

6.4 Future Research

With respect to the first limitation of this research described in the previous section, *update of a background model*, most of the existing studies [177-184] try automatically to update a model of dynamic background, with focus on periodic motions such as swaying trees, which may be confused with moving foreground objects. This research, however, also needs to detect static clutter, which would become part of the background using those methods for background update. For the particular system proposed in this research, another background update method needs to be developed to cope with static foreground objects.

For better results, the work in this thesis needs to employ a robust background subtraction method to segment almost perfectly foreground objects in images so as not to incur further problems such as a single object corresponding to multiple regions. A single object can still, however, split into multiple regions by being occluded by a thin background object. Thus, the handling of the regional merges and splits may need to be able to distinguish between a single object and multiple objects.

In order to be effectively used for a long time, the system proposed in this research needs to be able to classify human regions into a toddler and a caregiver to cover the case where a caregiver appears with a toddler on the scene but does not give full attention to the toddler. It would, furthermore, be advantageous if the system could automatically turn off when the caregiver allows the toddler's behaviours which might cause fall injuries, for advancement of motor skills under thorough supervision.

The research presented in this thesis deals only with fall injuries since falls account for almost half of accidental injuries to children in the home [1]. There are, however, other causes of home injuries such as burns and scalds, poisoning, drowning, and animal bites [3], whose risk factors could be recognised by vision-based analysis. Multiple cameras could also be used to cover all the house areas where a toddler generally moves around.

References

[1] CAPT, "Factsheet - Home Accidents", Child Accident Prevention Trust, London 2008.

[2] CAPT, "Factsheet - Child Accident Facts", Child Accident Prevention Trust, London 2008.

[3] KidsafeWA, "Kidsafe Home: A Community Action Kit for Home Safety", Kidsafe Western Australia, Subico 2002.

[4] E. Towner, T. Dowswell, C. Mackereth, and S. Jarvis, "What Works in Preventing Unintentional Injuries in Children and Young Adolescents?" Health Development Agency, London 2001.

[5] L. M. Millward, A. Morgan, and M. P. Kelly, "Prevention and Reduction of Accidental Injury in Children and Older People", Health Development Agency, London 2003.

[6] J. Lindon, *Child Development from Birth to Eight*. London: National Children's Bureau, 1993.

[7] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The Active Badge Location System", *Acm Transactions on Information Systems,* vol. 10*,* no. 1, pp. 91-102, 1992.

[8] J. Yang, W. Yang, M. Denecke, and A. Waibel, "Smart Sight: A Tourist Assistant System", in *the 3rd International Symposium on Wearable Computers*, San Francisco, USA, 1999, pp. 73-78.

[9] M. H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24*,* no. 1, pp. 34-58, 2002.

[10] T. B. Moeslund, A. Hilton, and V. Kruger, "A Survey of Advances in Vision-Based Human Motion Capture and Analysis", *Computer Vision and Image Understanding,* vol. 104*,* no. 2-3, pp. 90-126, 2006.

[11] H. Sidenbladh, "Detecting Human Motion with Support Vector Machines", in *the 17th International Conference on Pattern Recognition*, Cambridge, UK, 2004, pp. 188-191.

[12] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[13] T. Zhao and R. Nevatia, "Tracking Multiple Humans in Complex Situations", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 26*,* no. 9, pp. 1208-1221, 2004.

[14] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian Detection in Crowded Scenes", in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, USA, 2005, pp. 878-885.

[15] G. F. Fuller, "What Causes Falls in the Elderly? How Can I Prevent a Fall?" *American Family Physician,* vol. 61*,* p. 2173, 1 Apr. 2000.

[16] S. Luo and Q. Hu, "A Dynamic Motion Pattern Analysis Approach to Fall Detection", in *IEEE International Workshop on Biomedical Circuits and Systems*, 2004, pp. 1- 5-8a.

[17] B. Jansen and R. Deklerck, "Context Aware Inactivity Recognition for Visual Fall Detection", in *Pervasive Health Conference and Workshops*, 2006, pp. 1-4.

[18] J. Y. Hwang, J. M. Kang, Y. W. Jang, and H. C. Kim, "Development of Novel Algorithm and Real-Time Monitoring Ambulatory System using Bluetooth Module for Fall Detection in the Elderly", in *the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Francisco, USA, 2004, vol. 26*,* pp. 2204-2207.

[19] J. Chen, K. Kwong, D. Chang, J. Luk, and R. Bajcsy, "Wearable Sensors for Reliable Fall Detection", in *the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Shanghai, China, 2005, pp. 3551-3554.

[20] T. Tamura, J. X. Zhou, H. Mizukami, and T. Togawa, "A System for Monitoring Temperature Distribution in Bed and its Application to the Assessment of Body Movement", *Physiological Measurement,* vol. 14*,* no. 1, pp. 33-41, 1993.

[21] T. Harada, A. Saito, T. Sato, and T. Mori, "Infant Behavior Recognition System Based on Pressure Distribution Image", in *IEEE International Conference on Robotics and Automation*, San Francisco, USA, 2000, vol. 4*,* pp. 4082-4088.

[22] 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", in *the 2nd IEEE International Conference on Information and Communication Technologies*, Damascus, Syria, 2006, vol. 1*,* pp. 1003-1007.

[23] L. A. Wang, W. M. Hu, and T. N. Tan, "Recent Developments in Human Motion Analysis", *Pattern Recognition,* vol. 36*,* no. 3, pp. 585-601, 2003.

[24] A. Sixsmith and N. Johnson, "A Smart Sensor to Detect the Falls of the Elderly", *IEEE Pervasive Computing,* vol. 3*,* no. 2, pp. 42-47, 2004.

[25] J. P. Foster, M. S. Nixon, and A. Prugel-Bennett, "Automatic Gait Recognition using Area-Based Metrics", *Pattern Recognition Letters,* vol. 24*,* no. 14, pp. 2489-2497, 2003.

[26] L. Wang, H. Z. Ning, T. N. Tan, and W. M. Hu, "Fusion of Static and Dynamic Body Biometrics for Gait Recognition", *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 14*,* no. 2, pp. 149-158, 2004.

[27] C. Stauffer and W. E. L. Grimson, "Learning Patterns of Activity using Real-Time Tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22*,* no. 8, pp. 747-757, 2000.

[28] H. Nait-Charif and S. J. McKenna, "Activity Summarisation and Fall Detection in a Supportive Home Environment", in *the 17th International Conference on Pattern Recognition*, 2004, vol. 4*,* pp. 323-326.

[29] N. Noury, A. Dittmar, C. Corroy, R. Baghai, J. L. Weber, D. Blanc, F. Klefstat, A. Blinovska, S. Vaysse, and B. Comet, "A Smart Cloth for Ambulatory Telemonitoring of Physiological Parameters and Activity: The VTAMN Project", in *the 6th International Workshop on Enterprise Networking and Computing in Healthcare Industry*, 2004, pp. 155-160.

[30] 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*,* no. 1, pp. 42-54, 2005.

[31] F. R. Allen, E. Ambikairajah, N. H. Lovell, and B. G. Celler, "An Adapted Gaussian Mixture Model Approach to Accelerometry-Based Movement Classification Using Time-Domain Features", in *the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, USA, 2006, pp. 3600-3603.

[32] D. Anderson, J. M. Keller, M. Skubic, X. Chen, and Z. He, "Recognizing Falls from Silhouettes", in *the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, USA, 2006, pp. 6388-6391.

[33] S.-G. Miaou, P.-H. Sung, and C.-Y. Huang, "A Customized Human Fall Detection System Using Omni-Camera Images and Personal Information", in *the 1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare*, Arlington, USA, 2006, pp. 39-42.

[34] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Monocular 3D Head Tracking to Detect Falls of Elderly People", in *the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, USA, 2006, pp. 6384-6387.

[35] M.-L. Wang, C.-C. Huang, and H.-Y. Lin, "An Intelligent Surveillance System Based on an Omnidirectional Vision Sensor", in *IEEE Conference on Cybernetics and Intelligent Systems*, Bangkok, Thailand, 2006, pp. 1-6.

[36] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Fall Detection from Human Shape and Motion History Using Video Surveillance", in *the 21st International Conference on Advanced Networking and Applications Workshops/Symposia*, Niagara Falls, Canada, 2007, vol. 2*,* pp. 875-880.

[37] Z. M. Fu, E. Culurciello, P. Lichtsteiner, and T. Delbruck, "Fall Detection Using an Address-Event Temporal Contrast Vision Sensor", in *IEEE International Symposium on Circuits and Systems*, New York, USA, 2008, pp. 424-427.

[38] M. N. Nyan, F. E. H. Tay, and M. Z. E. Mah, "Application of Motion Analysis System in Pre-Impact Fall Detection", *Journal of Biomechanics* vol. 41*,* no. 10, pp. 2297-2304, 2008.

[39] A. K. Bourke and G. M. Lyons, "A Threshold-Based Fall-Detection Algorithm Using a Bi-Axial Gyroscope Sensor", *Medical Engineering & Physics,* vol. 30*,* no. 1, pp. 84-90, 2008.

[40] R. Casas, A. Marco, I. Plaza, Y. Garrido, and J. Falco, "ZigBee-based alarm system for pervasive healthcare in rural areas", *IET Communications,* vol. 2*,* no. 2, pp. 208-214, 2008.

[41] Y. Lee and M. Lee, "Accelerometer Sensor Module and Fall Detection Monitoring System Based on Wireless Sensor Network for e-Health Applications", *Telemedicine and e-Health,* vol. 14*,* no. 6, pp. 587-592, 2008.

[42] G. Wu and S. W. Xue, "Portable Preimpact Fall Detector with Inertial Sensors", *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 16*,* no. 2, pp. 178-183, 2008.

[43] K. Fukaya, "Fall Detection Sensor for Fall Protection Airbag", in *the 41st Annual Conference of the Society of Instrument and Control Engineers (SICE 2002)*, Osaka, Japan, 2002, pp. 419-420.

[44] D. P. Colvin, C. J. Lord, G. G. Bishop, T. W. Engel, and A. L. Patra, "A Fall Intervention Mobility Aid System for Elderly and Rehabilitative Populations", in *the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Orlando, USA, 1991, pp. 1936-1937.

[45] K. Ashby and M. Corbo, "Child Fall Injuries: An Overview", Monash University Accident Research Centre, Victoria Hazard (Edition No. 44), September 2000.

[46] E. Cassell and A. Clapperton, "Preventing Unintentional Injury in Victorian Children Aged 0-14 Years: A Call to Action", Monash University Accident Research Centre, Victoria Hazard (Edition No. 65), Autumn 2007.

[47] SafeKidsWorldwide, "What We Do", Safe Kids Worldwide, available from www.safekids.org/ [Accessed: 19 Jan. 2009]

[48] CAPT, "Factsheet - Child Accident Prevention Trust", Child Accident Prevention Trust, London 2008.

[49] CAPT, "Active Steps to Safety", Child Accident Prevention Trust, London 2004.

[50] CAPT, "How Safe is Your Child from a Serious Fall?" Child Accident Prevention Trust, London 2008.

[51] SafeKidsWorldwide, "About Us", Safe Kids Worldwide, available from http://www.safekids.org/about/about.html [Accessed: 19 Jan. 2009]

[52] KidsafeNSW, "Falls Prevention", Kidsafe New South Wales, available from http://www.kidsafensw.org/homesafety/falls_prevention.htm [Accessed: 19 Jan. 2009]

[53] KidsafeQLD, "Factsheet - Falls", Kidsafe Queensland, Herston 2006.

[54] KidsafeWA, "Factsheet - Toddlers (1 and 2 year olds)", Kidsafe Western Australia, Subico 2005.

[55] SafeKidsWorldwide, "Safety Tips - Falls Safety", Safe Kids Worldwide, available from http://www.safekids.org/tips/tips_falls.html [Accessed: 19 Jan. 2009]

[56] SafekidsNZ, "For Parents & Caregivers - Preventing Falls in and around the Home", Safekids New Zealand, available from http://www.safekids.org.nz/index.php/pi_pageid/104 [Accessed: 19 Jan. 2009]

[57] ECSA, "Priorities for Child Safety in the European Union : Agenda for Action", European Child Safety Alliance, Amsterdam 2004.

[58] ECSA, "Keeping Children Safe at Home: Falls", European Child Safety Alliance, EuroSafe, Amsterdam 2006.

[59] MUARC, "Annual Report 2007", Monash University Accident Research Centre, Victoria 2008.

[60] N. Paramanls and K. Harvey, "Preventing Injury. Saving Lives." Monash University Accident Research Centre, Victoria 2007.

[61] A. Gunatilaka, A. Clapperton, and E. Cassell, "Preventing Home Fall Injuries: Structural and Design Issues and Solutions", Monash University Accident Research Centre, Victoria Hazard (Edition No. 59), Summer 2005.

[62] V. Routley, "Injuries in Child Care Settings", Monash University Accident Research Centre, Victoria Hazard (Edition No. 16), September 1993.

[63] CDC, "About CDC", Centers for Disease Control and Prevention, available from http://www.cdc.gov/about/ [Accessed: 19 Jan. 2009]

[64] K. L. Smith, "Factsheet - Falls in the Home", Ohio State University Extension, Columbus AEX-691.1-92, 1992.

[65] HSC, "Home Safety Council Study: Falls are Leading Cause of Home Injury, but Americans' Home Safety Concerns may be Misguided", Home Safety Council, available from http://www.homesafetycouncil.org/home/home_sep05_w001.aspx [Accessed: 19 Jan. 2009]

[66] NAMIC, "Preventing Falls in the Home", National Association of Mutual Insurance Companies, available from http://www.namic.org/consumer/falls.asp [Accessed: 19 Jan. 2009]

[67] B. Fan and Z. F. Wang, "Pose Estimation of Human Body Based on Silhouette Images ", in *International Conference on Information Acquisition*, Hefei, China, 2004, pp. 296-300.

[68] D. Schleicher, L. M. Bergasa, R. Barea, and E. Lopez, "People Tracking and Recognition Using the Multi-Object Particle Filter Algorithm and Hierarchical PCA Method", in *Eurocon 2005: the International Conference on Computer as a Tool*, Belgrade, Serbia Monteneg, 2005, vol. 2, pp. 999-1002.

[69] G. H. Dunteman, *Principal Components Analysis*. SAGE Publications, 1989.

[70] D. Ramanan, D. A. Forsyth, and A. Zisserman, "Tracking People by Learning their Appearance", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, no. 1, pp. 65-81, 2007.

[71] N. D. Bird, O. Masoud, N. P. Papanikolopoulos, and A. Isaacs, "Detection of Loitering Individuals in Public Transportation Areas", *IEEE Transactions on Intelligent Transportation Systems,* vol. 6, no. 2, pp. 167-177, 2005.

[72] P. Fihl, R. Corlin, S. Park, T. B. Moeslund, and M. M. Trivedi, "Tracking of Individuals in Very Long Video Sequences", *Advances in Visual Computing,* vol. 4291, pp. 60-69, 2006.

[73] A. S. Micilotta, "Detection and Tracking of Humans for Visual Interaction", in Centre for Vision Speech and Signal Processing, School of Electronics and Physical Sciences. PhD Guildford, UK: University of Surrey, 2005.

[74] M. T. Yang, S. C. Wang, and Y. Y. Lin, "A Multimodal Fusion System for People Detection and Tracking ", *International Journal of Imaging Systems and Technology,* vol. 15*,* no. 2, pp. 131-142, 2005.

[75] S. Pszczolkowski and A. Soto, "Human Detection in Indoor Environments Using Multiple Visual Cues and a Mobile Robot", in *Progress in Pattern Recognition, Image Analysis and Applications*. vol. 4756, L. Rueda, D. Mery, and J. Kittler, Eds. Berlin: Springer-Verlag Berlin, 2007, pp. 350-359.

[76] C. Choi, J. Ahn, S. Lee, and H. Byun, "Disparity Weighted Histogram-Based Object Tracking for Mobile Robot Systems", in *Advances in Artificial Reality and Tele-Existence*. vol. 4282 Berlin: Springer-Verlag Berlin, 2006, pp. 584-593.

[77] S. Ammouri and G. A. Bilodeau, "Face and Hands Detection and Tracking Applied to the Monitoring of Medication Intake ", in *the 5th Canadian Conference on Computer and Robot Vision*, Windsor, Canada, 2008, pp. 147-154.

[78] B. D. Kang, J. S. Eom, J. H. Kim, C. S. Kim, S. H. Ahn, B. J. Shin, and S. K. Kim, "Human Motion Modeling Using Multivision ", in *Human-Computer Interaction*. vol. 4552, J. A. Jacko, Ed. Berlin: Springer-Verlag Berlin, 2007, pp. 659-668.

[79] G. Medioni, A. R. J. Francois, M. Siddiqui, K. Kim, and H. Yoon, "Robust Real-Time Vision for a Personal Service Robot", *Computer Vision and Image Understanding,* vol. 108*,* no. 1-2, pp. 196-203, 2007.

[80] Q. C. Pham, Y. Dhome, L. Gond, and P. Sayd, "Video Monitoring of Vulnerable People in Home Environment", in *Smart Homes and Health Telematics*. vol. 5120, S. Helal, S. Mitra, J. Wong, C. Chang, and M. Mokhtari, Eds. Berlin: Springer-Verlag Berlin, 2008, pp. 90-98.

[81] S. Pellegrini and L. Iocchi, "Human Posture Tracking and Classification through Stereo Vision and 3D Model Matching", *Journal on Image and Video Processing,* vol. 8*,* no. 2, 2008.

[82] R. Kehl and L. V. Gool, "Markerless Tracking of Complex Human Motions from Multiple Views", *Computer Vision and Image Understanding,* vol. 104*,* no. 2-3, pp. 190-209, 2006.

[83] B. Wu and R. Nevatia, "Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet Based Part Detectors", *International Journal of Computer Vision,* vol. 75*,* no. 2, pp. 247-266, 2007.

[84] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, USA, 2005, vol. 1*,* pp. 886-893.

[85] P. Viola, M. J. Jones, and D. Snow, "Detecting Pedestrians Using Patterns of Motion and Appearance", *International Journal of Computer Vision,* vol. 63*,* no. 2, pp. 153-161, 2005.

[86] S. Kwak, B. Ko, and H. Byun, "Salient Human Detection for Robot Vision", *Pattern Analysis and Applications*, vol. 10*,* no. 4, pp. 291-299, 2007.

[87] H. Schneiderman and T. Kanade, "Object Detection Using the Statistics of Parts", *International Journal of Computer Vision,* vol. 56*,* no. 3, pp. 151-177, 2004.

[88] Y. Ishii, H. Hongo, K. Yamamoto, and Y. Niwa, "Real-Time Face and Head Detection Using Four Directional Features", in *the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul, Korea, 2004, pp. 403-408.

[89] C. Fanti, L. Zelnik-Manor, and P. Perona, "Hybrid Models for Human Motion Recognition", in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, USA, 2005, vol. 1*,* pp. 1166-1173.

[90] A. G. Hochuli, L. E. S. Oliveira, A. S. Britto, and A. L. Koerich, "Detection and Classification of Human Movements in Video Scenes", in *Advances in Image and Video Technology*. vol. 4872, D. Mery and L. Rueda, Eds. Berlin: Springer-Verlag Berlin 2007, pp. 678-691.

[91] G. Mori and J. Malik, "Recovering 3D Human Body Configurations Using Shape Contexts", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28*,* no. 7, pp. 1052-1062, 2006.

[92] G. Antonini, S. V. Martinez, M. Bierlaire, and J. P. Thiran, "Behavioral Priors for Detection and Tracking of Pedestrians in Video Sequences", *International Journal of Computer Vision,* vol. 69*,* no. 2, pp. 159-180, 2006.

[93] M. Dimitrijevic, V. Lepetit, and P. Fua, "Human Body Pose Detection Using Bayesian Spatio-Temporal Templates ", *Computer Vision and Image Understanding,* vol. 104*,* no. 2, pp. 127-139, 2006.

[94] J. W. Davis, A. M. Morison, and D. D. Woods, "An Adaptive Focus-of-Attention Model for Video Surveillance and Monitoring", *Machine Vision and Applications,* vol. 18*,* no. 1, pp. 41-64, 2007.

[95] C. J. Pai, H. R. Tyan, Y. M. Liang, H. Y. M. Liao, and S. W. Chen, "Pedestrian Detection and Tracking at Crossroads ", *Pattern Recognition,* vol. 37*,* no. 5, pp. 1025-1034, 2004.

[96] R. Munoz-Salinas, E. Aguirre, and M. Garcia-Silvente, "People Detection and Tracking Using Stereo Vision and Color", *Image and Vision Computing,* vol. 25*,* no. 6, pp. 995-1007, 2007.

[97] K. Sato and J. K. Aggarwal, "Temporal Spatio-Velocity Transform and its Application to Tracking and Interaction", *Computer Vision and Image Understanding,* vol. 96*,* no. 2, pp. 100-128, 2004.

[98] W. M. Hu, T. N. Tan, L. Wang, and S. Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviors", *IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews,* vol. 34*,* no. 3, pp. 334-352, 2004.

[99] A. Utsumi, H. Mori, J. Ohya, and M. Yachida, "Multiple-View-Based Tracking of Multiple Humans", in *the 14th International Conference on Pattern Recognition*, Brisbane, Australia, 1998, pp. 597-601.

[100] A. Mittal and L. S. Davis, "M(2)Tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene", *International Journal of Computer Vision,* vol. 51*,* no. 3, pp. 189-203, 2003.

[101] J. P. Batista, "Tracking Pedestrians under Occlusion Using Multiple Cameras ", in *International Conference on Image Analysis and Recognition*, Porto, Portugal, 2004, vol. 3212*,* pp. 552-562.

[102] H. B. Kang and S. H. Cho, "Multi-Modal Face Tracking in Multi-Camera Environments", in *the 11th International Conference on Computer Analysis of Images and Patterns*, 2005, vol. 3691*,* pp. 814-821.

[103] Q. M. Zhou and J. K. Aggarwal, "Object Tracking in an Outdoor Environment Using Fusion of Features and Cameras", *Image and Vision Computing,* vol. 24*,* no. 11, pp. 1244-1255, 2006.

[104] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people", *Computer Vision and Image Understanding,* vol. 80*,* no. 1, pp. 42-56, 2000.

[105] T. H. Chen, T. Y. Chen, and Z. X. Chen, "An Intelligent People-Flow Counting Method for Passing through a Gate", in *IEEE Conference on Robotics, Automation and Mechatronics*, Bangkok, Thailand, 2006, pp. 97-102.

[106] G. Medioni, I. Cohen, F. Bremond, S. Hongeng, and R. Nevatia, "Event Detection and Analysis from Video Streams", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 23*,* no. 8, pp. 873-889, 2001.

[107] X. F. Song and R. Nevatia, "Robust Vehicle Blob Tracking with Split/Merge Handling", in *the 1st International Evaluation Workshop on Classification of Events, Activities and Relationships*, Southampton, England, 2007, vol. 4122*,* pp. 216-222.

[108] P. Guha, A. Mukerjee, K. S. Venkatesh, and P. Mitra, "Activity Discovery from Surveillance Videos", in *the 18th International Conference on Pattern Recognition*, Hong Kong, China, 2006, pp. 433-436.

[109] P. Kumar, S. Ranganath, K. Sengupta, and W. M. Huang, "Cooperative Multitarget Tracking with Efficient Split and Merge Handling", *IEEE Transactions on Circuits and Systems for Video Technology* vol. 16*,* no. 12, pp. 1477-1490, 2006.

[110] S. W. Joo and R. Chellappa, "A Multiple-Hypothesis Approach for Multiobject Visual Tracking", *IEEE Transactions on Image Processing,* vol. 16*,* pp. 2849-2854, 2007.

[111] D. Ramanan, D. A. Forsyth, and K. Barnard, "Building Models of Animals from Video", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28*,* no. 8, pp. 1319-1334, 2006.

[112] R. Polana and R. C. Nelson, "Detection and Recognition of Periodic, Nonrigid Motion", *International Journal of Computer Vision,* vol. 23*,* no. 3, pp. 261-282, 1997.

[113] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, "Recognising Human and Animal Movement by Symmetry", in *International Conference on Image Processing*, Thessaloniki, Greece, 2001, pp. 290-293.

[114] Q. Jiang and C. Daniell, "Recognition of Human and Animal Movement Using Infrared Video Streams", in *International Conference on Image Processing*, Singapore, 2004, pp. 1265-1268.

[115] Z. Y. Zhang, "A Flexible New Technique for Camera Calibration", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22*,* no. 11, pp. 1330-1334, 2000.

[116] T. N. Schoepflin and D. J. Dailey, "Dynamic Camera Calibration of Roadside Traffic Management Cameras for Vehicle Speed Estimation", *IEEE Transactions on Intelligent Transportation Systems*, vol. 4*,* no. 2, pp. 90-98, 2003.

[117] Y. Yuan, Y. Zhaoxuan, H. Yinghua, and W. Zengmin, "The Application of An Improved Camera Calibration Algorithm to the Video-Based Traffic Information Detection", in *the 6th International Conference on ITS Telecommunications*, Chengdu, China, 2006, pp. 808-811.

[118] P. Allard, I. A. F. Stokes, and J.-P. Blanchi, *Three-dimensional Analysis of Human Movement*. Human Kinetics, 1995.

[119] D. Farin, J. G. Han, and P. H. N. de With, "Fast Camera Calibration for the Analysis of Sport Sequences", in *IEEE International Conference on Multimedia and Expo*, Amsterdam, Netherlands, 2005, pp. 482-485.

[120] H. Zhong, F. Mai, and Y. S. Hung, "Camera Calibration Using Circle and Right Angles", in *the 18th International Conference on Pattern Recognition*, Hong Kong, China, 2006, vol. 1, pp. 646-649.

[121] J. S. Kim, P. Gurdjos, and I. S. Kweon, "Geometric and Algebraic Constraints of Projected Concentric Circles and their Applications to Camera Calibration", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 4, pp. 637-642, 2005.

[122] Q. Liu and H. Su, *Correction of the Asymmetrical Circular Projection in DLT Camera Calibration* vol. 2. Los Alamitos: IEEE Computer Soc 2008.

[123] J. H. Wang, F. H. Shi, J. Zhang, and Y. C. Liu, "A New Calibration Model and Method of Camera Lens Distortion", in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, 2006, pp. 5713-5718.

[124] M. Baba, M. Mukunoki, and N. Asada, "A Unified Camera Calibration Using Geometry and Blur of Feature Points", in *the 18th International Conference on Pattern Recognition*, Hong Kong, China, 2006, vol. 1, pp. 816-819.

[125] B. Zhang, Y. F. Li, and Y. H. Wu, "Self-Recalibration of a Structured Light System via Plane-Based Homography", *Pattern Recognition,* vol. 40, no. 4, pp. 1368-1377, 2007.

[126] D. Xin and Z. Xiao-guang, "A Camera Calibration Technique Based on Plane Square", in *IEEE International Geoscience and Remote Sensing Symposium*, 2005, vol. 5, pp. 3663-3666.

[127] M. Fiala and C. Shu, "Self-Identifying Patterns for Plane-Based Camera Calibration", *Machine Vision and Applications,* vol. 19, no. 4, pp. 209-216, 2008.

[128] X. H. Ying and H. B. Zha, "Geometric Interpretations of the Relation between the Image of the Absolute Conic and Sphere Images", *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, vol. 28*,* no. 12, pp. 2031-2036, 2006.

[129] N. Avinash and S. Murali, "Perspective Geometry Based Single Image Camera Calibration", *Journal of Mathematical Imaging and Vision,* vol. 30*,* no. 3, pp. 221-230, 2008.

[130] L. Grammatikopoulos, G. Karras, and E. Petsa, "An Automatic Approach for Camera Calibration from Vanishing Points", *Isprs Journal of Photogrammetry and Remote Sensing,* vol. 62*,* no. 1, pp. 64-76, 2007.

[131] K. Y. K. Wong, P. R. S. Mendonca, and R. Cipolla, "Camera Calibration from Surfaces of Revolution", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 25*,* no. 2, pp. 147-161, 2003.

[132] Z. J. Zhao and Y. C. Liu, "New Multi-Camera Calibration Algorithm Based on 1D Objects", *Journal of Zhejiang University-Science A,* vol. 9*,* no. 6, pp. 799-806, 2008.

[133] H. Chen, H. Yu, and A. Long, "A New Camera Calibration Algorithm Based on Rotating Object", in *Robot Vision*. vol. 4931, G. Sommer and R. Klette, Eds. Berlin: Springer-Verlag Berlin, 2008, pp. 403-411.

[134] F. J. Lv, T. Zhao, and R. Nevatia, "Camera Calibration from Video of a Walking Human", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28*,* no. 9, pp. 1513-1518, 2006.

[135] F. Lu, X. C. Cao, Y. P. Shen, and H. Foroosh, "Camera Calibration from Two Shadow Trajectories", in *the 18th International Conference on Pattern Recognition*, Hong Kong, China, 2006, pp. 1-4.

[136] RoSPA, "About RoSPA", The Royal Society for the Prevention of Accidents, available from http://www.rospa.com/aboutrospa/index.htm [Accessed: 19 Jan. 2009]

[137] DTI, "24th Report of the Home and Leisure Accident Surveillance System", The Royal Society for the Prevention of Accidents, London 2004.

[138] M. D. Harvey, "Models for Accident Investigation", Alberta Workers Health Safety and Compensation, Edmonton, Canada 1985.

[139] R. Barkan, D. Zohar, and I. Erev, "Accidents and Decision Making under Uncertainty: A Comparison of Four Models", *Organizational Behavior and Human Decision Processes,* vol. 74*,* no. 2, pp. 118-144, 1998.

[140] OECD, "Committee on the Safety of Nuclear Installations", in *the International Workshop Building the New HRA: Errors of Commission - From Research to Application*, Rockville, USA, 2001.

[141] M. Ferjencik and R. Kuracina, "MORT WorkSheet or How to Make MORT Analysis Easy", *Journal of Hazardous Materials,* vol. 151*,* pp. 143-154, 2008.

[142] L. Benner, "Accident Investigations: Multilinear Events Sequencing Methods", *Journal of Safety Research,* vol. 7*,* no. 2, pp. 67-73, 1975.

[143] C. W. Runyan, "Introduction: Back to the Future - Revisiting Haddon's Conceptualization of Injury Epidemiology and Prevention", *Epidemiologic Reviews,* vol. 25*,* pp. 60-64, 2003.

[144] Y. Holder, M. Peden, E. Krug, J. Lund, G. Gururaj, and O. Kobusingye, "Injury Surveillance Guidelines", World Health Organization, Geneva, Switzerland 2001.

[145] W. Haddon, "Advances in the Epidemiology of Injuries as a Basis for Public Policy", *Public Health Reports,* vol. 95*,* no. 5, pp. 411-421, 1980.

[146] B. Gillham, *Developing a Questionnaire*. London: Continuum, 2000.

[147] J. H. Joo, E. J. Lenze, B. H. Mulsant, A. E. Begley, E. M. Weber, J. A. Stack, S. Mazumdar, C. F. Reynolds, and B. G. Pollock, "Risk Factors for Falls during Treatment of Late-Life Depression", *Journal of Clinical Psychiatry,* vol. 63*,* no. 10, pp. 936-941, 2002.

[148] M. R. Perracini and L. R. Ramos, "Fall-Related Factors in a Cohort of Elderly Community Residents", *Revista De Saude Publica,* vol. 36*,* no. 6, pp. 709-716, 2002.

[149] A. H. Abdelhafiz and C. A. Austin, "Visual Factors should be Assessed in Older People Presenting with Falls or Hip Fracture", *Age and Ageing,* vol. 32*,* no. 1, pp. 26-30, 2003.

[150] S. Buatois, R. Gueguen, G. C. Gauchard, A. Benetos, and P. P. Perrin, "Posturography and Risk of Recurrent Falls in Healthy Non-Institutionalized Persons Aged over 65", *Gerontology,* vol. 52*,* no. 6, pp. 345-352, 2006.

[151] U. Laessoe, H. C. Hoeck, O. Simonsen, T. Sinkjaer, and M. Voigt, "Fall Risk in an Active Elderly Population - Can it be Assessed?" *Journal of Negative Results in BoiMedicine,* vol. 6*,* p. 2, 2007.

[152] R. Kikuchi, K. Kozaki, T. Nakamura, and K. Toba, "Muscle and Bone Health as a Risk Factor of Fall Among the Elderly. An Approach to Identify High-Risk Fallers by Risk Assessment", *Clinical Calcium,* vol. 18*,* no. 6, pp. 784-787, 2008.

[153] D. H. Solomon, "Will Absolute Fracture Risk Prediction Facilitate Treatment of Osteoporosis?" *Nature Clinic Practice Endocrinology & Metabolism,* vol. 4*,* no. 9, pp. 480-481, 2008.

[154] Y. Balash, C. Peretz, G. Leibovich, T. Herman, J. M. Hausdorff, and N. Giladi, "Falls in Outpatients with Parkinson's Disease - Frequency, Impact and Identifying Factors", *Journal of Neurology,* vol. 252*,* no. 11, pp. 1310-1315, 2005.

[155] D. A. Ganz, Y. Bao, P. G. Shekelle, and L. Z. Rubenstein, "Will my Patient Fall?" *Journal of the American Medical Association,* vol. 297*,* no. 1, pp. 77-86, 2007.

[156] S. van Helden, C. E. Wyers, P. C. Dagnelie, M. C. van Dongen, G. Willems, P. R. G. Brink, and P. P. Geusens, "Risk of Falling in Patients with a Recent Fracture", *BMC Musculoskeletal Disorders,* vol. 8*,* p. 55, 2007.

[157] M. Y. C. Pang and J. J. Eng, "Fall-Related Self-Efficacy, not Balance and Mobility Performance, is Related to Accidental Falls in Chronic Stroke Survivors with Low Bone Mineral Density", *Osteoporosis International,* vol. 19*,* no. 7, pp. 919-927, 2008.

[158] E. K. Kim, J. C. Lee, and M. R. Eom, "Falls Risk Factors of Inpatients", *Journal of Korean Academy of Nursing,* vol. 38*,* no. 5, pp. 676-684, 2008.

[159] P. F. Agran, C. Anderson, D. Winn, R. Trent, L. Walton-Haynes, and S. Thayer, "Rates of Pediatric Injuries by 3-Month Intervals for Children 0 to 3 Years of Age", *Pediatrics,* vol. 111*,* no. 6, pp. E683-E692, 2003.

[160] B. Hansoti and T. Beattie, "Can the Height of Fall Predict Long Bone Fracture in Children under 24 Months?" *European Journal of Emergency Medicine,* vol. 12*,* no. 6, pp. 285-286, 2005.

[161] Intel, "Open Source Computer Vision Library: Reference Manual", Intel Corporation 2000.

[162] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", in *the 7th International Joint Conference on Artificial Intelligence*, Vancouver, Canada, 1981, pp. 674-679.

[163] F. K. Nejadasl, B. G. H. Gorte, and S. P. Hoogendoorn, "Optical Flow Based Vehicle Tracking Strengthened by Statistical Decisions", *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 61*,* pp. 159-169, 2006.

[164] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of Optical Flow Techniques", *International Journal of Computer Vision,* vol. 12*,* no. 1, pp. 43-77, 1994.

[165] Hamleys, available from http://www.hamleys.com [Accessed: 19 Jan. 2009]

[166] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct Least Square Fitting of Ellipses", *Pattern Analysis and Machine Intelligence,* vol. 21*,* no. 5, pp. 476-480, 1999.

[167] S. F. Qin, I. N. Jordanov, and D. K. Wright, "Freehand Drawing System Using a Fuzzy Logic Concept", *Computer-Aided Design,* vol. 31*,* no. 5, pp. 359-360, 1999.

[168] J. Avison, *Physics for CXC*. Gloucestershire: Nelson Thornes Ltd., 2000.

[169] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[170] "Entropy", in *Columbia Encyclopedia*, 6th ed, P. Lagassé, Ed. New York: Columbia University Press, 2008.

[171] H. Rheingold, *Tools for Thought: The History and Future of Mind-Expanding Technology*. Cambridge, USA: The MIT Press, 2000.

[172] C. E. Shannon, "A Mathematical Theory of Communication", *The Bell System Technical Journal,* vol. 27*,* pp. 379-423, 1948.

[173] Y. Zhou and H. Leung, "Minimum Entropy Approach for Multisensor Data Fusion", in *IEEE Signal Processing Workshop on Higher-Order Statistics*, 1997, pp. 336-339.

[174] G. Kern-Isberner and W. Rödder, "Fusing Probabilistic Information on Maximum Entropy", in *KI 2003: Advances in Artificial Intelligence*. vol. 2821/2003: Springer Berlin/Heidelberg, 2003, pp. 407-420.

[175] W. Hsu and S. F. Chang, "Generative, Discriminative, and Ensemble Learning on Multi-Modal Perceptual Fusion towards News Video Story Segmentation", in *IEEE International Conference on Multimedia and Expo*, 2004, vol. 2*,* pp. 1091-1094.

[176] A. Rényi, "On Measures of Entropy and Information", in *the 4th Berkeley Symposium on Mathematics, Statistics, and Probability*, 1961, pp. 547-561.

[177] M. Seki, H. Fujiwara, and K. Sumi, "A Robust Background Subtraction Method for Changing Background", in *the 5th IEEE Workshop on Applications of Computer Vision*, 2000, pp. 207-213.

[178] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, "Foreground Object Detection from Videos Containing Complex Background", in *the 11th ACM International Conference on Multimedia*, 2003, pp. 2-10.

[179] Q. Zang and R. Klette, "Robust Background Subtraction and Maintenance", in *the 17th International Conference on Pattern Recognition*, 2004, vol. 2, pp. 90-93.

[180] T. Thongkamwitoon, S. Aramvith, and T. H. Chalidabhongse, "Non-Linear Learning Factor Control for Statistical Adaptive Background Subtraction Algorithm", in *IEEE International Symposium on Circuits and Systems*, 2005, vol. 4, pp. 3785-3788.

[181] E. Zhang, F. Chen, and W. Zhang, "A Novel Particle Filter Based Background Subtraction Method", in *International Conference on Computational Intelligence and Security*, 2006, vol. 2, pp. 1837-1840.

[182] H. Zhang and D. Xu, "Robust Estimation for Background Subtraction", in *the 1st International Conference on Innovative Computing, Information and Control*, 2006, vol. 1, pp. 660-664.

[183] P.-M. Jodoin, M. Mignotte, and J. Konrad, "Statistical Background Subtraction Using Spatial Cues", *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 17, no. 12, pp. 1758-1763, 2007.

[184] L. Maddalena and A. Petrosino, "A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications ", *IEEE Transactions on Image Processing,* vol. 17, no. 7, pp. 1168-1177, 2008.

Appendix A

Questionnaire for Evaluation of Fall Risk Factors

According to the UK Child Accident Prevention Trust, in the UK over two million children every year are taken to hospital due to accidental injuries, and around half of these accidents are domestic. Falls account for over 40% of all home accidental injuries of children, and young children aged under five are most vulnerable to injuries in the home environment, where they spend most of their time.

As young children are not able to assess risks for themselves, the best way to prevent their fall injuries would be continuous supervision and instruction from their parents. However, this is not always practical. A smart vision system is proposed in my research to assist the parents' supervision by detecting risk factors of a toddler's falls from environmental or behavioural changes within an indoor home environment so that a caregiver can be alerted to eliminate the factors for preventing fall injuries.

The identification of the fall risk factors was based on 4377 fall stories of toddlers at home, collected by the Royal Society for the Prevention of Accidents (RoSPA), and the suggestions of several related organisations[6] to prevent the falls of toddlers in the home environment. The suggestions requesting environmental modification such as installation of safety gates, window locks or non-skid rugs, were excluded because the goal of this vision system is occasional substitution of parents' supervision on situational or behavioural changes which may cause a toddler's falls. The stories from RoSPA revealed that many toddlers fell down just whilst going up or down stairs alone and could easily trip while moving around. Also their resulting impact with furniture or the edges of a room may have caused severe injuries. The suggestions of related organisations indicate similar points:

[6] UK Child Accident Prevention Trust, Safekids Worldwide, USA Home Safety Council, USA national Safety Council, European Child Safety Alliance

□ Keep floors clear of toys and other clutter which might trip toddlers when they walk around.

□ Ensure there are no sharp or hard edges near them that could cause injuries when they fall.

□ Discourage children from climbing on furniture.

Please imagine that we are looking at a home environment which has been modified (e.g. small gaps between vertical railings of banisters or balconies) and installed safety products (e.g. safety gates, window locks and non-skid rugs) enough to satisfy specialists in home safety for toddlers. In this environment a toddler is playing with toys on the floor and a parent is looking after him or her with doing housework.

Do you think the above three factors are major to be continuously supervised by the parent? Please write ‘v’ in front of each factor if you think it is major to be supervised in order to prevent the toddler’s fall injuries.

Do you think the above three factors are enough? If not, please fill the below space with another situational or behavioural changes to be constantly watched in order to prevent the toddler’s fall injuries.

Appendix B
Questionnaire Responses

<Response 1>

According to the UK Child Accident Prevention Trust, in the UK over two million children every year are taken to hospital due to accidental injuries, and around half of these accidents are domestic. Falls account for over 40% of all home accidental injuries of children, and young children aged under five are most vulnerable to injuries in the home environment, where they spend most of their time.

As young children are not able to assess risks for themselves, the best way to prevent their fall injuries would be continuous supervision and instruction from their parents. However, this is not always practical. A smart vision system is proposed in my research to assist the parents' supervision by detecting risk factors of a toddler's falls from environmental or behavioural changes within an indoor home environment so that a caregiver can be alerted to eliminate the factors for preventing fall injuries.

The identification of the fall risk factors was based on 4377 fall stories of toddlers at home, collected by the Royal Society for the Prevention of Accidents (RoSPA), and the suggestions of several related organisations[7] to prevent the falls of toddlers in the home environment. The suggestions requesting environmental modification such as installation of safety gates, window locks or non-skid rugs, were excluded because the goal of this vision system is occasional substitution of parents' supervision on situational or behavioural changes which may cause a toddler's falls. The stories from RoSPA revealed that many toddlers fell down just whilst going up or down stairs alone and could easily trip while moving around. Also their resulting impact with furniture or the edges of a room may have caused severe injuries. The suggestions of related organisations indicate similar points:

[7] UK Child Accident Prevention Trust, Safekids Worldwide, USA Home Safety Council, USA national Safety Council, European Child Safety Alliance

☑ Keep floors clear of toys and other clutter which might trip toddlers when they walk around.

☑ Ensure there are no sharp or hard edges near them that could cause injuries when they fall.

☑ Discourage children from climbing on furniture.

Please imagine that we are looking at a home environment which has been modified (e.g. small gaps between vertical railings of banisters or balconies) and installed safety products (e.g. safety gates, window locks and non-skid rugs) enough to satisfy specialists in home safety for toddlers. In this environment a toddler is playing with toys on the floor and a parent is looking after him or her with doing housework.

Do you think the above three factors are major to be continuously supervised by the parent? Please write 'v' in front of each factor if you think it is major to be supervised in order to prevent the toddler's fall injuries.

Do you think the above three factors are enough? If not, please fill the below space with another situational or behavioural changes to be constantly watched in order to prevent the toddler's fall injuries.

I am pleased that you are focusing of falls prevention among children. I would like to recommend one other factor for your consideration:

Ensure that the floor is not slippery due to the materials used for its construction (e.g., marble), cleaning product residue (e.g., oil) or liquid on the floor.

<Response 2>

According to the UK Child Accident Prevention Trust, in the UK over two million children every year are taken to hospital due to accidental injuries, and around half of these accidents are domestic. Falls account for over 40% of all home accidental injuries of children, and young children aged under five are most vulnerable to injuries in the home environment, where they spend most of their time.

As young children are not able to assess risks for themselves, the best way to prevent their fall injuries would be continuous supervision and instruction from their parents. However, this is not always practical. A smart vision system is proposed in my research to assist the parents' supervision by detecting risk factors of a toddler's falls from environmental or behavioural changes within an indoor home environment so that a caregiver can be alerted to eliminate the factors for preventing fall injuries.

The identification of the fall risk factors was based on 4377 fall stories of toddlers at home, collected by the Royal Society for the Prevention of Accidents (RoSPA), and the suggestions of several related organisations[8] to prevent the falls of toddlers in the home environment. The suggestions requesting environmental modification such as installation of safety gates, window locks or non-skid rugs, were excluded because the goal of this vision system is occasional substitution of parents' supervision on situational or behavioural changes which may cause a toddler's falls. The stories from RoSPA revealed that many toddlers fell down just whilst going up or down stairs alone and could easily trip while moving around. Also their resulting impact with furniture or the edges of a room may have caused severe injuries. The suggestions of related organisations indicate similar points:

[8] UK Child Accident Prevention Trust, Safekids Worldwide, USA Home Safety Council, USA national Safety Council, European Child Safety Alliance

☑ Keep floors clear of toys and other clutter which might trip toddlers when they walk around.

☑ Ensure there are no sharp or hard edges near them that could cause injuries when they fall.

☑ Discourage children from climbing on furniture.

Please imagine that we are looking at a home environment which has been modified (e.g. small gaps between vertical railings of banisters or balconies) and installed safety products (e.g. safety gates, window locks and non-skid rugs) enough to satisfy specialists in home safety for toddlers. In this environment a toddler is playing with toys on the floor and a parent is looking after him or her with doing housework.

Do you think the above three factors are major to be continuously supervised by the parent? Please write 'v' in front of each factor if you think it is major to be supervised in order to prevent the toddler's fall injuries.

Do you think the above three factors are enough? If not, please fill the below space with another situational or behavioural changes to be constantly watched in order to prevent the toddler's fall injuries.

- Maintenance of stairs generally, repairing damaged carpets
- Stairs should always be well lit.
- Fixing tall or heavy furniture securely to the wall; there have been instances of children being crushed through pulling items down on top of themselves.
- Babies in high chairs and prams need to be strapped in with at least a three-point harness, five-point is even better.
- Top level on bunk beds are for 6 years old and older
- Safety training can start as young as 3 years old when toddlers are open to simple rules;this training is continuous and incremental as children develop; this is one of the hardest tasks for any parent

<Response 3>

According to the UK Child Accident Prevention Trust, in the UK over two million children every year are taken to hospital due to accidental injuries, and around half of these accidents are domestic. Falls account for over 40% of all home accidental injuries of children, and young children aged under five are most vulnerable to injuries in the home environment, where they spend most of their time.

As young children are not able to assess risks for themselves, the best way to prevent their fall injuries would be continuous supervision and instruction from their parents. However, this is not always practical. A smart vision system is proposed in my research to assist the parents' supervision by detecting risk factors of a toddler's falls from environmental or behavioural changes within an indoor home environment so that a caregiver can be alerted to eliminate the factors for preventing fall injuries.

The identification of the fall risk factors was based on 4377 fall stories of toddlers at home, collected by the Royal Society for the Prevention of Accidents (RoSPA), and the suggestions of several related organisations[9] to prevent the falls of toddlers in the home environment. The suggestions requesting environmental modification such as installation of safety gates, window locks or non-skid rugs, were excluded because the goal of this vision system is occasional substitution of parents' supervision on situational or behavioural changes which may cause a toddler's falls. The stories from RoSPA revealed that many toddlers fell down just whilst going up or down stairs alone and could easily trip while moving around. Also their resulting impact with furniture or the edges of a room may have caused severe injuries. The suggestions of related organisations indicate similar points:

[9] UK Child Accident Prevention Trust, Safekids Worldwide, USA Home Safety Council, USA national Safety Council, European Child Safety Alliance

□ Keep floors clear of toys and other clutter which might trip toddlers when they walk around.

☑ Have toys in a low traffic area. (This reduces the need for as much supervision)

☑ Ensure there are no sharp or hard edges near them that could cause injuries when they fall.

☑ Discourage children from climbing on furniture.

Please imagine that we are looking at a home environment which has been modified (e.g. small gaps between vertical railings of banisters or balconies) and installed safety products (e.g. safety gates, window locks and non-skid rugs) enough to satisfy specialists in home safety for toddlers. In this environment a toddler is playing with toys on the floor and a parent is looking after him or her with doing housework.

Do you think the above three factors are major to be continuously supervised by the parent? Please write 'v' in front of each factor if you think it is major to be supervised in order to prevent the toddler's fall injuries.

Do you think the above three factors are enough? If not, please fill the below space with another situational or behavioural changes to be constantly watched in order to prevent the toddler's fall injuries.

- Ensure children are not left unattended on high surfaces such as change tables & kitchen benches. (parents often pop demanding children onto the kitchen bench while they are cooking or preparing meals and falls are a common result of this practice)
- Always use 5 point harness in high chairs and strollers.
- There has been an increase in the number of childhood injuries in Australia because of falls from Windows. Preventative measures are to place furniture away from windows and consider having window openings limited to 100mm to reduce the risk.

Appendix C
Ethical Approval

Brunel
UNIVERSITY
WEST LONDON

19 December 2007

Dear Hana Na

Thank you for your responses to the questions concerning your request for ethical approval of your project “A smart vision sensor for detecting risk factors of a toddler’s fall in a home environment”

I accept that under the circumstances it does not seem to be possible to use home-video recordings that already exist and thus your request should in principle be approved. I have two further requirements:

1. I assume that *at all times* one or more of the parents/guardians/adult carers of the child in question will be present during filming and that *at no time* will you be in sole charge of the child. If this is not true, then you will have to agree to a Criminal Records Bureau check in advance of any work undertaken. I strongly recommend that you adopt the former approach.
2. It is vitally important that to take all proper precautions to preserve the anonymity of the subject(s) that you are recording, this includes any material that you might wish to include in a PhD thesis or publication. In particular I suggest that you use pixelation or blurring of all facial features. If this is not possible then I must ask you to justify why it cannot be done. You must of course also comply with the

requirements of the Data Protection Act at all times, and all of your primary material must be kept secure (ideally encrypted and physically secure) and then destroyed as soon as is practical after the end of your research.

Yours sincerely

Professor Peter R Hobson

Chair, SED Research Ethics Committee

School of Engineering & Design
Tower D
Brunel University, Uxbridge UB8 3PH

Tel: +44(0)1895 266799
Peter.Hobson@brunel.ac.uk

Appendix D
List of Publications

Conference Publications

A number of conference publications have been written describing results obtained and are listed below:

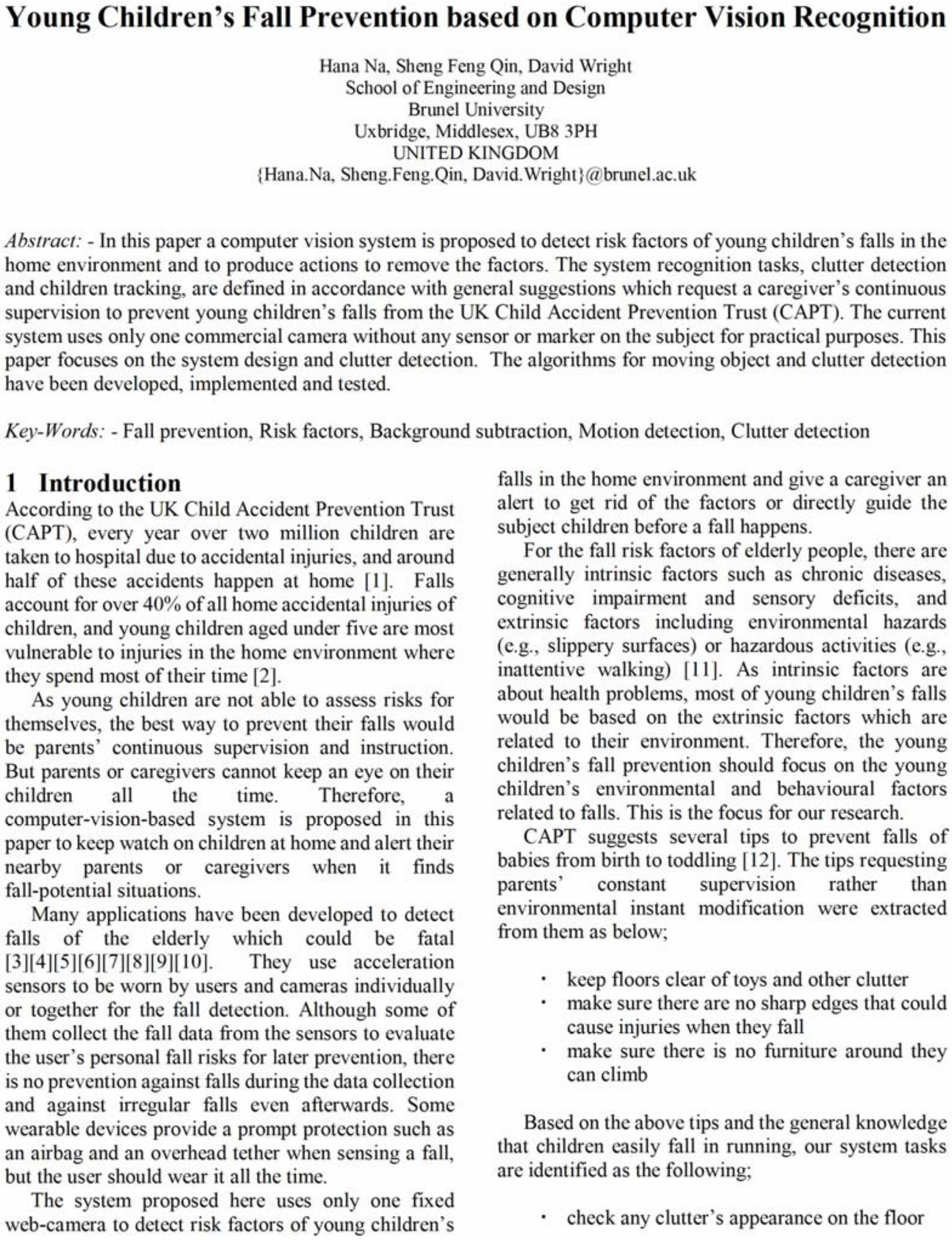
- “Young Children’s Fall Prevention based on Computer Vision Recognition”, Proceedings of the 6th WSEAS International Conference on Robotics, Control and Manufacturing Technology (ROCOM ’06), Hangzhou, China, April 16-18 2006, pp. 193-198. This paper is attached in Appendix E.
- “A Smart Vision Sensor for Detecting Risk Factors of a Toddler’s Fall in a Home Environment”, Proceedings of the 2007 IEEE International Conference on Networking, Sensing and Control (ICNSC ’07), London, UK, 15-17 April 2007, pp. 656-661. This paper is attached in Appendix F.
- “Vision-Based Tracking a Toddler at Home”, Proceedings of IEEE EUROCON 2007 the International Conference on Computer as a Tool, Warsaw, Poland, 9-12 September 2007, pp. 375-382. This paper is attached in Appendix G.

Journal Publications

A journal paper was submitted to IEEE Pervasive Computing and is awaiting review at present. This paper is attached in Appendix H.

Appendix E

WSEAS ROCOM '06



Young Children's Fall Prevention based on Computer Vision Recognition

Hana Na, Sheng Feng Qin, David Wright
School of Engineering and Design
Brunel University
Uxbridge, Middlesex, UB8 3PH
UNITED KINGDOM
{Hana.Na, Sheng.Feng.Qin, David.Wright}@brunel.ac.uk

Abstract: - In this paper a computer vision system is proposed to detect risk factors of young children's falls in the home environment and to produce actions to remove the factors. The system recognition tasks, clutter detection and children tracking, are defined in accordance with general suggestions which request a caregiver's continuous supervision to prevent young children's falls from the UK Child Accident Prevention Trust (CAPT). The current system uses only one commercial camera without any sensor or marker on the subject for practical purposes. This paper focuses on the system design and clutter detection. The algorithms for moving object and clutter detection have been developed, implemented and tested.

Key-Words: - Fall prevention, Risk factors, Background subtraction, Motion detection, Clutter detection

1 Introduction

According to the UK Child Accident Prevention Trust (CAPT), every year over two million children are taken to hospital due to accidental injuries, and around half of these accidents happen at home [1]. Falls account for over 40% of all home accidental injuries of children, and young children aged under five are most vulnerable to injuries in the home environment where they spend most of their time [2].

As young children are not able to assess risks for themselves, the best way to prevent their falls would be parents' continuous supervision and instruction. But parents or caregivers cannot keep an eye on their children all the time. Therefore, a computer-vision-based system is proposed in this paper to keep watch on children at home and alert their nearby parents or caregivers when it finds fall-potential situations.

Many applications have been developed to detect falls of the elderly which could be fatal [3][4][5][6][7][8][9][10]. They use acceleration sensors to be worn by users and cameras individually or together for the fall detection. Although some of them collect the fall data from the sensors to evaluate the user's personal fall risks for later prevention, there is no prevention against falls during the data collection and against irregular falls even afterwards. Some wearable devices provide a prompt protection such as an airbag and an overhead tether when sensing a fall, but the user should wear it all the time.

The system proposed here uses only one fixed web-camera to detect risk factors of young children's falls in the home environment and give a caregiver an alert to get rid of the factors or directly guide the subject children before a fall happens.

For the fall risk factors of elderly people, there are generally intrinsic factors such as chronic diseases, cognitive impairment and sensory deficits, and extrinsic factors including environmental hazards (e.g., slippery surfaces) or hazardous activities (e.g., inattentive walking) [11]. As intrinsic factors are about health problems, most of young children's falls would be based on the extrinsic factors which are related to their environment. Therefore, the young children's fall prevention should focus on the young children's environmental and behavioural factors related to falls. This is the focus for our research.

CAPT suggests several tips to prevent falls of babies from birth to toddling [12]. The tips requesting parents' constant supervision rather than environmental instant modification were extracted from them as below;

- keep floors clear of toys and other clutter
- make sure there are no sharp edges that could cause injuries when they fall
- make sure there is no furniture around they can climb

Based on the above tips and the general knowledge that children easily fall in running, our system tasks are identified as the following;

- check any clutter's appearance on the floor

- check if the children run too fast
- check if the children move close to any structure in their environment
- check if the children climb any furniture

From the perspective of computer vision, the above tasks could be divided into clutter detection and tracking of moving children, and this paper only presents the clutter detection.

2 Related Work

The proposed system should watch both of small objects and human beings to detect any clutter's appearance on the floor and children's fall-prone behaviours. Therefore, a human being and an object must be distinguished accurately using not only motion but also other reasonable cues. This section gives a brief overview of existing works to detect clutter and human beings based on different cues and track them individually.

Lipton et al [13] detect moving targets by using the pixel wise difference between consecutive image frames and classify them into human, vehicle or background clutter, based on the target size and the shape dispersedness as humans are smaller than vehicles and have more complex shapes. This method is somewhat simple which is good for real-time motion analysis, however it seems only good enough to distinguish humans from big vehicles and the tiny motion of trees.

VIGOUR of Sherrah and Gong [14] finds skin colour clusters and tracks three boxes respectively bounding a head and two hands for one person. The head box tracker is initialised using Support Vector Machine face detection and the hand box trackers are initialised heuristically with respect to the head position for tracking of multiple people. VIGOUR also uses a simple method using a colour cue, but the subjects should be initially facing the camera and the faces should not be occluded.

The single view tracking of Cai and Aggarwal [15] is composed of background subtraction, human segmentation and the human feature correspondence between adjacent frames. After the background subtraction, human and non-human moving regions are distinguished using moment invariants based on Principal Component Analysis (PCA). And location, intensity and geometric information of the human are extracted for the tracking. The use of the three features to track human achieves much better tracking than the use of any individual feature, but the occlusion is a major obstacle.

After background subtraction, Schleicher et al [16] use a Particle Filter (PF) algorithm to identify and individually track any moving objects, and apply PCA to the each object to classify into person or non-person using geometrical constraints of several body parts. This system is relatively robust at occlusion but long-term occlusion and the lateral view of a person cause some failures.

Micilotta [17] also uses PF to track each human body after fitting a torso primitive to foreground regions and segmenting skin tone regions for the face and hands. Meanwhile, he presents a more robust method of tracking a human. Body part detectors trained by AdaBoost, detect several body parts by using skin colour cues to reduce false detections, and RANSAC assembles the parts into body configurations.

The more cues are used to detect and track human beings, the more accurate the results are. However, the use of many cues or complex methods would require expensive computation and take too long for real-time applications.

3 System Overview

The objectives of clutter detection are to find standstill objects on the floor which might trip children and to find objects which are thrown towards and hit children. The current system detects any appearance of clutter and just discriminates between standstill objects on the floor and moving objects in the scene. The information about the moving objects will be used to check if they are moving towards children after the future work, tracking of children is done.

The clutter detection consists of background subtraction, updating background image and motion detection to detect and classify clutter, and respective action against moving and standstill objects. The system workflow is presented below.

Firstly, background subtraction detects anything on the floor by comparing current images and an initial background image containing a clear floor. To focus on the floor area, non-floor area is masked in the background image.

The background image needs to keep being updated due to extraneous changes like the illumination variance. There might be small motions in the background, but they are not general in the indoor environment like swaying branches outdoors and the system only deals with lighting changes. The slight changes of the sun light are discarded using thresholding and the dramatic changes in switching on/off a lamp are judged as an overall change.

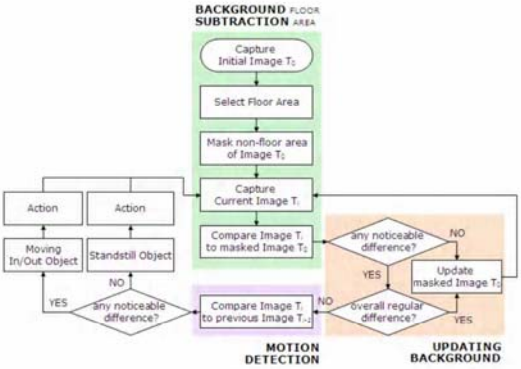

BACKGROUND SUBTRACTION
FLOOR AREA
Capture Initial Image T_0
Select Floor Area
Mask non-floor area of Image T_0
Capture Current Image T_i
Compare Image T_i to masked Image T_0
any noticeable difference?
NO
YES
Update masked Image T_0
overall regular difference?
YES
NO
Compare Image T_i to previous Image T_{i-1}
any noticeable difference?
NO
YES
Action
Action
Moving In/Out Object
Standstill Object
MOTION DETECTION
UPDATING BACKGROUND

Fig. 1 System Work Flow

In the detection of any standstill objects on the floor, background subtraction is not enough because it detects both standstill and moving regions if they do not belong to the background image. So to classify standstill and moving regions, comparison of consecutive frames to detect motion is used.

Lastly, the system produces different actions in detection of still objects on the floor and moving objects in the scene.

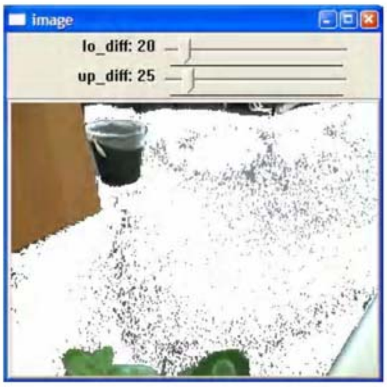
4 Implementation

A single Logitech Quickcam Pro4000, which has the highest performance among commercial CCD webcams, is used to capture real-time images. The image size is 320x240 pixels and the developed software has dialog-based interfaces to set up and control the system.

Sub-tasks of clutter detection using image processing are floor selection, background subtraction including background update and motion detection. The details of the tasks and actions after recognition are described in this section.

4.1 Floor Selection

To focus on the floor area for detecting clutter, a mask image to indicate the floor is necessary. As a fixed camera is used here, the floor detection is required only one time, when the camera is set up, and the software lets the user select the floor region in the initial background image.



lo_diff: 20

up_diff: 25

Fig. 2 FloodFill

After capturing a background image, a window of the image (Fig.2) pops up. In this window, FloodFill fills neighbour pixels whose values are close to the pixel clicked by the user. The pixel will belong to the repainted domain if its value v meets the following condition;

$$v_0 - d_{lw} \leq v \leq v_0 + d_{up} \qquad (1)$$

where v_0 is the value of one of the pixels in the repainted domain beginning with the clicked pixels [18]. $d_{lw,}$ the maximal lower difference and d_{up}, the maximal upper difference between the pixels, can be defined by the user as in Fig.2, and the user can select the floor area with several clicks.

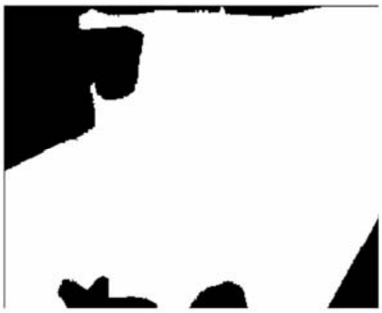


Fig. 3 Floor Mask Image

As the selected area gets lots of tiny chinks, when the user submits the floor-selected image, the system finds the contours of the area and fills them in the mask image like Fig.3.

4.2 Background Subtraction

Background subtraction is a way to find the difference between current images and the background image. In this system, an initial background image with nothing on the floor is captured to consider a noticeable difference from current images as clutter.

Firstly, a simple background model is built up by accumulating several dozens of frames (N) and calculating the mean of each summed pixel value ($bgSum_{(x,y)}$) to get mean brightness [18].

$$bgMean_{(x,y)} = bgSum_{(x,y)}/N \qquad (2)$$

Then, the absolute differences ($diff_{(x,y)}$) between the background model and the current image ($Cur_{(x,y)}$) are calculated after non-floor region is masked in the both images, using the floor mask image built previously.

$$diff_{(x,y)} = abs(bgMean_{(x,y)} - Cur_{(x,y)}) \qquad (3)$$

To get rid of noise, differences smaller than a threshold value are returned to 0, and a binary image is created by returning the others to 255.

$$diff_{(x,y)} = \begin{cases} 255, \text{ if } diff_{(x,y)} > threshold \\ 0, \text{ otherwise} \end{cases} \qquad (4)$$

Whenever this binary image becomes null, the background model is updated to cope with slight lighting changes which are ignored by thresholding. For the dramatic lighting changes by turning on/off a lamp, the background model is also updated when the difference is averagely general.

4.3 Motion Detection

As background subtraction finds both standstill and moving objects, the system also detects the difference between successive frames to only find motion.

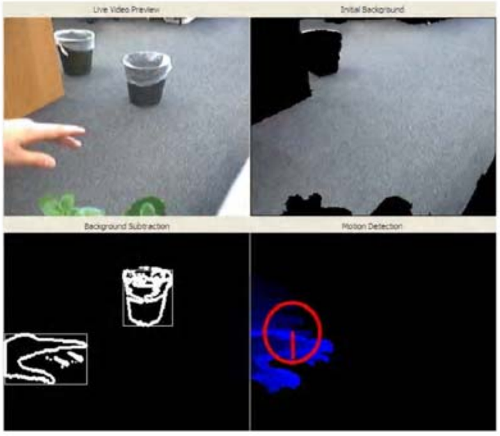


Fig. 4 Standstill vs Moving objects

In Fig.5, a slipper and a hand both appear on the background-subtracted image (left-bottom), but the difference between successive frames shows only the moving hand (right-bottom).

For the cases that there are several objects on the floor at the same time, a bounding box of each noticeable region subtracted from the background image is set as Region of Interest (ROI), and the each ROI is checked if there is any motion inside. If there is any ROI without any motion, the system considers it as still clutter on the floor. And any noticeable motion is considered as a moving object.

4.4 Action

When the system detects any still clutter on the floor, it produces a voice alert, "There is a clutter on the floor. Please move it from the floor". And the voice tells "There is a moving object" for any moving clutter.

5 Result

The floor selection works well no matter how many separate regions correspond to the floor in the background image. If there is more than one individual floor region, the contour of the each region is detected and filled respectively. As the floor is detected only once in the beginning, if any structure in the room moves in the middle of the clutter detection, the floor mask image should be updated manually. The tiny motion of a structure in the environment, however, would not be considered as a clutter because small differences would not be set as ROIs.

As background subtraction and motion detection both compare pixel values, if the colour and texture of the clutter are very similar to the floor, the system is unlikely to consider it as clutter. So it is assumed that there is no clutter with the floor's colour and texture. Also, as the background image is updated only when the dramatic lighting changes are overall in the scene, it is assumed that there is no small lamp, only a ceiling fixture lighting the whole room.

The clutter detection works well unless a ROI is occluded by a moving region. As the bounding boxes are used as ROIs to check motion inside, if a standstill object's bounding box overlaps a moving object too much, the standstill object would be defined as a moving object. Usually the occlusion vanishes soon because one keeps the position and the other is moving. As soon as this happens, the still object is detected and the warning sound is produced.

6 Conclusion

This paper focuses on detecting risk factors of young children's falls in the home environment, to prevent falls. This is different from the previous papers that focus on detecting the actual falls and the target subject is elderly people. The risk factors are determined as environmental and behavioural ones, which are dynamic and require a caregiver's constant supervision. They include any appearance of clutter on the floor, whether the children are running fast, and whether children approach, or climbs, any structure from which they may have suffer an injury if they fall. Based on these, the tasks of the proposed system are identified as clutter detection and tracking of moving children. This paper only presents the clutter detection.

A single commercial camera is used to detect clutter for practical use without any sensor or marker to be attached on body. The clutter detection works well unless too much occlusion happens between a still object's bounding box and a moving region. The action of the system when it finds any clutter is to produce a warning voice alert at present, but it could be modified to send a message to a caregiver's or parents' portable device.

The second task of tracking moving children is being planned to check if they approach or climb furniture by measuring their position, trajectory and velocity. As the task does not require understanding of the children's posture, the method should be robust enough to classify moving regions into human or non-human and obtain accurate tracking data. Also it should be efficient and fast enough to run in real time. As the subject is young children, the classifying and tracking of them could use different features from those proposed in former papers targeting adults.

References:

[1] Child Accident Prevention Trust, Factsheet: Home Accidents, *http://www.capt.org.uk/pdfs/factsheet home accidents.pdf*, United Kingdom, 2004

[2] E. Towner, T. Dowswell, C. Mackereth and S. Jarvis, What Works in Preventing Unintentional Injuries in Children and Young Adolescents? An Updated Systematic Review, *London Health Development Agency*, 2001

[3] D. Colvin, C. Lord, G. Bishop, T. Engel and A. Patra, A Fall Intervention/Mobility Aid System for Elderly and Rehabilitative Populations, *Annual Internaltional Conference of the IEEE Engineering in Medicine and Biology Society*, Vol.13, No.4, 1991, pp. 1936-1937

[4] G. Williams, K. Doughty, K. Cameron and D. Bradley, A Smart Fall & Activity Monitor for Telecare Applications, *20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol.20, No.3, 1998, pp. 1151-1154

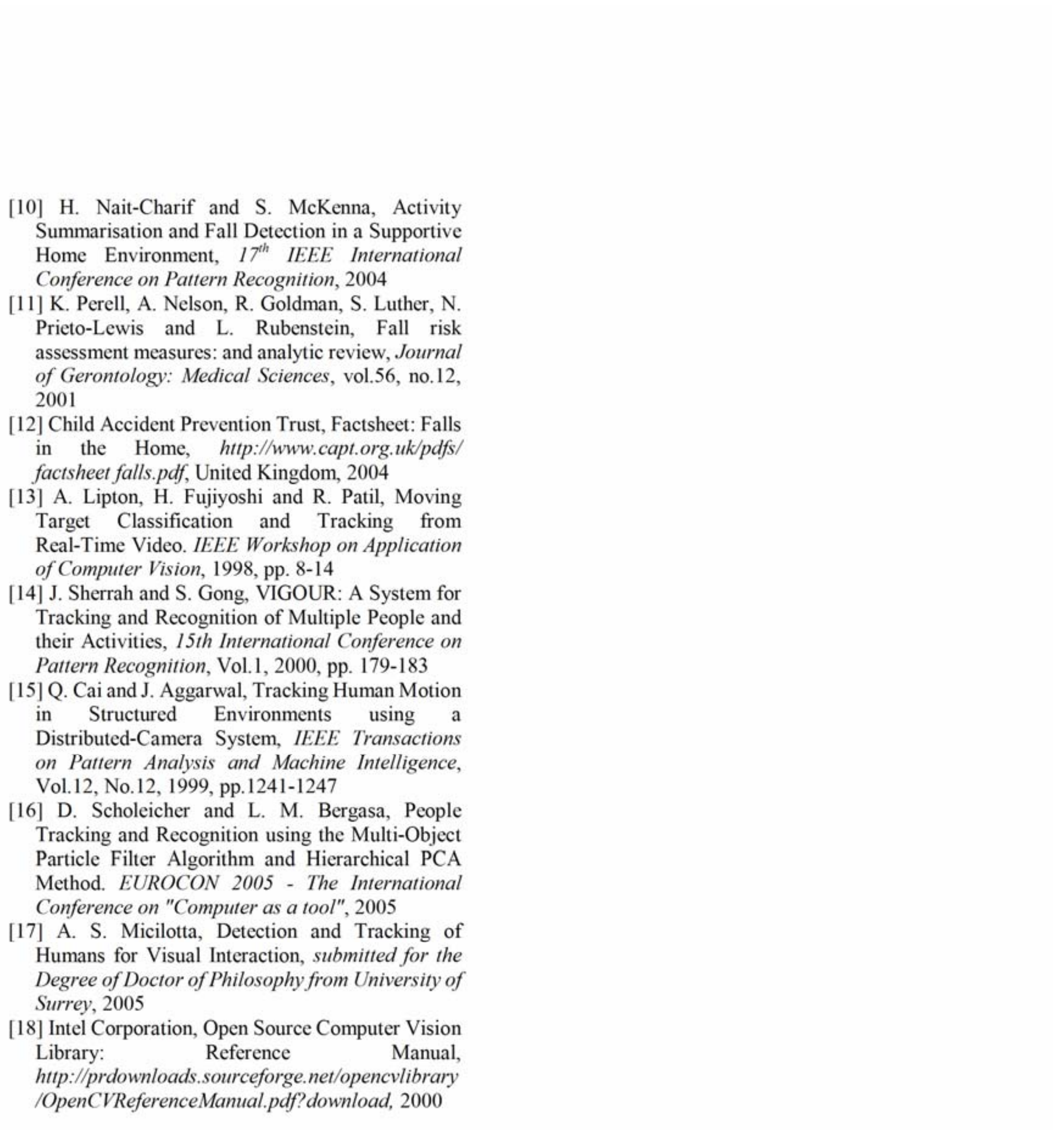
[5] T. Tamura, T. Yoshimura, F. Horiuchi, Y. Higashi and T. Fujimoto, An Ambulatory Fall Monitor for the Elderly, *22nd Annual EMBS International Conference*, 2000, pp. 2608-2610

[6] K. Fukaya, Fall Detection sensor for Fall Protection airbag, *Annual Conference of The Society of Instrument and Control Engineers*, 2002, pp. 419-420

[7] N. Noury, A Smart Sensor for the Remote Follow Up of Activity and Fall Detection of the Elderly, *2nd Annual International IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine & Biology*, 2002, pp. 314-317

[8] B. Najafi, K. Aminian, F. Loew, Y. Blanc and P. Robert, Measurement of Stand-Sit and Sit-Stand Transitions using a Miniature Gyroscope and its Application in Fall Risk Evaluation in the Elderly, IEEE Transactions on Biomedical Engineering, Vol.49, No.8, 2002, pp. 843-851

[9] A. Sixsmith and N. Johnson, A Smart Sensor to Detect the Falls of the Elderly, *IEEE Pervasive Computing*, Vol.3, No.2, 2004, pp. 42-47



[10] H. Nait-Charif and S. McKenna, Activity Summarisation and Fall Detection in a Supportive Home Environment, *17th IEEE International Conference on Pattern Recognition*, 2004

[11] K. Perell, A. Nelson, R. Goldman, S. Luther, N. Prieto-Lewis and L. Rubenstein, Fall risk assessment measures: and analytic review, *Journal of Gerontology: Medical Sciences*, vol.56, no.12, 2001

[12] Child Accident Prevention Trust, Factsheet: Falls in the Home, *http://www.capt.org.uk/pdfs/ factsheet falls.pdf*, United Kingdom, 2004

[13] A. Lipton, H. Fujiyoshi and R. Patil, Moving Target Classification and Tracking from Real-Time Video. *IEEE Workshop on Application of Computer Vision*, 1998, pp. 8-14

[14] J. Sherrah and S. Gong, VIGOUR: A System for Tracking and Recognition of Multiple People and their Activities, *15th International Conference on Pattern Recognition*, Vol.1, 2000, pp. 179-183

[15] Q. Cai and J. Aggarwal, Tracking Human Motion in Structured Environments using a Distributed-Camera System, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.12, No.12, 1999, pp.1241-1247

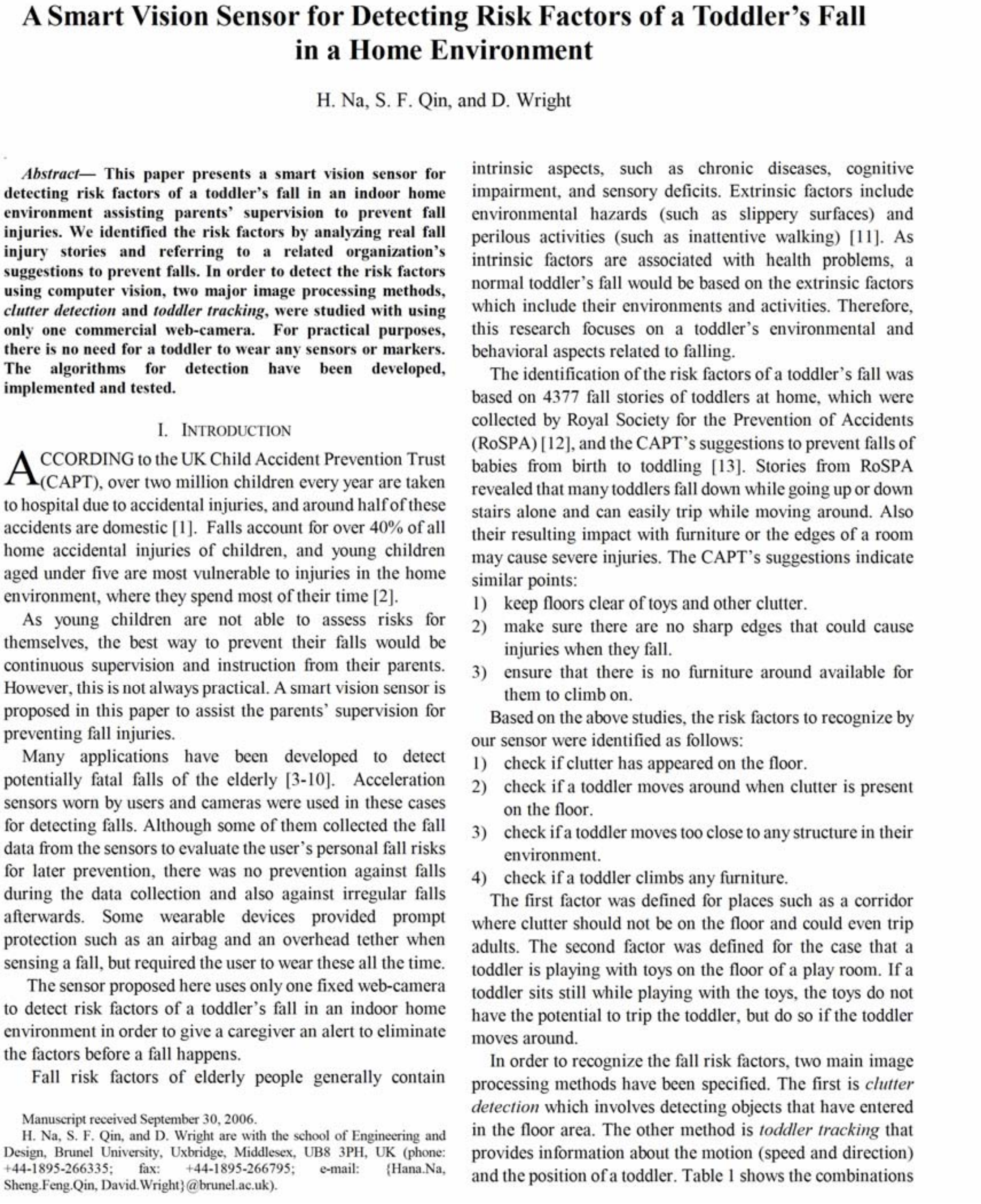
[16] D. Scholeicher and L. M. Bergasa, People Tracking and Recognition using the Multi-Object Particle Filter Algorithm and Hierarchical PCA Method. *EUROCON 2005 - The International Conference on "Computer as a tool"*, 2005

[17] A. S. Micilotta, Detection and Tracking of Humans for Visual Interaction, *submitted for the Degree of Doctor of Philosophy from University of Surrey*, 2005

[18] Intel Corporation, Open Source Computer Vision Library: Reference Manual, *http://prdownloads.sourceforge.net/opencvlibrary /OpenCVReferenceManual.pdf?download,* 2000

Appendix F

IEEE ICNSC ’07



A Smart Vision Sensor for Detecting Risk Factors of a Toddler’s Fall in a Home Environment

H. Na, S. F. Qin, and D. Wright

***Abstract*— This paper presents a smart vision sensor for detecting risk factors of a toddler’s fall in an indoor home environment assisting parents’ supervision to prevent fall injuries. We identified the risk factors by analyzing real fall injury stories and referring to a related organization’s suggestions to prevent falls. In order to detect the risk factors using computer vision, two major image processing methods, *clutter detection* and *toddler tracking*, were studied with using only one commercial web-camera. For practical purposes, there is no need for a toddler to wear any sensors or markers. The algorithms for detection have been developed, implemented and tested.**

I. Introduction

ACCORDING to the UK Child Accident Prevention Trust (CAPT), over two million children every year are taken to hospital due to accidental injuries, and around half of these accidents are domestic [1]. Falls account for over 40% of all home accidental injuries of children, and young children aged under five are most vulnerable to injuries in the home environment, where they spend most of their time [2].

As young children are not able to assess risks for themselves, the best way to prevent their falls would be continuous supervision and instruction from their parents. However, this is not always practical. A smart vision sensor is proposed in this paper to assist the parents’ supervision for preventing fall injuries.

Many applications have been developed to detect potentially fatal falls of the elderly [3-10]. Acceleration sensors worn by users and cameras were used in these cases for detecting falls. Although some of them collected the fall data from the sensors to evaluate the user’s personal fall risks for later prevention, there was no prevention against falls during the data collection and also against irregular falls afterwards. Some wearable devices provided prompt protection such as an airbag and an overhead tether when sensing a fall, but required the user to wear these all the time.

The sensor proposed here uses only one fixed web-camera to detect risk factors of a toddler’s fall in an indoor home environment in order to give a caregiver an alert to eliminate the factors before a fall happens.

Fall risk factors of elderly people generally contain intrinsic aspects, such as chronic diseases, cognitive impairment, and sensory deficits. Extrinsic factors include environmental hazards (such as slippery surfaces) and perilous activities (such as inattentive walking) [11]. As intrinsic factors are associated with health problems, a normal toddler’s fall would be based on the extrinsic factors which include their environments and activities. Therefore, this research focuses on a toddler’s environmental and behavioral aspects related to falling.

The identification of the risk factors of a toddler’s fall was based on 4377 fall stories of toddlers at home, which were collected by Royal Society for the Prevention of Accidents (RoSPA) [12], and the CAPT’s suggestions to prevent falls of babies from birth to toddling [13]. Stories from RoSPA revealed that many toddlers fall down while going up or down stairs alone and can easily trip while moving around. Also their resulting impact with furniture or the edges of a room may cause severe injuries. The CAPT’s suggestions indicate similar points:

1) keep floors clear of toys and other clutter.
2) make sure there are no sharp edges that could cause injuries when they fall.
3) ensure that there is no furniture around available for them to climb on.

Based on the above studies, the risk factors to recognize by our sensor were identified as follows:

1) check if clutter has appeared on the floor.
2) check if a toddler moves around when clutter is present on the floor.
3) check if a toddler moves too close to any structure in their environment.
4) check if a toddler climbs any furniture.

The first factor was defined for places such as a corridor where clutter should not be on the floor and could even trip adults. The second factor was defined for the case that a toddler is playing with toys on the floor of a play room. If a toddler sits still while playing with the toys, the toys do not have the potential to trip the toddler, but do so if the toddler moves around.

In order to recognize the fall risk factors, two main image processing methods have been specified. The first is *clutter detection* which involves detecting objects that have entered in the floor area. The other method is *toddler tracking* that provides information about the motion (speed and direction) and the position of a toddler. Table 1 shows the combinations

Manuscript received September 30, 2006.

H. Na, S. F. Qin, and D. Wright are with the school of Engineering and Design, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK (phone: +44-1895-266335; fax: +44-1895-266795; e-mail: {Hana.Na, Sheng.Feng.Qin, David.Wright}@brunel.ac.uk).

of information obtained by the methods to detect each risk factor.

TABLE I
IMAGE PROCESSING METHODS USED TO RECOGNIZE RISK FACTORS

Image Processing / Fall Risk Factors	Clutter Detection	Toddler Tracking (Motion)	Toddler Tracking (Position)
Clutter on the Floor	■		
Moving Toddler + Clutter on the Floor	■	■	
Toddler Moving near Furniture		■	■
Climbing Toddler		■	■

Clutter detection detects the appearance of clutter on the floor. The speed and direction information obtained by *tracking a toddler* is used to measure his or her movement. The position information against the floor area judges if a toddler moves near furniture or climbs a structure.

II. RELATED WORK

The proposed vision sensor should watch both small objects and the toddler to detect the appearance of clutter on the floor and the toddler's fall-prone behaviors. Therefore, the biggest problem is to distinguish between human and non-human artifacts. This section gives a brief overview of existing studies that differentiate between human beings and clutter and track them individually based on various cues.

Lipton et al. [14] detected moving targets by using the pixel wise difference between consecutive image frames. They then classified them into human, vehicle, or background clutter, based on the target size and shape dispersedness as people are smaller than vehicles and have more complex shapes. This method was relatively simple and sufficient for real-time motion analysis, but it seemed only adequate enough to distinguish people from big vehicles and the tiny motion of trees.

VIGOUR of Sherrah and Gong [15] found skin color clusters and tracked three boxes that bounded a head and two hands respectively for one person. The head box tracker was initialized using Support Vector Machine face detection and the hand box trackers were initialized heuristically with respect to the head position for tracking of multiple people. VIGOUR also used a simple method with a color cue, but the subjects had to be initially facing the camera and faces could not be occluded.

The single view tracking by Cai and Aggarwal [16] was composed of background subtraction, human segmentation, and human feature correspondence between adjacent frames. After background subtraction, human and non-human moving regions were distinguished using moment invariants based on Principal Component Analysis (PCA). Location, intensity, and geometric information of people were extracted for the tracking. The use of the three features to track a human body achieved much better tracking results than the use of any individual feature, but occlusion was a major obstacle.

After background subtraction, Schleicher et al. [17] used a Particle Filter (PF) algorithm to identify and track any moving objects individually. They applied PCA to each object in order to classify it into person or non-person categories by using geometrical constraints of several body parts. This system was relatively reliable at overcoming occlusion but long-term occlusions and lateral views of people still caused some problems.

Micilotta [18] also used PF to track each human body after fitting a torso primitive to human foreground regions and segmenting skin tone regions for the face and hands. Meanwhile, he presented a more robust method of tracking a human body. Body part detectors trained by AdaBoost detected several body parts by using skin color cues to reduce false detections, and RANSAC assembled the parts into body configurations.

The more cues that are used to detect and track human beings, the more accurate the results would be. However, the use of many cues or complex methods would require expensive computation and may be too time-consuming for real-time applications. The above studies used diverse cues to differentiate between a human and a non-human object, but all of the cues were related to human appearance and were therefore not very reliable at occlusions. In this research, we use motion cues to classify a human body.

III. SYSTEM OVERVIEW

A. Clutter Detection

The objective of *clutter detection* is to find stationary objects on the floor which may constitute a potential trip hazard. In order to achieve this, background subtraction, background image update, and motion detection are used. Its workflow is presented in Fig.1.

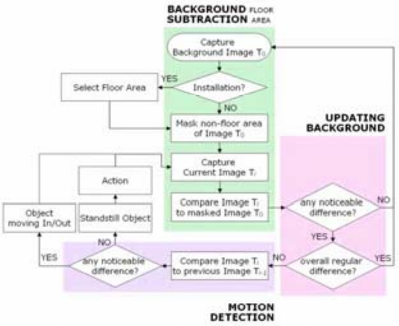

BACKGROUND SUBTRACTION
FLOOR AREA
Capture Background Image T0
Installation?
YES
Select Floor Area
NO
Mask non-floor area of Image T0
UPDATING BACKGROUND
Capture Current Image Ti
Action
Compare Image Ti to masked Image T0
any noticeable difference?
NO
Object moving In/Out
Standstill Object
YES
NO
YES
any noticeable difference?
Compare Image Ti to previous Image Ti-1
NO
overall regular difference?
YES
MOTION DETECTION

Fig. 1. Clutter detection consists of background subtraction, background image update, and motion detection.

Background subtraction detects anything on the floor by comparing current images with an initial background image that contains a clear floor. To focus on the floor area, the non-floor area is masked in the background image. The background image needs to be constantly refreshed due to extraneous changes such as illumination variance and swaying tree branches. As this sensor targets indoor home environments, only domestic lighting changes are dealt with.

As background subtraction detects anything which does not belong to the background image, it is not sufficient to distinguish stationary objects from moving ones. Therefore every two consecutive images are compared in order to classify stationary and moving regions.

B. Toddler Tracking

Tracking a toddler requires finding and matching identical toddler regions in consecutive frames in order to obtain the toddler's real-time motion and position information. After background subtraction, identical regions over frames are simply connected by detecting the closest region centers between frames. The biggest problem here is the discrimination of human and non-human artifacts to detect the toddler against clutter.

This sensor uses different moving characteristics of human and non-human objects to distinguish the human from the non-human artifacts. As the defined toddler's potential fall behaviors related to moving the whole body, the toddler's movement to be detected is assumed to have irregular motion vectors within, due to different motions of body parts when the whole body is mobile. Conversely, objects in an indoor home environment would have relatively constant motion vectors. Hence, a toddler is detected by calculating the similarity of the motion vectors in each region that is subtracted from the background image.

The necessary position information applies if a toddler is moving near or climbing furniture or a structure. Hence, the floor region identified from *clutter detection* is used to determine if the toddler region is near the boundary of the floor area or off the area with the assumption that the non-floor area is filled with furniture and the room structure.

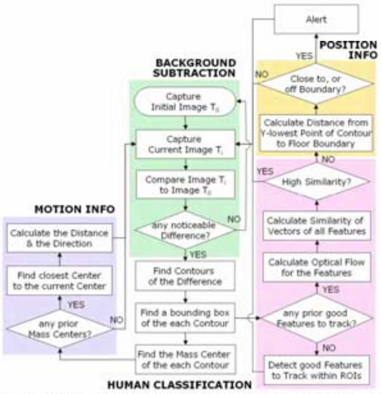

Alert
POSITION INFO
YES
NO
Close to, or off Boundary?
BACKGROUND SUBTRACTION
Capture Initial Image T0
Calculate Distance from Y-lowest Point of Contour to Floor Boundary
Capture Current Image Ti
NO
YES
Compare Image Ti to Image T0
High Similarity?
MOTION INFO
any noticeable Difference?
NO
Calculate the Distance & the Direction
Calculate Similarity of Vectors of all Features
YES
Find Contours of the Difference
Calculate Optical Flow for the Features
Find closest Center to the current Center
YES
YES
any prior Mass Centers?
NO
Find a bounding box of the each Contour
any prior good Features to track?
NO
Find the Mass Center of the each Contour
Detect good Features to Track within ROIs
HUMAN CLASSIFICATION

Fig. 2. *Toddler tracking* involves finding a toddler, matching identical regions in successive frames, and calculating speed, direction, and position of the toddler.

IV. Implementation

A single Logitech Quickcam Pro 5000 was used to capture real-time images. The image size is 640x480 pixels and the developed software has dialog-based interfaces for users to set up and control the sensor.

A. Clutter Detection

The sub-tasks of *clutter detection* contain floor selection, background subtraction including background image update, and motion detection.

1) Floor Selection: To focus on the floor area for detecting clutter, a mask image is required to indicate the floor is necessary. As a fixed camera is used, the floor detection is required only once when the camera is set up. The software lets the user select the floor region in the initial background image using the FloodFill method.

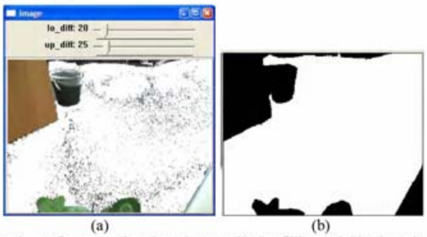


Fig. 3. A floor area is selected manually by filling neighboring pixels whose value is similar to the pixel clicked by the user.

The FloodFill method fills neighboring pixels whose values are close to the pixel clicked by the user. The pixels will belong to the repainted domain if their value v meets the following condition;

$$v_0 - \delta_{lw} \leq v \leq v_0 + \delta_{up,} \tag{1}$$

where v_0 is the value of one of the pixels in the repainted domain that begins with the selected pixels [19]. $\delta_{lw,}$ the maximal lower difference and δ_{up}, the maximal upper difference between the pixels can be defined by the user with the sliding bar controls in Fig. 3a. In this way the user can select the floor area with several clicks. As the selected area gathers lots of tiny chinks, when the user submits the floor-selected image, a mask image is returned with filled contours of the selected area on it (Fig. 3b).

2) Background Subtraction: this finds the difference between the current image and the background image. Firstly, a simple background model is built up when the floor area is clear by accumulating several frames (N) and calculating the mean value of each summed pixel ($bgSum_{(x,y)}$) to get their mean brightness.

$$\begin{aligned} &bgMean_{(x,\,y)} = bgSum_{(x,\,y)}/N \\ &diff_{(x,\,y)} = abs(bgMean_{(x,\,y)} - Cur_{(x,\,y)}) \end{aligned} \tag{2}$$

The absolute difference ($diff_{(x,y)}$) between the background model and the current image ($Cur_{(x,y)}$) is then calculated by pixels after the non-floor region is masked in the both images using the floor mask image built previously.

To eliminate noise, differences smaller than a threshold value are returned to 0, and a binary image is created by returning the others to 255. Whenever this binary image becomes null, the background model is updated to cope with slight changes of sunlight that are ignored by thresholding. For the dramatic lighting changes such as turning on/off a lamp, the background model is also updated when the differences before the thresholding are similar all over the image.

3) Motion Detection: In order to classify standstill objects from the background-subtracted regions, the sensor also detects the difference between successive frames to detect motion.

In Fig. 4, a bin and a hand both appear on the background-subtracted image, but the difference between successive frames only shows the moving hand. For the cases where there are several objects on the floor at the same time, a bounding box for each noticeable background-subtracted region is set as a Region of Interest (ROI), and the ROI is checked if there is any motion inside it. If there is any ROI without any motion, the sensor considers it as still clutter placed on the floor.

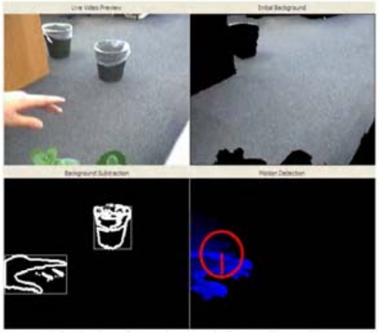


Fig. 4. (anticlockwise from the top-left) a) live video preview, b) background-subtracted regions, c) a moving region, and d) non-floor-marked background image

B. Toddler Tracking

The calculation of the motion and position information of a toddler is done by joining corresponding regions, which are classified as human, in successive frames. A simple background subtraction method is used same as the *clutter detection* and hence its details are omitted here.

1) Obtaining Motion Information: At first, the regions that are background-subtracted from the current image, are focused individually to detect each region's contour and center of mass (x_c, y_c), as calculated in (3).

$$x_c = \sum_x \sum_y xI(x,y) / \sum_x \sum_y I(x,y)$$
$$y_c = \sum_x \sum_y yI(x,y) / \sum_x \sum_y I(x,y) \quad (3)$$

I(x,y) is the pixel intensity value in the position *(x,y)* in the image where each contour is drawn [19]. This center of mass of each contour from one frame is saved to be connected to the center of mass of its corresponding region's contour on the next frame. In case that there is more than one contour detected in an image, distances from the current center of mass to all the centers detected from the previous frame are calculated, and the center is connected to the closest one from the previous frame. The relation of the connected centers provides the speed and direction information.

2) Human Classification: The classification of a human against clutter is based upon the irregular motions inside a human region due to the different motions of body parts. In order to capture the different internal motions, some features which are good to track are detected within the ROI of each background-subtracted region. Such features are actually the corner points which have relatively big eigenvalues in their pixels [19].

The detected features are tracked by calculating the optical flow for each feature using the iterative Lucas-Kanade method [19]. As a result, motion vectors connecting identical features between two adjacent frames can be obtained like in Fig. 5a. Therefore, using the dot product of any two vectors, $\vec{a} \bullet \vec{b} = a_x \times b_x + a_y \times b_y = ab\cos\theta$, the similarity of the motion vectors of all the features in one ROI is calculated over every two consecutive frames.

$$avg(\cos\theta) =$$
$$[\sum_{i=0}^{n-1} \sum_{j=i+1}^{n} \{(a_x^i \times a_x^j) + (a_y^i \times a_y^j)\} / a^i a^j] / \sum_{k=0}^{n} k \quad (4)$$

When all the vectors in one ROI are defined $\vec{a}^0, \vec{a}^1, ..., \vec{a}^n$, the average of cos$\theta$ can be calculated in (4), and the closer to 1 its value is, the closer is the similarity of the vectors. The threshold value to classify human and non-human is defined after tests.

3) *Obtaining Position Information:* As toddlers can barely jump, the vertically lowest point of a human's region contour becomes the focus to observe where a toddler stays on the floor. As the fall risk factors involve if a toddler moves near or climbs furniture, the shortest distance between the point and the boundary of the floor area detected previously is calculated at every frame. If the distance is too short, it is regarded as the toddler moves near furniture, and if the point leaves the area, it is considered as the toddler climbing furniture with the assumption that the off-floor area is filled with furniture and the room structure.

V. Results

A. Clutter Detection

The floor selection works well even when there is more than one separate region corresponding to the floor in the background image. This is because the contour of each region is detected and filled respectively. As the floor is detected only once at the beginning, if any structure in the room moves during the clutter detection, the floor mask image should be updated manually. Any tiny motion of a structure in the environment, however, is not detected as clutter because small differences would not be set as ROIs.

As background subtraction and motion detection both compare pixel intensity values, if the color and texture of the clutter is very similar to the floor, the sensor is unlikely to consider it as clutter. So it is assumed that there is no clutter corresponding with the floor color and texture. Also, as the background image is only updated when a significant lighting change is introduced to the scene, we assume that there is no spot light but a ceiling fixture that lights the whole room.

Detecting clutter works well unless a stationary object's ROI is obscured by a moving region. As each region bounding box is checked if it contains any motion, a standstill object's ROI could overlap a moving object even while the moving object actually does not occlude the still one. The occlusion vanishes soon thereafter because one object maintains its position while the other is mobile.

B. Toddler Tracking

Relating the centers of mass of regions over two successive frames works adequately as long as the background subtraction functions properly. Sometimes, one person's region can split into two or three connected components on the next frame, causing incorrect connections of centers of mass, and consequently generates the wrong speed and direction information.

The calculation of the similarity of all the motion vectors within a moving human region works adequately because any features that are tracked in the next frame but do not belong to the next ROI are discarded. Also when there are multiple ROIs in one image, each ROI is checked to remove features that have been tracked incorrectly, and new features are detected whenever there is any ROI containing less than two features inside.

Several tests have been carried out to identify the threshold value to classify a human with an adult, a ball, and a radio controlled car. As it was revealed from Hamleys, one of the largest toy shops in the world, that other toys that move more dynamically are for children older than toddlers, they were not used in these tests.

The averaged value of cosθ between every two vectors within one moving region over every two adjacent frames is fairly dynamic for a walking human (Fig. 5b) and is constantly close to 1 for a thrown ball (Fig. 6b) and a radio controlled model car moving forwards and backwards (Fig. 7b). As the cosθ value sometimes gets considerably close to 1 for a human motion and somewhat lower than 1 for an object motion, the average of the cosθ values from the past frames is calculated. Based on the tests, we found that the average of every past frame's cosθ value stays under 0.75 for a walking human (Fig. 5c) and over 0.9 for a rolling ball (Fig. 6c) and a moving model car (Fig. 7c). Therefore the threshold to classify a human and a non-human is defined as 0.8.

As the necessary position information of a toddler is his or her relative position to the floor, this is easily acquired by calculating the shortest distance between the lowest point of the toddler's contour and the floor's contour.

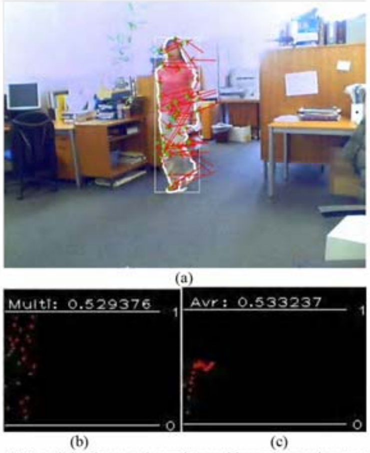


Fig. 5. (a) A walking human shows internal irregular motion vectors. (b) The graph shows the average of cosθ between every two vectors from the each frame during the human walk and it is also irregular. c) The graph presents the average of all the past frames' cosθ values.

VI. Conclusion

This research focuses on detecting risk factors of a toddler's fall in an indoor home environment in order to prevent falling injuries. This is different from the studies conducted previously, which focused on detecting the actual falls and were specifically tailored towards elderly people. In our research, the risk factors are determined as environmental and behavioral ones, which are dynamic and would require a caregiver's constant supervision. These detections include the identification of any clutter on the floor and whether a toddler approaches or climbs furniture. Two main image-processing methods are studied to detect these factors, namely *clutter detection* and *toddler tracking*.

A single commercial camera is used without having sensors or markers attached on the body for practicality. The *clutter detection* works adequately unless there is too much overlap between a still object's bounding box and a moving region. The *Toddler tracking* method shows that the novel concept of human classification using irregular motions of different body parts, works effectively. Based on several tests, the threshold of the average cosθ value is identified as 0.8 to differentiate a human from a non-human objects. In order to reinforce this discrimination even from adults and pets whose motions are also dynamic, other cues related to toddlers will be used, such as size and motion history.

The correct motion and position information of each foreground region can be obtained provided that the background subtraction functions and a human region does not split into multiple regions. If this smart sensing technique is further developed to cope with occlusion of multiple toddlers as well as the regional splits of a single object in the future, it could be extended to be used in other environments, such as a nursery school.

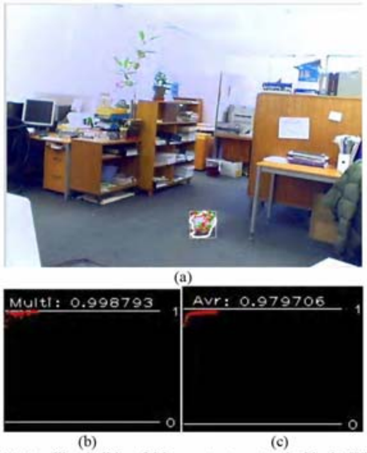


Fig. 6. (a) A rolling ball has fairly constant vectors inside. b) While the ball is thrown and rolls, the averaged cosθ from each frame is close to 1. c) The average of every frame's cosθ value is also very close to 1.

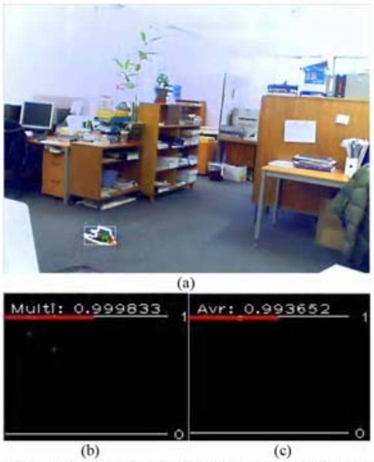


Fig. 7. (a) A radio controlled model car has almost parallel vectors. b) While the car is moving forwards and backwards, the averaged cosθ from each frame is fairly close to 1. c) The average of every frame's cosθ value is also considerably close to 1.

References

[1] Child Accident Prevention Trust. (2004). Factsheet: Home Accidents. Available: http://www.capt.org.uk/pdfs/factsheet home accidents.pdf

[2] E. Towner, T. Dowswell, C. Mackereth, and S. Jarvis, *What Works in Preventing Unintentional Injuries in Children and Young Adolescents? An Updated Systematic Review*, London Health Development Agency, 2001

[3] D. Colvin, C. Lord, G. Bishop, T. Engel, and A. Patra, "A Fall Intervention/Mobility Aid System for Elderly and Rehabilitative Populations," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 13, no. 4, 1991, pp. 1936-1937

[4] G. Williams, K. Doughty, K. Cameron, and D. Bradley, "A Smart Fall & Activity Monitor for Telecare Applications," *20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 20, no. 3, 1998, pp. 1151-1154

[5] T. Tamura, T. Yoshimura, F. Horiuchi, Y. Higashi, and T. Fujimoto, "An Ambulatory Fall Monitor for the Elderly," *22nd Annual EMBS International Conference*, 2000, pp. 2608-2610

[6] K. Fukaya, "Fall Detection sensor for Fall Protection airbag," *Annual Conference of The Society of Instrument and Control Engineers*, 2002, pp. 419-420

[7] N. Noury, "A Smart Sensor for the Remote Follow Up of Activity and Fall Detection of the Elderly," *2nd Annual International IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine & Biology*, 2002, pp. 314-317

[8] B. Najafi, K. Aminian, F. Loew, Y. Blanc, and P. Robert, "Measurement of Stand-Sit and Sit-Stand Transitions using a Miniature Gyroscope and its Application in Fall Risk Evaluation in the Elderly," *IEEE Transactions on Biomedical Engineering*, vol.49, no.8, 2002, pp. 843-851

[9] A. Sixsmith and N. Johnson, "A Smart Sensor to Detect the Falls of the Elderly," *IEEE Pervasive Computing*, vol. 3, no. 2, 2004, pp. 42-47

[10] H. Nait-Charif and S. McKenna, "Activity Summarization and Fall Detection in a Supportive Home Environment," *17th IEEE International Conference on Pattern Recognition*, 2004

[11] K. Perell, A. Nelson, R. Goldman, S. Luther, N. Prieto-Lewis, and L. Rubenstein, "Fall risk assessment measures: and analytic review," *Journal of Gerontology: Medical Sciences*, vol. 56, no. 12, 2001

[12] Child Accident Prevention Trust. (2004). Factsheet: Falls in the Home. Available: http://www.capt.org.uk/pdfs/factsheet falls.pdf

[13] Royal Society for the Prevention of Accidents. (2000-2002). Home and Leisure Accident Surveillance System - Annual Reports. Available: http://www.hassandlass.org.uk/query/reports.htm (need to contact the Information Center for details of individual accidents)

[14] A. Lipton, H. Fujiyoshi, and R. Patil, "Moving Target Classification and Tracking from Real-Time Video," *IEEE Workshop on Application of Computer Vision*, 1998, pp. 8-14

[15] J. Sherrah and S. Gong, "VIGOUR: A System for Tracking and Recognition of Multiple People and their Activities," *15th International Conference on Pattern Recognition*, vol. 1, 2000, pp. 179-183

[16] Q. Cai and J. Aggarwal, Tracking Human Motion in Structured Environments using a Distributed-Camera System, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.12, No.12, 1999, pp.1241-1247

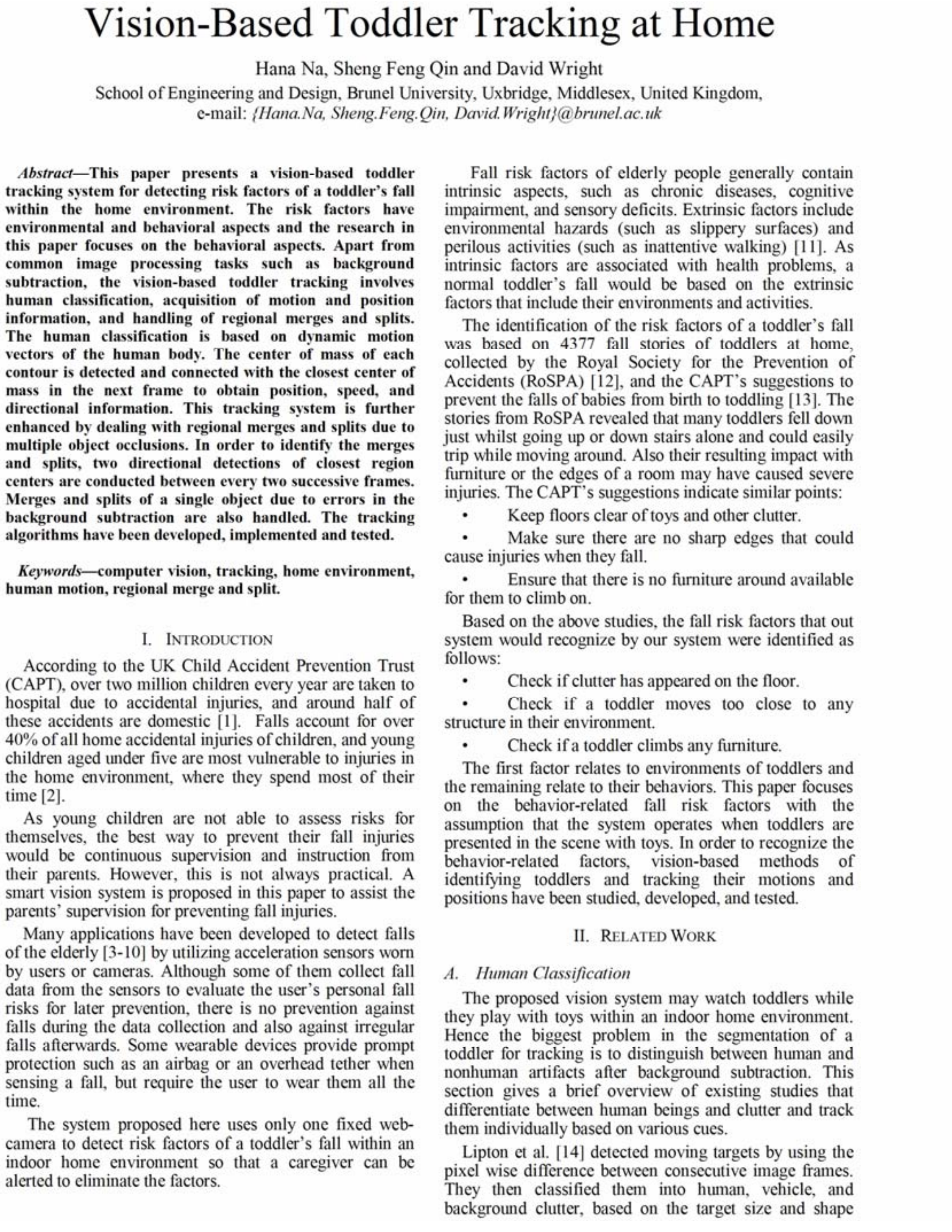
[17] D. Scholeicher and L. M. Bergasa, "People Tracking and Recognition using the Multi-Object Particle Filter Algorithm and Hierarchical PCA Method", *EUROCON 2005 - The International Conference on "Computer as a tool"*, 2005

[18] A. S. Micilotta, "Detection and Tracking of Humans for Visual Interaction," Ph.D. dissertation, School of Electronics and Physical Science, University of Surrey, Surrey, United Kingdom, 2005

[19] Intel Corporation. (2000). Open Source Computer Vision Library: Reference Manual. Available: http://switch.dl.sourceforge.net/sourceforge/opencvlibrary/OpenCVReferenceManual.pdf

Appendix G

IEEE EUROCON ’07



Vision-Based Toddler Tracking at Home

Hana Na, Sheng Feng Qin and David Wright
School of Engineering and Design, Brunel University, Uxbridge, Middlesex, United Kingdom,
e-mail: *{Hana.Na, Sheng.Feng.Qin, David.Wright}@brunel.ac.uk*

***Abstract*—This paper presents a vision-based toddler tracking system for detecting risk factors of a toddler's fall within the home environment. The risk factors have environmental and behavioral aspects and the research in this paper focuses on the behavioral aspects. Apart from common image processing tasks such as background subtraction, the vision-based toddler tracking involves human classification, acquisition of motion and position information, and handling of regional merges and splits. The human classification is based on dynamic motion vectors of the human body. The center of mass of each contour is detected and connected with the closest center of mass in the next frame to obtain position, speed, and directional information. This tracking system is further enhanced by dealing with regional merges and splits due to multiple object occlusions. In order to identify the merges and splits, two directional detections of closest region centers are conducted between every two successive frames. Merges and splits of a single object due to errors in the background subtraction are also handled. The tracking algorithms have been developed, implemented and tested.**

***Keywords*—computer vision, tracking, home environment, human motion, regional merge and split.**

I. Introduction

According to the UK Child Accident Prevention Trust (CAPT), over two million children every year are taken to hospital due to accidental injuries, and around half of these accidents are domestic [1]. Falls account for over 40% of all home accidental injuries of children, and young children aged under five are most vulnerable to injuries in the home environment, where they spend most of their time [2].

As young children are not able to assess risks for themselves, the best way to prevent their fall injuries would be continuous supervision and instruction from their parents. However, this is not always practical. A smart vision system is proposed in this paper to assist the parents' supervision for preventing fall injuries.

Many applications have been developed to detect falls of the elderly [3-10] by utilizing acceleration sensors worn by users or cameras. Although some of them collect fall data from the sensors to evaluate the user's personal fall risks for later prevention, there is no prevention against falls during the data collection and also against irregular falls afterwards. Some wearable devices provide prompt protection such as an airbag or an overhead tether when sensing a fall, but require the user to wear them all the time.

The system proposed here uses only one fixed web-camera to detect risk factors of a toddler's fall within an indoor home environment so that a caregiver can be alerted to eliminate the factors.

Fall risk factors of elderly people generally contain intrinsic aspects, such as chronic diseases, cognitive impairment, and sensory deficits. Extrinsic factors include environmental hazards (such as slippery surfaces) and perilous activities (such as inattentive walking) [11]. As intrinsic factors are associated with health problems, a normal toddler's fall would be based on the extrinsic factors that include their environments and activities.

The identification of the risk factors of a toddler's fall was based on 4377 fall stories of toddlers at home, collected by the Royal Society for the Prevention of Accidents (RoSPA) [12], and the CAPT's suggestions to prevent the falls of babies from birth to toddling [13]. The stories from RoSPA revealed that many toddlers fell down just whilst going up or down stairs alone and could easily trip while moving around. Also their resulting impact with furniture or the edges of a room may have caused severe injuries. The CAPT's suggestions indicate similar points:

- Keep floors clear of toys and other clutter.
- Make sure there are no sharp edges that could cause injuries when they fall.
- Ensure that there is no furniture around available for them to climb on.

Based on the above studies, the fall risk factors that out system would recognize by our system were identified as follows:

- Check if clutter has appeared on the floor.
- Check if a toddler moves too close to any structure in their environment.
- Check if a toddler climbs any furniture.

The first factor relates to environments of toddlers and the remaining relate to their behaviors. This paper focuses on the behavior-related fall risk factors with the assumption that the system operates when toddlers are presented in the scene with toys. In order to recognize the behavior-related factors, vision-based methods of identifying toddlers and tracking their motions and positions have been studied, developed, and tested.

II. Related Work

A. Human Classification

The proposed vision system may watch toddlers while they play with toys within an indoor home environment. Hence the biggest problem in the segmentation of a toddler for tracking is to distinguish between human and nonhuman artifacts after background subtraction. This section gives a brief overview of existing studies that differentiate between human beings and clutter and track them individually based on various cues.

Lipton et al. [14] detected moving targets by using the pixel wise difference between consecutive image frames. They then classified them into human, vehicle, and background clutter, based on the target size and shape dispersedness as humans are smaller than vehicles and have more complex shapes. This method was relatively simple and sufficient for real-time motion analysis but performed adequate enough only to distinguish humans from big vehicles and the tiny motion of trees.

VIGOUR of Sherrah and Gong [15] tracked a head and two hands of one person by seeking skin color clusters and by utilizing a Support Vector Machine face detector and human body structural information between the head and two hands. VIGOUR required that the subjects had to be initially facing the camera and the faces could not be occluded.

The single view tracking by Cai and Aggarwal [16] was composed of background subtraction, human segmentation, and the human feature correspondence between adjacent frames. After background subtraction, human and nonhuman moving regions were distinguished using moment invariants based on Principal Component Analysis (PCA). Location, intensity, and geometric information of people were extracted for the tracking. The use of the three features to track a human body achieved much better tracking results than the use of any individual feature. However occlusion was a remained major obstacle.

After background subtraction, Schleicher et al. [17] used a Particle Filter (PF) algorithm to identify and individually track any moving objects. They applied PCA to each object in order to classify it into a person or nonperson category by using geometrical constraints of several body parts. This system was relatively reliable at overcoming occlusion but long-term occlusions and lateral views of people still caused some problems.

Micilotta [18] also used PF to track each human body after fitting a torso primitive to human foreground regions and segmenting skin tone regions for the face and hands. Meanwhile, he presented a more robust method of tracking a human. Body part detectors trained by AdaBoost, detected several body parts by using skin color cues to reduce false detections, and RANSAC assembled the parts into body configurations.

The more cues that are used to detect and track human beings, the more accurate the results would become. However, the use of many cues or complex methods would require expensive computation and may be too time-consuming for real-time applications. The above studies used diverse cues to differentiate between a human and a nonhuman object, but all of the cues were related to human appearance and were therefore not very reliable at occlusions. In this research, we use motion cues to classify a human body.

B. Handling of Merges and Splits

In practice, self-occlusion and occlusions between different moving objects or between moving objects and the background are inevitable [19]. Multiple camera systems offer promising methods to reduce ambiguities due to occlusion. Multiple cameras have been used to choose the best view regarding occlusion or to estimate the 3D information of each object for coping with occlusion [20-24]. However, using multiple cameras required complex computation to match identical objects from different cameras or to calibrate the cameras for 3D information.

There were also several studies [25-28] that proposed ideas to tackle the occlusion problems using a single camera by handling regional merges. They dealt with another similar problem that a single object can be split into multiple blobs that yield separate measurements due to errors in background subtraction.

The algorithm of Medioni et al. [25] developed an algorithm that coped with splits by measuring the gray-level similarity between a moving region at one frame and a set of regions at the next frame in its neighborhood, but it did not handle merges of multiple objects.

An approach to handle both merges and splits was to associate prediction based on previous measurements. The method used in [26] for the association was based on virtual measurements that superseded and extended a set of measurements and the set was chosen to optimally fit the set of predicted measurements at each time step. Kumar et al. [27] used Kalman filter based trackers, which predicted and estimated states of objects, so that the predicted shape and position of the objects gave rise to a new synthesized blob when the predicted objects merged. Then, a geometric shape matching algorithm was used to match the predicted blob with the real segmented blob. These association methods worked well as long as the position and motion of target objects were predictable.

McKenna et al. [28] only dealt with regions which belonged to, corresponded to, or included a human being. In order to form a person, multiple regions had to be in close proximity, their projections onto the x-axis had to overlap, and they had to have a total area larger than a threshold. If regions in a group that indicated one or more people grouped together, had not met any of the above conditions, the group would have been split up.

The proposed system does not need to recognize minute postures of toddlers and to identify each toddler and each piece of clutter. It just needs to discriminate any toddler from clutter in an indoor home environment. The clutter may be smaller than a toddler's body and the environment is fairly restricted. Therefore, this research seeks another method to handle visual merges and splits in images from one fixed camera using simple cues rather than associating with prediction.

III. System Overview

The toddler tracking involves background subtraction, human classification, obtaining motion and position information, and handling of merges and splits. The whole workflow is presented in Fig. 1.

Due to the usage of a fixed camera, a simple background subtraction is used to segment both moving and stationary objects. The background image used for the subtraction needs to be constantly refreshed due to extraneous changes such as the swaying branches of trees and illumination variance. As this system targets indoor home environments, only domestic lighting changes are dealt with.

Once all foreground regions are segmented, toddlers need to be separated from clutter which may be toys that they play with. This classification uses different moving characteristics of human and nonhuman objects. As this system starts with capturing a background image that only includes an environment, the system's supervision begins when toddlers and clutter move in the scene. Therefore a toddler's movement to be detected at first is supposed to have irregular internal motion vectors due to the different motions of body parts when the whole body is mobile. Conversely, clutter within an indoor home environment may have relatively constant motion vectors. Hence, toddlers are detected by calculating the similarity of the motion vectors in each region that is subtracted from the background image.

Meanwhile, each foreground region is tracked simply by connecting the closest region centers between consecutive frames, and its speed and direction are calculated with the relation of the connected centers for its motion information.

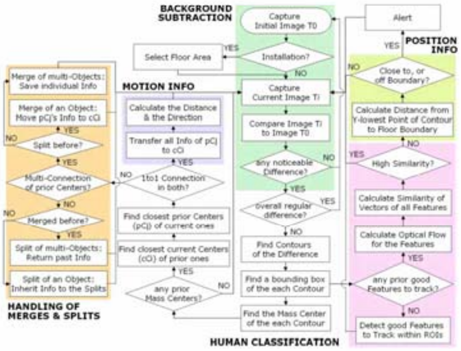

BACKGROUND SUBTRACTION
Capture Initial Image T0
Installation?
YES
Select Floor Area
NO
Capture Current Image Ti
Compare Image Ti to Image T0
any noticeable Difference?
NO
YES
overall regular difference?
YES
NO
Find Contours of the Difference
Find a bounding box of the each Contour
Find the Mass Center of the each Contour
HUMAN CLASSIFICATION
Alert
YES
POSITION INFO
NO
Close to, or off Boundary?
Calculate Distance from Y-lowest Point of Contour to Floor Boundary
NO
YES
High Similarity?
Calculate Similarity of Vectors of all Features
Calculate Optical Flow for the Features
YES
any prior good Features to track?
NO
Detect good Features to Track within ROIs
MOTION INFO
Calculate the Distance & the Direction
Transfer all Info of pCj to cCi
YES
NO
1to1 Connection in both?
Find closest prior Centers (pCj) of current ones
Find closest current Centers (cCi) of prior ones
YES
any prior Mass Centers?
NO
Merge of multi-Objects: Save individual Info
Merge of an Object: Move pCj's Info to cCi
YES
NO
Split before?
YES
Multi-Connection of prior Centers?
NO
NO
Merged before?
YES
Split of multi-Objects: Return past Info
Split of an Object: Inherit Info to the Splits
HANDLING OF MERGES & SPLITS

Fig. 1. System work flow

The necessary position information is if a toddler is moving near or climbing furniture or the room structure. Hence, the floor region is manually selected during installation of the system and used to determine if any toddler region is near the boundary of the floor area or off the area with the assumption that the no floor area is filled with furniture and the room structure.

As this is only a single camera system, regional merges and splits are inevitable. In order to connect identical objects over frames in spite of the merges and splits, closest region centers are detected in two directions between every two consecutive frames. Furthermore, each region's size and its history of merges and splits are used to distinguish between multiple objects and a single object in merges and splits.

IV. Implementation

A single Logitech Quickcam Pro 5000 was used to capture real-time images at the rate of 30 frames per second. The image size is 640x480 pixels and the developed software written in C++ has dialog-based interfaces for users to set up and control the system.

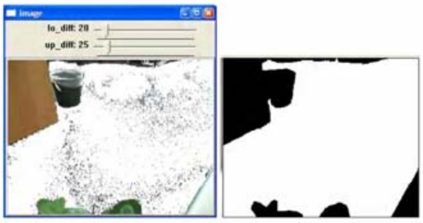
A. Installation (Floor Selection)

A mask image is required to indicate the floor to estimate each toddler's relative position to the floor. As a fixed camera is used, the floor detection is required only once, when the camera is set up. The software lets the user select the floor region in the initial background image using the FloodFill method.

The FloodFill method fills neighboring pixels whose values are close to the pixel clicked by the user. The pixel will belong to the repainted domain if its value *v* meets the following condition:

$$\nu_0 - \delta_{lw} \leq \nu \leq \nu_0 + \delta_{up} , \qquad (1)$$

where v_0 is the value of one of the pixels in the repainted domain that begins with the selected pixels [29]. $\delta_{lw,}$ the maximal lower difference and δ_{up}, the maximal upper difference between the pixels, can be defined by the user with the sliding bar controls in Fig. 2a. In this way the user can select the floor area with several clicks. As the selected area contains lots of tiny chinks (Fig. 2a), when the user submits the floor-selected image, a mask image is returned with filled contours of the selected area on it (Fig. 2b).



(a) (b)

Fig. 2. (a) Selection of the floor area and (b) floor-masked image

B. Background Subtraction

Background subtraction finds the difference between the current image and the background image. Firstly, a simple background model is built up when the floor area is clear by accumulating several frames(*N*) and calculating the mean value of each summed pixel ($bgSum_{(x\ y)}$) to get their mean brightness.

$$\begin{aligned} bgMean_{(x,\ y)} &= bgSum_{(x,\ y)}/N \\ diff_{(x,\ y)} &= abs(bgMean_{(x,\ y)} - Cur_{(x,\ y)}) \end{aligned} \qquad (2)$$

The absolute difference ($diff_{(x,y)}$) between the background model and the current image ($Cur_{(x,y)}$) is then calculated after the nonfloor region is masked in both images, using the floor mask image built previously.

To eliminate noise, differences smaller than a threshold value are set to 0, and a binary image is created by setting the others to 255. Whenever this binary image becomes null, the background model is updated to cope with slight changes of sunlight that are ignored by thresholding. For dramatic lighting changes such as turning on/off a lamp, the background model is also updated when the differences before the thresholding are similar all over the image.

C. Human Classification

The classification of a human against objects is based upon the irregular motions inside a human region due to the different motions of body parts. In order to capture the different internal motions, some features that are good to track, are detected within each Region of Interest (ROI), which is a bounding box of each noticeable background-subtracted region. Such features are actually the corner points that have relatively big eigenvalues in the pixels and have a satisfied distance from one another [29].

The detected features are tracked by calculating the optical flow between every two successive frames for each feature using the iterative Lucas-Kanade method [19]. If any features detected by the optical flow calculation get out of any ROIs found on the same frame, the features are discarded to focus on the ROIs. Also, whenever there are less than five features left within a ROI, the feature detection is executed anew in the ROI to avoid capturing very few motions in one region.

As a result, the relation from one feature's coordinate to its new position detected on the next frame is presented as an arrow indicating a motion vector and one ROI gets multiple motion vectors like in Fig. 5a. Therefore, using the dot product of any two vectors, $\vec{a} \bullet \bar{b} = a_x \times b_x + a_y \times b_y = ab\cos\theta$, the similarity of the motion vectors in one ROI is calculated over two adjacent frames.

$$avg(\cos\theta) = \\ [\sum_{i=0}^{n-1}\sum_{j=i+1}^{n}\{(a_x^i \times a_x^j) + (a_y^i \times a_y^j)\} / a^i a^j] / \sum_{k=0}^{n} k \tag{3}$$

When all the vectors in one ROI are defined $\vec{a}^0, \vec{a}^1, ..., \vec{a}^n$, the average of cos$\theta$ can be calculated in (3). As the vectors are parallel when θ is 0, the closer to 1 the average of cosθ is, the closer the similarity of the vectors. The threshold value to classify human and clutter is defined after tests.

D. Motion and Position Information

At first, the regions that are background-subtracted from each current image are focused individually to detect each region's contour and center of mass (x_c, y_c), as calculated in (4).

$$x_c = \sum_x \sum_y xI(x,y) / \sum_x \sum_y I(x,y) \\ y_c = \sum_x \sum_y yI(x,y) / \sum_x \sum_y I(x,y) \tag{4}$$

I(x,y) is the pixel intensity value in the position *(x,y)* in the image where each contour is drawn [29]. This center of mass coordinate of each region's contour from one frame is saved to be connected to the center of mass of its corresponding region's contour on the next frame. The distances between a center of mass from a frame and all the centers from the previous frame are calculated, and the center is connected to the closest one from the previous frame. This connection is separately conducted on every contour's center detected on each frame. The speed and direction of each contour is calculated for motion information using the coordinates of two connected centers over two consecutive frames.

Whereas the center of mass of a background-subtracted region is used to obtain the motion information, the vertically lowest point of a toddler's region contour becomes the focus here. As a toddler cannot jump, the vertically lowest point of the contour is considered as where the toddler stands on the floor. As this vision system needs to check if a toddler moves near furniture or climbs it, the lowest point is checked for every frame to see if it is close to the boundary of the floor area detected during the system installation or if it gets out of the floor area considering that the no floor area is filled with furniture or the room structure.

E. Handling of Merges and Splits

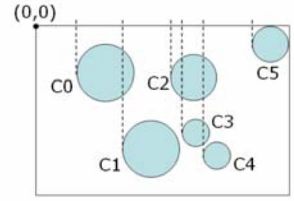
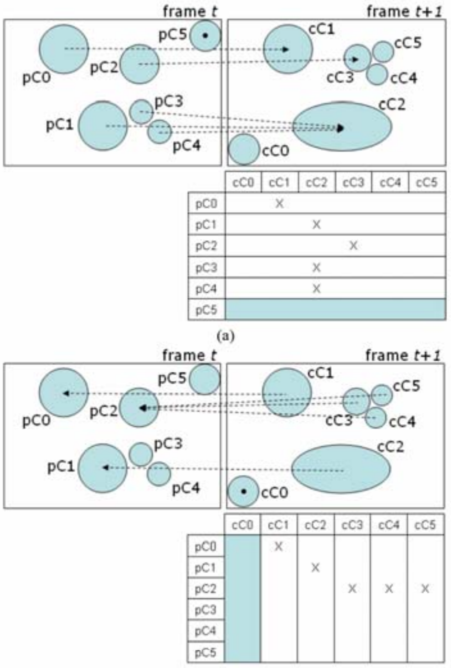


Fig. 3. Contour indexing

Every foreground region on each frame is indexed by the smallest x-coordinate of its contour. Fig. 3 shows an example of the indexing. All the information obtained from a region, such as the coordinates of its center and bounding box, is also tagged with the region's index and kept over every two successive frames to be used in comparing the two frames.



	cC0	cC1	cC2	cC3	cC4	cC5
pC0		X				
pC1			X			
pC2				X		
pC3			X			
pC4			X			
pC5	🟦	🟦	🟦	🟦	🟦	🟦

(a)

	cC0	cC1	cC2	cC3	cC4	cC5
pC0	🟦	X				
pC1	🟦		X			
pC2	🟦			X	X	X
pC3	🟦					
pC4	🟦					
pC5	🟦					

(b)

Fig. 4. (a) Detections of closest centers from the previous frame and (b) from the current frame

As this indexing is conducted anew on every frame, an object can get a different index on the next frame due to regional merges and splits as well as due to simple position changes. In order to connect correct regions for an identical object over two consecutive frames, the closest center detection is carried out from the previous frame to the current frame and vice versa.

For instance, when regions merge, split, and move in and out at the same time as shown in Fig. 4, the closest center of each contour center on the frame *t* is detected on the frame *t+1* (Fig. 4a), and the reverse detection is conducted (Fig. 4b). In order to prevent any wrong connection due to an appearance or a disappearance that does not have any identical region to be connected on the previous frame or the current frame, the distance between the closest centers over two frames is limited within the half length of the diagonal line connecting two opposite points of the bounding box of each region. This is because the moving speed of a toddler and a toy is assumed to be slow enough to catch up within the limitation at the rate of 30 frames per second.

These two different connections are compared to check where a merge, a split, an appearance, a disappearance, or a one-to-one connection happens. The one-to-one connection, which means tracking of a region without any merge or split, is confirmed when the two connections are both singular. For example, when all the regions in the frame *t* are indexed with pC0, pC1, …, pC*n* and the regions in the frame *t+1* are indexed with cC0, cC1, …, cC*n*, only pC0 is connected to cC1 in Fig. 4a and only cC1 is connected to pC0 in Fig. 4b. In this case, all the past speeds and averaged cosθ values tagged with 0 become indexed with 1. The speed and the cosθ value of a region at every frame is the information that should be kept and tagged to correct regions of an identical object during the whole observation to classify and focus on toddlers.

When a region in the current frame has a multiple connection in the closest center detection from the previous frame like cC2, it is considered as a merge, and a region in the other way around like pC2 in Fig. 4 is considered as a split. These merge and split need to be further examined in case that they result from occlusion of multiple objects or from separated blobs of one object due to errors in background subtraction. The differentiation is based on a history of merges and splits for each region as a split should happen at first for a merge to happen in a single object and a merge should come before a split in multiple objects.

When a split happens to a single object, all the split regions inherit the past information tagged to the region before the split. When they merge afterwards, the information of any region before the merge is transferred to the merged region.

When a merge of multiple objects happens, all the past information of each region before the merge is saved individually with the region's size, and the merged region starts with null information unless the multiple objects include a toddler. If a toddler is included there, the toddler's region information is kept on the merged region because for instance, a toddler carrying toys needs to be classified as a toddler and focused upon. Then, when any of the objects becomes separated from the merge, among the pieces of information saved before the merge, a correct piece is returned to the split object by comparing its regional size with the saved regional sizes. The region size allows a ten percent error margin.

A region with no connection in the closest center detection from the previous frame like pC5 is regarded as an object's disappearance and that region's information is removed. A region with no connection in the opposite way like cC0 is regarded as an appearance and begins a new data collection.

V. Results

A. *Installation (Floor Selection)*

The floor selection works well even when there is more than one separate region corresponding to the floor in the background image. This is because the contour of each region is detected and filled respectively. As the floor is detected only once at the beginning, if any structure in the room moves in the middle of the toddler tracking, the floor mask image should be updated manually. Any tiny motion of any structure in the environment, however, is ignored by thresholding.

B. *Background Subtraction*

The simple method of background subtraction works fine with 640x480 images from the QuickCam Pro 5000. Logitech's other web cameras of lower or higher performances such as the QuickCam Pro 4000 or the Ultra Vision are more prone to noise due to low resolution or visible compression artifacts. However, the method occasionally has the problem of splitting one object into multiple blobs mainly for toddlers due to their dynamic posture changes. So as mentioned previously, this split of one object is handled with a split of multiple occluded objects.

As the background image is only updated when a significant lighting change is introduced to the scene, we assumed that there is no spot light, but a ceiling fixture that lights the whole room.

C. *Human Classification*

The calculation of the similarity of all the vectors within a human motion works adequately because any features that are tracked in the next frame but do not belong to the next ROI are discarded. In case of multiple ROIs in one frame, this discarding is conducted in each ROI and new features are detected in any ROI with less than five features.

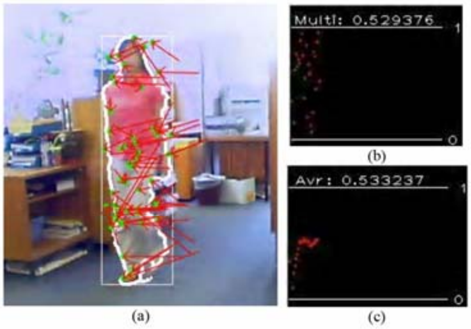


Fig. 5. (a) A walking human region with internal motion vectors (b) its graph showing the average of cosθ between every two vectors from each frame during the human walk and (c) graph presenting the average of all the past frames' cosθ values.

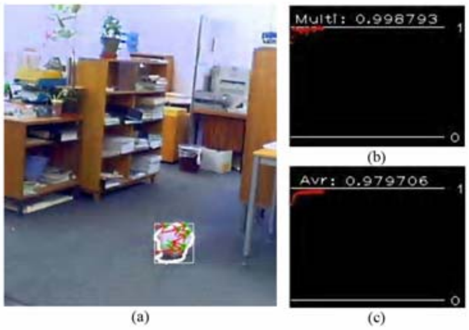


Fig. 6. (a) A rolling ball region and (b) its graph of each frame's cosθ, and (c) graph of the averages of past frames' cosθ values

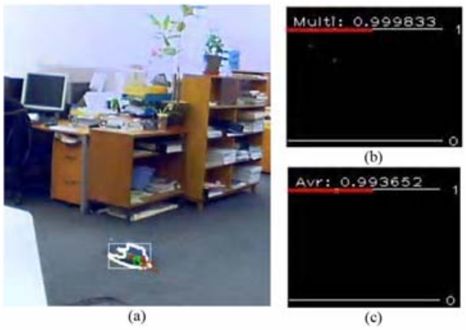


Fig. 7. (a) A region of a radio controlled model car and (b) its graph of each frame's cosθ while the car is moving forwards and backwards and (c) graph of the averages of past frame's cosθ values

Several tests have been carried out to detect the threshold value to classify a human with an adult, a ball, and a radio controlled model car that represent human, rolling, and straight motions respectively. As it was revealed from Hamleys [30], one of the largest toy shops in the world, that other toys that move more dynamically are for children over three years old who are no longer toddlers anymore, they were not used to these tests.

The averaged value of cosθ between every two vectors within one moving region in each frame was fairly dynamic for a walking human (Fig. 5b) and was mostly close to 1 for a rolling ball (Fig. 6b) and a radio controlled model car (Fig. 7b). As sometimes the cosθ value gets considerably close to 1 for human motion and somewhat lower than 1 for object motion, the average of the cosθ values from the past frames is also calculated at every frame. Based on several tests, we found that the average of cosθ value of every past frame stays under 0.75 for a walking human (Fig. 5c) and over 0.9 for a rolling ball (Fig. 6c) and a moving model car (Fig. 7c). Therefore the threshold to classify human and nonhuman is defined as 0.8. Fig. 8 shows a person region bounded by a red box that means it is classified as a human based on its motion vectors.

D. Motion and Position Information

The connection of corresponding regions' centers of masses over two successive frames works well, but incorrect speed and direction information is generated when a regional merge or split occurs. Therefore, the motion information of a region is ignored when it splits or merges with other region.

A toddler's position information, if the toddler moves near or climbs furniture is easily identified by calculating the shortest distance between the vertically lowest point of the toddler's region contour and the contour of the floor area defined during installation. The number underneath the person's bounding box in Fig. 8 indicates the shortest distance from the floor's contour that is drawn in blue.

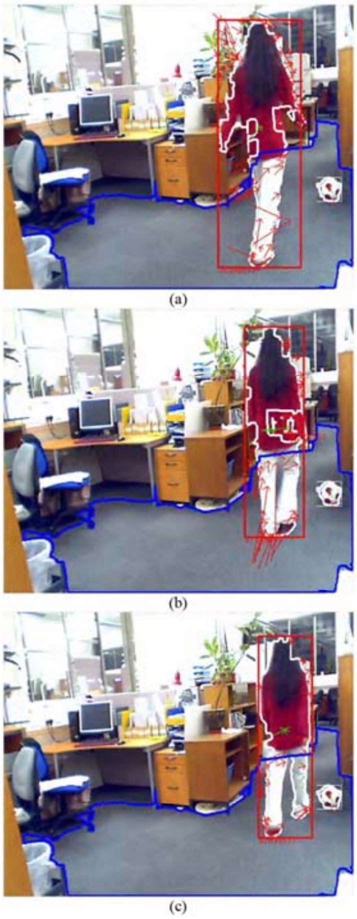


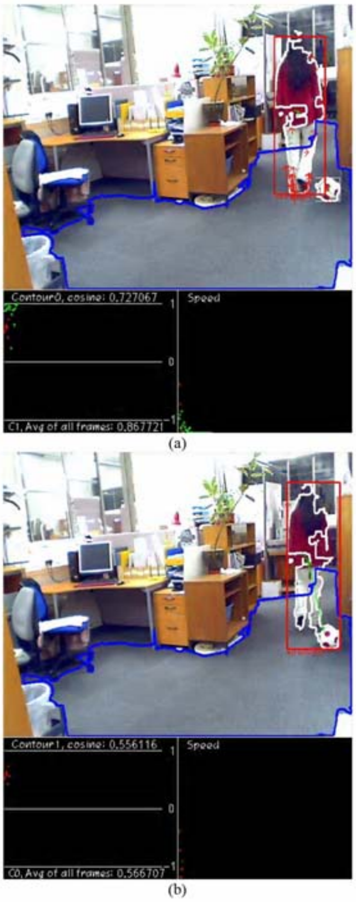
Fig. 8. A split and a merge of a single object: (a) a region of a walking person, (b) its split regions both bounded by red boxes due to information inheritance, and (c) a merge of the split regions

E. Handling of Merges and Splits

Capturing regional merges and splits works well by detecting the closest contour center in the current frame for each contour center in the previous frame and again in the inverse way. Fig. 8 and Fig. 9 present successful cases to recognize a merge and a split of a human and multiple objects respectively.

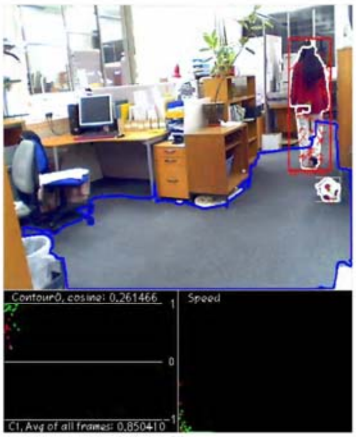
In Fig. 8, a walking person is classified as a human by the dynamic internal motion vectors and bounded by a red box. The person's region splits (Fig. 8b) and the two split regions (one inside the other) both are bounded by a red box because the person region's data is inherited. The split regions merge immediately (Fig. 8c) and the two green arrows heading the merged region's center represents the merge.

In Fig. 9, a person is passing by a ball and their region's past information, which is averaged cosθ values and speeds, is recorded in graphs in red and green respectively (Fig. 9a). When the regions merge only the person's region data are kept (Fig. 9b), and when they split the ball's region data are returned in the graphs (Fig. 9c).


Contour0, cosine: 0.727067
Speed
1
0
C1, Avg of all frames: 0.867721
Contour1, cosine: 0.556116
Speed
1
0
C0, Avg of all frames: 0.566707

(a)

(b)



(c)

Fig. 9. A merge and a split of multiple objects: (a) regions of a person and a ball and their averaged cosθ and speed graphs in red and green respectively, (b) a merge of the regions and graphs only keeping data of the person's region, and (c) a split of the regions and graphs restoring the ball region's past data

However, the system occasionally has problems with differentiating merges and splits of a single object from the ones of multiple objects based on each region's size and history of merges and splits. A person's region splits, for instance, while the person occludes a ball, and the size of the split region from the person is fairly similar to the ball region size. This split would be regarded as the one of multiple objects due to the person's merge history, and the past information of the ball is transferred to the split region.

VI. Conclusion

This research focuses on toddler tracking in an indoor home environment in order to detect risk factors of a toddler's fall. This is different from the studies conducted previously that focused on detecting the actual falls and was specifically tailored towards elderly people. The risk factors are determined as behavioral ones that are dynamic and would require a caregiver's constant supervision. The vision-based tracking methods for real-time detection of the risk factors involve background subtraction, human classification, acquisition of motion and position information, and handling of regional merges and splits.

A single commercial camera is used without having any sensors or markers to be attached on a toddler's body for practical use. The background subtraction works well with a Quickcam Pro 5000 but occasionally produces an error on a human region by splitting it, invoking problems with regional merges and splits.

The human classification has a novel concept by using irregular motions of different body parts. Based on several tests, the threshold of the average cosθ value is identified as 0.8 to differentiate humans from nonhuman objects. In order to reinforce this discrimination to work even against adults or pets, other cues related to toddlers will be used, such as sizes, body ratios, and motion history.

Correct motion and position information can be obtained separately from each foreground object in general, but when the object region merges or splits, so it is ignored at that time. Detection of regional merges and splits works well by connecting closest region centers twice from the previous frame to the current frame and vice versa. But Distinguishing between a single object and multiple objects in merges and splits has problems occasionally only based on each region's size and history of merges and splits. In the future, other cues regarding a toddler's key postures will be used to tackle these problems.

REFERENCES

[1] Child Accident Prevention Trust. (2004). Factsheet: Home Accidents. Available: http://www.capt.org.uk/pdfs/factsheet home accidents.pdf

[2] E. Towner, T. Dowswell, C. Mackereth, and S. Jarvis, What Works in Preventing Unintentional Injuries in Children and Young Adolescents? An Updated Systematic Review, London Health Development Agency, 2001

[3] D. Colvin, C. Lord, G. Bishop, T. Engel, and A. Patra, "A Fall Intervention/Mobility Aid System for Elderly and Rehabilitative Populations," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 13, no. 4, 1991, pp. 1936-1937

[4] G. Williams, K. Doughty, K. Cameron, and D. Bradley, "A Smart Fall & Activity Monitor for Telecare Applications," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 20, no. 3, 1998, pp. 1151-1154

[5] T. Tamura, T. Yoshimura, F. Horiuchi, Y. Higashi, and T. Fujimoto, "An Ambulatory Fall Monitor for the Elderly," in *Proceedings of the 22nd Annual EMBS International Conference*, 2000, pp. 2608-2610

[6] K. Fukaya, "Fall Detection sensor for Fall Protection airbag," in *Proceedings of the Annual Conference of The Society of Instrument and Control Engineers*, 2002, pp. 419-420

[7] N. Noury, "A Smart Sensor for the Remote Follow Up of Activity and Fall Detection of the Elderly," in *Proceedings of the 2nd Annual International IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine & Biology*, 2002, pp. 314-317

[8] B. Najafi, K. Aminian, F. Loew, Y. Blanc, and P. Robert, "Measurement of Stand-Sit and Sit-Stand Transitions using a Miniature Gyroscope and its Application in Fall Risk Evaluation in the Elderly," *IEEE Transactions on Biomedical Engineering*, vol.49, no.8, 2002, pp. 843-851

[9] A. Sixsmith and N. Johnson, "A Smart Sensor to Detect the Falls of the Elderly," *IEEE Pervasive Computing*, vol. 3, no. 2, 2004, pp. 42-47

[10] H. Nait-Charif and S. McKenna, "Activity Summarization and Fall Detection in a Supportive Home Environment," in *Proceedings of the 17th IEEE International Conference on Pattern Recognition*, 2004

[11] K. Perell, A. Nelson, R. Goldman, S. Luther, N. Prieto-Lewis, and L. Rubenstein, "Fall risk assessment measures: and analytic review," *Journal of Gerontology: Medical Sciences*, vol. 56, no. 12, 2001

[12] Child Accident Prevention Trust. (2004). Factsheet: Falls in the Home. Available: http://www.capt.org.uk/pdfs/factsheet falls.pdf

[13] Royal Society for the Prevention of Accidents. (2000-2002). Home and Leisure Accident Surveillance System - Annual Reports. Available: http://www.hassandlass.org.uk/query/reports.htm (need to contact the Information Center for details of individual accidents)

[14] A. Lipton, H. Fujiyoshi, and R. Patil, "Moving Target Classification and Tracking from Real-Time Video," *IEEE Workshop on Application of Computer Vision*, 1998, pp. 8-14

[15] J. Sherrah and S. Gong, "VIGOUR: A System for Tracking and Recognition of Multiple People and their Activities," in *Proceedings of the 15th International Conference on Pattern Recognition*, vol. 1, 2000, pp. 179-183

[16] Q. Cai and J. Aggarwal, "Tracking Human Motion in Structured Environments using a Distributed-Camera System," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.12, No.12, 1999, pp.1241-1247

[17] D. Scholeicher and L. M. Bergasa, "People Tracking and Recognition using the Multi-Object Particle Filter Algorithm and Hierarchical PCA Method," in *Proceedings of EUROCON 2005 - The International Conference on "Computer as a tool"*, 2005

[18] A. S. Micilotta, "Detection and Tracking of Humans for Visual Interaction," Ph.D. dissertation, School of Electronics and Physical Science, University of Surrey, Surrey, United Kingdom, 2005

[19] W. Hu, T. Tan, L. Wang, and S. Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviors," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 34, no. 3, 2004, pp. 334-352

[20] A. Utsumi, H. Mori, J. Ohya, and M. Yachida, "Multiple-View-Based Tracking of Multiple Humans," in *Proceedings of the 14th International Conference on Pattern Recognition*, 1998, vol. 1, pp. 597-601

[21] A. Mittal and L. S. Davis, "M2 Tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene," *International Journal of Computer Vision*, vol. 51, no. 3, 2003, pp. 189-203

[22] J. P. Batista, "Tracking Pedestrians under Occlusion Using Multiple Cameras," in *Proceedings of the International Conference on Image Analysis and Recognition*, vol. 3212, 2004, pp. 552-562

[23] H. B. Kang and S. H. Cho, "Multi-modal Face Tracking in Multi-camera Environments," in *Proceedings of the 11th International Conference on Computer Analysis of Images and Patterns*, vol. 3691, 2005, pp. 814-821

[24] Q. Zhou and J. K. Aggarwal, "Object Tracking in an outdoor environment using fusion of features and cameras," *Image and Vision Computing*, vol. 24, issue. 11, 2006, pp. 1244-1255

[25] G. Medioni, I. Cohen, F. Bremond, S. Hongeng, and R. Nevatia, "Event Detection and Analysis from Video Streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, 2001, pp. 873-889

[26] A. Genovesio and J. Olivo-Marin, "Split and Merge Data Association Filter for Dense Multi-Target Tracking," in *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 4, 2004, pp. 677-680

[27] P. Kumar, S. Ranganath, K. Sengupta, and H. Weimin, "Cooperative Multitarget Tracking with Efficient Split and Merge Handling," IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 12, 2006, pp. 1477-1490

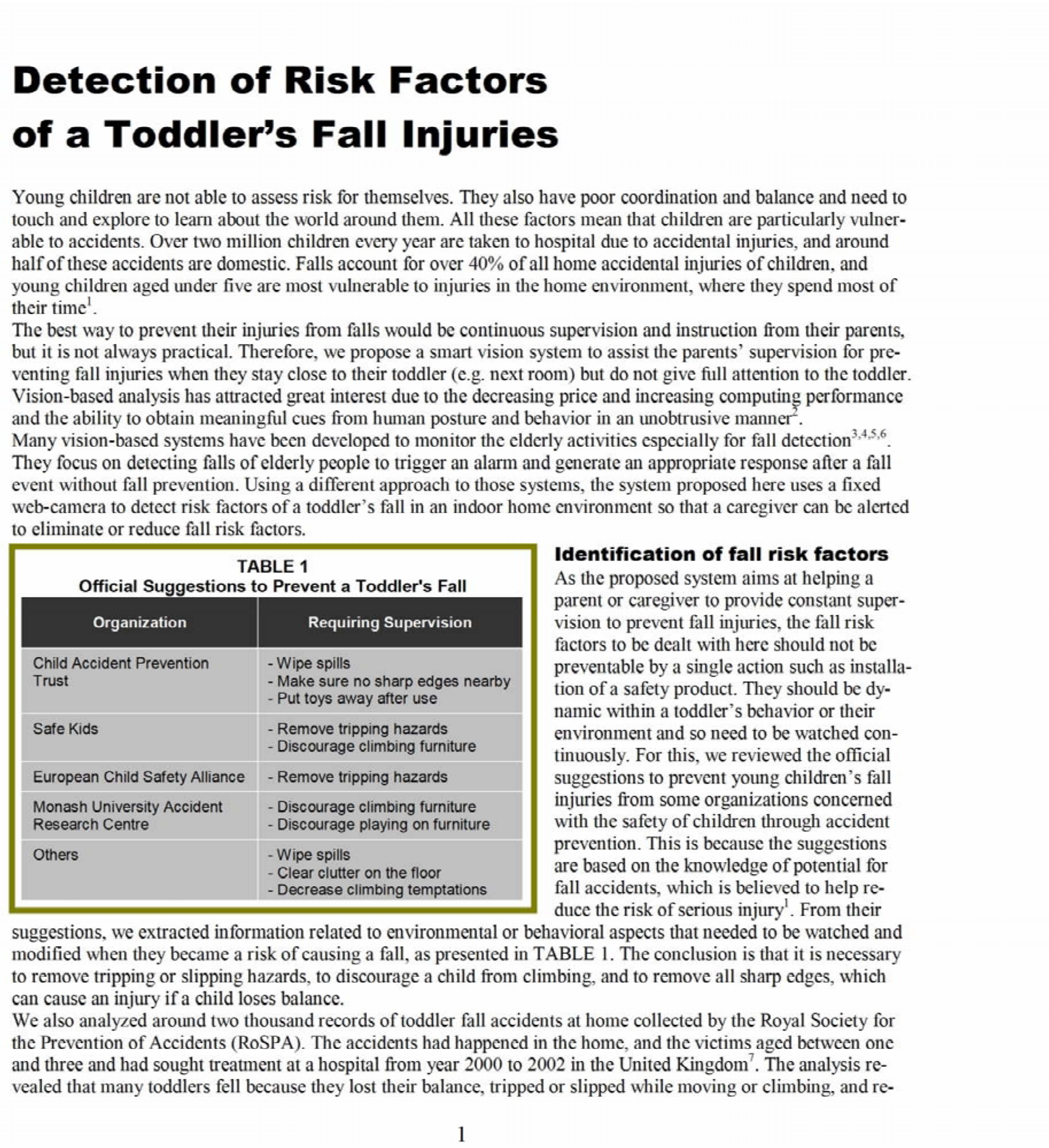
[28] S. J. McKenna, S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld, "Tracking Groups of People," Computer Vision and Image Understanding, vol. 80, issue. 1, 2000, pp. 42-56

[29] Intel Corporation, "Open Source Computer Vision Library: Reference Manual," Available: http://switch.dl.sourceforge.net/sourceforge/opencvlibrary/OpenCVReferenceManual.pdf, 2000

[30] Hamleys, Available: https://www.hamleys.com/

Appendix H

IEEE Pervasive Computing



Detection of Risk Factors of a Toddler's Fall Injuries

Young children are not able to assess risk for themselves. They also have poor coordination and balance and need to touch and explore to learn about the world around them. All these factors mean that children are particularly vulnerable to accidents. Over two million children every year are taken to hospital due to accidental injuries, and around half of these accidents are domestic. Falls account for over 40% of all home accidental injuries of children, and young children aged under five are most vulnerable to injuries in the home environment, where they spend most of their time[1].

The best way to prevent their injuries from falls would be continuous supervision and instruction from their parents, but it is not always practical. Therefore, we propose a smart vision system to assist the parents' supervision for preventing fall injuries when they stay close to their toddler (e.g. next room) but do not give full attention to the toddler. Vision-based analysis has attracted great interest due to the decreasing price and increasing computing performance and the ability to obtain meaningful cues from human posture and behavior in an unobtrusive manner[2].

Many vision-based systems have been developed to monitor the elderly activities especially for fall detection[3,4,5,6]. They focus on detecting falls of elderly people to trigger an alarm and generate an appropriate response after a fall event without fall prevention. Using a different approach to those systems, the system proposed here uses a fixed web-camera to detect risk factors of a toddler's fall in an indoor home environment so that a caregiver can be alerted to eliminate or reduce fall risk factors.

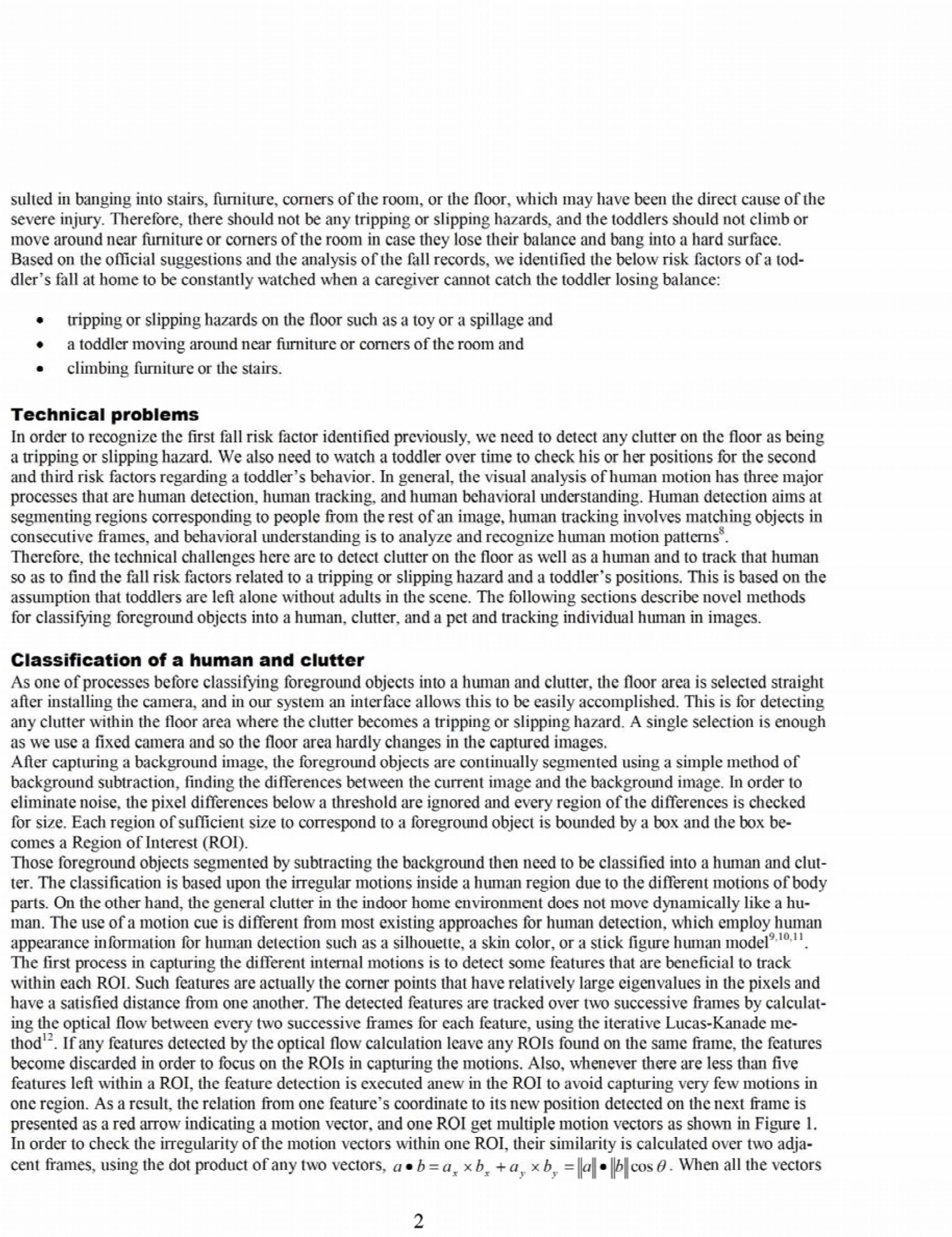
TABLE 1
Official Suggestions to Prevent a Toddler's Fall

Organization	Requiring Supervision
Child Accident Prevention Trust	- Wipe spills - Make sure no sharp edges nearby - Put toys away after use
Safe Kids	- Remove tripping hazards - Discourage climbing furniture
European Child Safety Alliance	- Remove tripping hazards
Monash University Accident Research Centre	- Discourage climbing furniture - Discourage playing on furniture
Others	- Wipe spills - Clear clutter on the floor - Decrease climbing temptations

Identification of fall risk factors

As the proposed system aims at helping a parent or caregiver to provide constant supervision to prevent fall injuries, the fall risk factors to be dealt with here should not be preventable by a single action such as installation of a safety product. They should be dynamic within a toddler's behavior or their environment and so need to be watched continuously. For this, we reviewed the official suggestions to prevent young children's fall injuries from some organizations concerned with the safety of children through accident prevention. This is because the suggestions are based on the knowledge of potential for fall accidents, which is believed to help reduce the risk of serious injury[1]. From their suggestions, we extracted information related to environmental or behavioral aspects that needed to be watched and modified when they became a risk of causing a fall, as presented in TABLE 1. The conclusion is that it is necessary to remove tripping or slipping hazards, to discourage a child from climbing, and to remove all sharp edges, which can cause an injury if a child loses balance.

We also analyzed around two thousand records of toddler fall accidents at home collected by the Royal Society for the Prevention of Accidents (RoSPA). The accidents had happened in the home, and the victims aged between one and three and had sought treatment at a hospital from year 2000 to 2002 in the United Kingdom[7]. The analysis revealed that many toddlers fell because they lost their balance, tripped or slipped while moving or climbing, and re-



sulted in banging into stairs, furniture, corners of the room, or the floor, which may have been the direct cause of the severe injury. Therefore, there should not be any tripping or slipping hazards, and the toddlers should not climb or move around near furniture or corners of the room in case they lose their balance and bang into a hard surface.
Based on the official suggestions and the analysis of the fall records, we identified the below risk factors of a toddler's fall at home to be constantly watched when a caregiver cannot catch the toddler losing balance:

- tripping or slipping hazards on the floor such as a toy or a spillage and
- a toddler moving around near furniture or corners of the room and
- climbing furniture or the stairs.

Technical problems

In order to recognize the first fall risk factor identified previously, we need to detect any clutter on the floor as being a tripping or slipping hazard. We also need to watch a toddler over time to check his or her positions for the second and third risk factors regarding a toddler's behavior. In general, the visual analysis of human motion has three major processes that are human detection, human tracking, and human behavioral understanding. Human detection aims at segmenting regions corresponding to people from the rest of an image, human tracking involves matching objects in consecutive frames, and behavioral understanding is to analyze and recognize human motion patterns[8].

Therefore, the technical challenges here are to detect clutter on the floor as well as a human and to track that human so as to find the fall risk factors related to a tripping or slipping hazard and a toddler's positions. This is based on the assumption that toddlers are left alone without adults in the scene. The following sections describe novel methods for classifying foreground objects into a human, clutter, and a pet and tracking individual human in images.

Classification of a human and clutter

As one of processes before classifying foreground objects into a human and clutter, the floor area is selected straight after installing the camera, and in our system an interface allows this to be easily accomplished. This is for detecting any clutter within the floor area where the clutter becomes a tripping or slipping hazard. A single selection is enough as we use a fixed camera and so the floor area hardly changes in the captured images.

After capturing a background image, the foreground objects are continually segmented using a simple method of background subtraction, finding the differences between the current image and the background image. In order to eliminate noise, the pixel differences below a threshold are ignored and every region of the differences is checked for size. Each region of sufficient size to correspond to a foreground object is bounded by a box and the box becomes a Region of Interest (ROI).

Those foreground objects segmented by subtracting the background then need to be classified into a human and clutter. The classification is based upon the irregular motions inside a human region due to the different motions of body parts. On the other hand, the general clutter in the indoor home environment does not move dynamically like a human. The use of a motion cue is different from most existing approaches for human detection, which employ human appearance information for human detection such as a silhouette, a skin color, or a stick figure human model[9,10,11].

The first process in capturing the different internal motions is to detect some features that are beneficial to track within each ROI. Such features are actually the corner points that have relatively large eigenvalues in the pixels and have a satisfied distance from one another. The detected features are tracked over two successive frames by calculating the optical flow between every two successive frames for each feature, using the iterative Lucas-Kanade method[12]. If any features detected by the optical flow calculation leave any ROIs found on the same frame, the features become discarded in order to focus on the ROIs in capturing the motions. Also, whenever there are less than five features left within a ROI, the feature detection is executed anew in the ROI to avoid capturing very few motions in one region. As a result, the relation from one feature's coordinate to its new position detected on the next frame is presented as a red arrow indicating a motion vector, and one ROI get multiple motion vectors as shown in Figure 1.

In order to check the irregularity of the motion vectors within one ROI, their similarity is calculated over two adjacent frames, using the dot product of any two vectors, $a \bullet b = a_x \times b_x + a_y \times b_y = \|a\| \bullet \|b\| \cos\theta$. When all the vectors

2

in one ROI are defined $a^0, a^1, ..., a^n$, the average of cosθ between the vectors can be calculated as $avg(\cos\theta) = [\sum_{i=0}^{n-1} \sum_{j=i+1}^{n} \{(a_x^i \times a_x^j) + (a_y^i \times a_y^j)\} / \|a^i\| \bullet \|a^j\|] / \sum_{k=1}^{n} k$. As the vectors head in the same direction when θ is 0 and cos0 is 1, the closer to 1 the average of cosθ is, the closer is the similarity of the vectors. We conducted several tests and defined the threshold value to classify a human and clutter.

Classification of a human and a pet

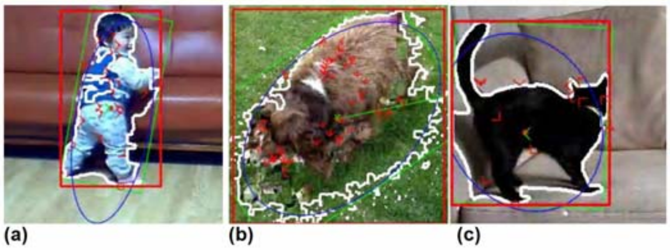


Figure 1. Angle changes of the ellipse fitted to a contour as a cue to differentiate a human from a pet; (a) a toddler, (b) a dog, and (c) a cat

Many families raise children with a dog or a cat. In case when both a toddler and a pet appear together, the toddler classification using the different internal motions would not work well since a pet also moves its body parts differently. Hence, we added another dynamic motion cue to reinforce the human classification.

We observed that the contour of a dog or a cat changes dynamically when moving around because they tend to turn around frequently in a limited space such as the indoor home environment. A cat also stretches the body or its tail widely when it moves around. In the meantime, the posture of a toddler is usually upright when walking. From this, we determined to use the angle changes of the ellipse fitted to a foreground region over the time, shown as the blue ellipse in Figure 1, as one of the additional motion cues. In order to capture the diversity, the changes in angle of the ellipse fitted to each contour are recorded only when the whole body moves, and the mean and the standard deviation of the history is calculated.

The other cue used for the differentiation of a human and a pet is the actual height information because toddlers are generally taller than pets. For this, the camera used in this system is calibrated using a 2D chessboard, the size of whose square unit is known, when the camera is installed. Although the actual height is 3D information and we have only 2D images from a single camera, it is possible to calculate the 3D information when the camera is calibrated with a 2D chessboard lying on the floor and the height always starts from the floor and is perpendicular to the floor. Since toddlers and pets barely jump but walk around in the indoor home environments, we decided to estimate the actual height of each foreground object using a 2D chessboard with the assumption that toddlers and pets never jump on the floor. The relationship between a point on the ground of the 3D world space, (X, Y, 1) and its projection onto the image plane, (u, v) can be defined with the corner points of a 2D chessboard as $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \\ 1 \end{bmatrix}$. Then the relationship parameters, h_{ij} (i=1,2,3,4, j=1,2,3) are used to calculate the actual height, h, which is the distance between the lowest point, (a, b, 1) and the highest point, (a, b, h+1) of a toddler or a pet. As toddlers can become shorter than their standing height when they bend their body, for instance, the heights are calculated and recorded when they move, and the history is used for this classification. The calculated heights of a pet are often different from the actual lengths because pets do not keep the trunk upright like a human when the whole body is mobile. But the relatively large changes of a pet's heights would be useful for the classification.

The two cues, the angle changes of an ellipse fitted to each foreground contour and the 3D actual heights of the foreground object, then need to be applied together to recognize the difference between a human and a pet. For the recognition with the lower uncertainty based on the two cues, we used the Rényi Entropy, which quantifies the uncertainty of a system based on the system's probabilities[13]. The Rényi Entropy of order α is $H_{\alpha}(X) = \frac{1}{1-\alpha}\log(\sum_{i=1}^{n} p_i^{\alpha})$, where p_i are the probabilities. For the entropy calculation, we built our system to set up a reference of probabilities to be a toddler and the one to be a pet separately by estimating the ellipse angle changes and actual heights from a brief footage of the subject toddler and a footage of his or her pet. This is done when the system is installed to get the system customized to different subjects. During real-time image analysis, the references are checked with the data regarding the ellipse angle changes and actual heights from each nonclutter foreground region to obtain the region's probabilities to be the toddler and those to be the pet. Those two groups of probabilities are respectively used in the Rényi Entropy calculation to return the quantified uncertainty to be a toddler and the one to be a pet, and the smaller uncertainty determines the classification.

Tracking

Now detected toddlers individually need to be tracked over time to find the fall risk factors related to their positions. Video tracking is the process of locating a moving object in time by matching identical objects between two consecutive images. Our method for tracking is to detect closest foreground objects over two frames. This is because our frame rate is high enough to capture a quickly moving toddler at a very close location from where the toddler is found at the previous frame. The center of mass of each region from one frame is used to calculate the distances from all the centers of mass detected on the next frame and is connected to the one with the shortest distance. This connection is conducted separately on every center detected on each frame for tracking of individual objects.

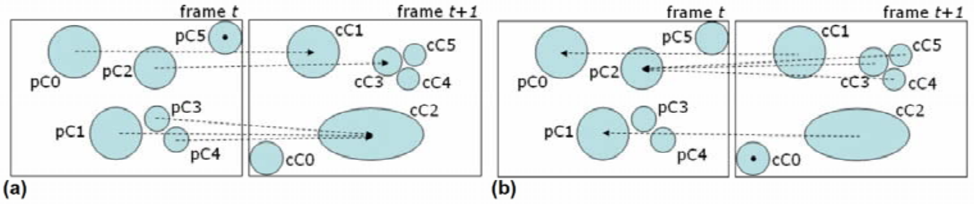


Figure 2. Double closest center connections over two consecutive frames; (a) from the former to the latter frames, (b) from the latter to the former frames

Handling of regional merges and splits

In practice, self-occlusion, and occlusions between different moving objects or between moving objects and the background are inevitable[14]. Due to those occlusions, multiple objects can appear to be one connected component and a single object can split into multiple blobs, which also can be caused by errors in the background subtraction. Our method of detecting those regional merges and splits is to connect the closest center connection over two frames twice from the previous to the latter frames and vice versa as presented in Figure 2. Merges can be found in the first connection such as pC1, pC3, and pC4 shown in Figure 2(a), and splits can be found in the second connection such as cC3, cC4, and cC5 in Figure 2(b). We also limited the closest center detection within each region so as to differentiate a disappearing region such as pC5 in Figure 2(a) and an appearing region such as cC0 in Figure 2(b).
The detected merges and splits need to be further examined to check if they resulted from occlusion of multiple objects or from separated blobs of an object. A regional merge can happen to a single object after it has split, but a split does not precede a merge in multiple objects. Meanwhile, a merge should come before a split in multiple objects but not in a single object. Hence, the differentiation is based on a history of merges and splits for each region with the assumption that every region first appears in the scene corresponds to a single object.

Detection of the fall risk factors

Since clutter, as one of the fall risk factors, is a tripping or slipping hazard on the floor, anything classified as clutter is checked if it is stationary on the floor. For this, any clutter region within the floor area is checked if there is any motion inside. For cases of clutter without tripping or slipping hazards such as toys being played with beside a sitting toddler, the movement of the subject toddler is checked at the same time.

As the rest of the fall risk factors are toddlers moving around of or climbing furniture or room structures, they can be detected by a toddler's relative position against furniture or room structures. Assuming that a toddler never jumps on the floor, we considered the vertically lowest point of the toddler's contour as being where the toddler stands in the environment. So the lowest point is checked every frame to see if the toddler approaches the boundary of the floor area or if leaves the floor area with the consideration that the nonfloor area is filled with furniture, hazardous corners, or stairs.

Evaluation of the methods

For the visual analysis to detect the fall risk factors, we employed one Logitech Quickcam Pro5000 in order to capture real-time images in a fixed position. The image size was 640x480 pixels, the frame rate was 30 frames per second, and the developed software had dialog-based interfaces to set up and control the system.

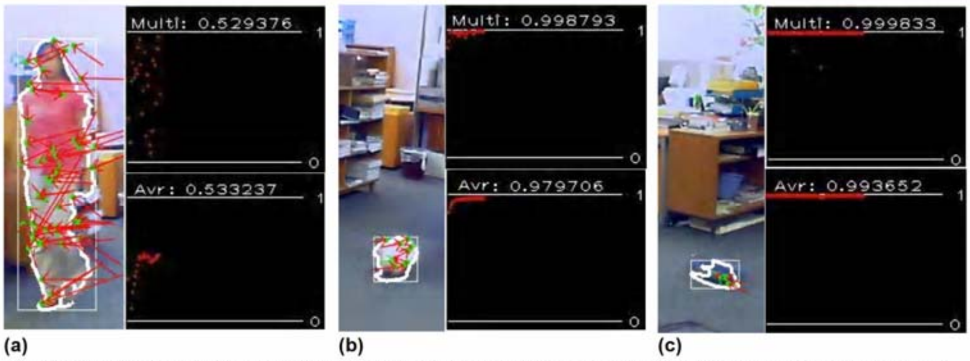


Figure 3. Tests for defining the threshold value to classify a human and clutter; (a) a walking human region with internal motion vectors, the graph showing the average of cosθ between every two vectors from each frame during the human walk, and the graph presenting the average of all the past frames' cosθ values, (b) a rolling ball and its graphs, and (c) a moving radio controlled model car and its graphs

Classification of a human and clutter

The simple method of background subtraction was satisfactory using 640x480 images from the QuickCam Pro 5000. Logitech's other web cameras of lower or higher performances such as the QuickCam Pro 4000 or the Ultra Vision were more prone to noise due to low resolution or visible compression artifacts. However, there are still some limitations because the background subtraction method was sensitive to lights and the scene needed to be bright but should not have contained direct rays.

As our method to classify foreground objects into a human and clutter is based on the dynamic internal motions of human body parts, several tests were carried out to detect the threshold value of the similarity of the motion vectors. We used an adult, a ball, and a radio controlled model car respectively, representing a human, rolling, and straight motions as presented in Figure 3. As Hamleys, one of the largest toy shops in the world, revealed that other toys moving more dynamically are for children over three years old who are not toddlers anymore, we did not use them for the tests.

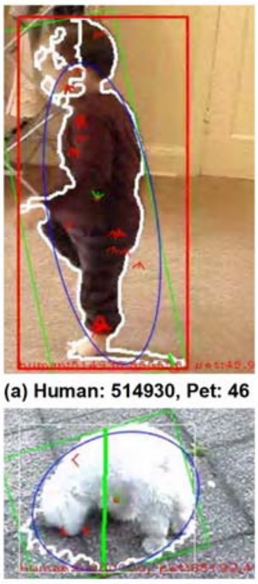
TABLE 2
Method Evaluation

Object in the Scene	Total Frame No.	Use of Cues		Percentage of Frames Correctly Recognized			
		Internal Motion Vectors	Ellipse Angle & Height	Merges & Splits	Classification		Human Status
Moving/Static Clutter	687	✓		99.42%	98.40%		
Pet	836	✓	✓	99.04%	98.69% (54.43% as Moving Clutter)		
Toddler	376	✓	✓	96.81%	91.22%		75.53%
Clutter + Toddler	556	✓		100%	Clutter	99.28%	
					Toddler	75.72%	74.64%
Pet + Toddler	570	✓	✓	98.25%	Pet	93.68% (68.25% as Clutter)	
					Toddler	76.67%	66.49%

The similarity of the entire vectors in one ROI was calculated as the averaged value of cosθ between every two vectors and presented over time in the upper graphs in Figure 3. It was fairly dynamic between 0 and 1 for a walking human as shown in the upper graph in Figure 3(a) and was constantly close to 1 for a rolling ball and a radio controlled model car as shown in Figure 3(b) and (c). As sometimes the cosθ value got considerably close to 1 for a human motion and somewhat lower than 1 for an object motion, the cosθ values from all the past frames were also averaged at every frame as presented in the lower graphs in Figure 3(a), (b), and (c). Based on several tests, we found that the average of every past frame's cosθ value stayed under 0.75 for a walking human and over 0.9 for a rolling ball and a moving model car. Hence, we defined the threshold to classify human and nonhuman as 0.8.
The classification method was evaluated with several short sequences whose foreground objects are adequately segmented by the simple background subtraction method used here. This is because ambiguous foreground segmentations generally cause errors in the classification. The sequences were manually extracted and the results are shown in TABLE 2. Clutter was classified well at high rates into moving and static also when it appeared with a toddler in the scene.

Classification of a human and a pet

After filtering out clutter using internal motion vectors, the classification of the rest foreground objects into a human and a pet performed well on the basis of the comparison of the quantified uncertainties to be a toddler and a pet as seen in TABLE 2. The system bounds any foreground object, whose uncertainty to be a



(a) Human: 514930, Pet: 46

(b) Human: 0, Pet: 85192

Figure 4. Rényi-entropy-based separate uncertainty calculations for a foreground object to be a human and a pet: (a) an object with lower uncertainty to be a human than uncertainty to be a pet and (b) an object with higher uncertainty to be a human. The displayed values are the results of $\sum_{i=1}^{n} p_i^2$, which is omitted $-\log$ from the original Rényi Entropy of order 2.

toddler is lower than the uncertainty to be a pet, with a thick red box, and Figure 4 shows a result that a toddler and a pet were classified correctly. The pet used for the evaluation in TABLE 2 was classified as moving clutter in around a half of the total frames. This is acceptable in our system because a pet can become a tripping hazard like clutter and our system generally focuses on a toddler.

Tracking

The connection of corresponding regions' centers of mass over two successive frames worked well to track objects and detect regional merges and splits. However, the system occasionally had problems with distinguishing merges and splits of a single object from the ones of multiple objects based on the history of merges and splits. When a person's region split, for instance, while the person occluded a ball, the split would be regarded as the one of multiple objects due to the person's merge history. However, splits of a single object tended to be merged soon as they were usually generated due to occasional errors in background subtraction and wrongly recognized region after a merge or a split could be reclassified with new information from the next few frames.

Issues

Some people might refute this system to detect fall risk factors by telling that young children, especially toddlers, who are learning to walk, learn risks and practice motor skills from falling. However, injury is the main cause of death and a major cause of ill health and disability in childhood[15], and injury due to falling is a focus here. This system cannot be a complete replacement of parents' supervision, but it is expected to aid their prevention of a toddler's fall injuries and let them know exactly what happened with a video even after a fall event so that parents can provide a prompt and appropriate treatment. As the images are captured and analyzed locally in the home environment, parents could be relieved of privacy issues.

In the technical side, this research has made novel approaches to classify humans, clutter, and pets in images using dynamic motion cues. The motion cues are relatively strong at occlusions since most approaches for human detection use cues related to human appearance and partial occlusion can make a human look like a nonhuman object based on the appearance cues. Our methods, however, have difficulties to be applied for a long time due to the use of the simple background subtraction method. Thus the first of the future work would be to improve the background subtraction method so as to segment foreground objects clearly without noise and well support the subsequent process, human detection.

Although we use only one camera in this paper, it can be implemented to use multiple cameras to cover several areas inside a house. As our methods are based on the use of a fixed camera, what parents need to do when installing an additional camera is to select the floor area and put a chessboard for calibration on the floor for a while at a single time. Also when they want to move a camera to another place, just the initial settings are required. As the floor selection is to define the safe area for a toddler to be alone, parents can include any other nonfloor object considered as a safe place to the selection. Although respective short footages of a toddler and a pet are also initially required for our human classification method, our system could be implemented to collect the footages at intervals during its real-time image analysis in order to adjust the system to the toddler's growth.

This research has not dealt with automatic actions after recognition. The actions after detection of the fall risk factors could be direct such as the parent's voice to instruct the subject toddler on what to do to avoid fall injuries or could be indirect by informing a nearby caregiver to come and remove the risk factors.

[1] CAPT, "Factsheet – Home Accidents," Child Accident Prevention Trust, Nov. 2004; http://www.capt.org.uk/pdfs/factsheet%20home%20accidents.pdf.

[2] L. Wang, "From Blob Metrics to Posture Classification to Activity Profiling," *Proc. 18th Int'l Conf. Pattern Recognition* (ICPR 06), 2006, vol. 4, pp. 736-739.

[3] A.G. Hauptmann, J. Gao, R. Yan, Y. Qin, J. Yang, and H.D. Wactlar, "Automated Analysis of Nursing Home Observations," *IEEE Pervasive Computing*, Apr.-Jun. 2004, vol. 3, no. 2, pp. 15-21.

[4] A. Sixsmith and N. Johnson, "A Smart Sensor to Detect the Falls of the Elderly," *IEEE Pervasive Computing*, Apr.-Jun. 2004, vol. 3, no. 2, pp. 42-47.

[5] D. Anderson, J.M. Keller, M. Skubic, X. Chen, and Z. He, "Recognizing Falls from Silhouettes," *Proc. 28th Int'l Conf. Eng. Med. Biol. Soc.* (EMBS 06), 2006, pp. 6388-6391.

[6] B. Jansen and R. Deklerck, "Context Aware Inactivity Recognition for Visual Fall Detection," *Pervasive Health Conf. and Workshops*, 2006, pp. 1-4.

[7] DTI, "24th Report of the Home and Leisure Accident Surveillance System," Royal Society for the Prevention of Accidents, Jan. 2004; http://www.hassandlass.org.uk/query/reports/2000_2002.pdf.

[8] L. Wang, W. Hu, and T. Tan, "Recent Developments in Human Motion Analysis," *Pattern Recognition*, Mar. 2003, vol. 36, no. 3, pp. 585-601.

[9] P. Fihl, R. Corlin, S. Park, T.B. Moeslund, and M.M. Trivedi, "Tracking of Individuals in Very Long Video Sequences," *2nd Int'l Symp. Advances in Visual Computing* (ISVC 06), 2006, pp. 60-69.

[10] M.T. Yang, Y.C. Shih, and S.C. Wang, "People Tracking by Integrating Multiple Features," *Proc. 17th Int'l Conf. Pattern Recognition*, 2004, vol. 4, pp. 929-932.

[11] B. Fan and Z.F. Wang, "Pose Estimation of Human Body Based on Silhouette Images," *Proc. Int'l Conf. Info. Acquisition*, 2004, pp. 296-300.

[12] B.D. Lucas and T. Kanade, "An Iterative Image Registration Technique with and Application to Stereo Vision," *Proc. 17th Int'l Joint Conf. Artificial Intelligence*, 1981, pp. 674-679.

[13] A. Rényi, "On Measures of Information and Entropy," *Proc.4th Berkeley Symp. Math. Stat. Probability*, 1960, vol. 1, pp. 547-561.

[14] W. Hu, T. Tan, L. Wang, and S. Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviors," *IEEE Trans. Systems Man Cybernetics*, 2004, vol. 34, no. 3, pp. 334-352.

[15] L.M. Millward, A. Morgan, and M.P. Kelly, *Prevention and Reduction of Accidental Injury in Children and Older People*, Health Development Agency, UK National Health Service, 2003.