Proceedings of the 2007 IEEE International Conference on
Networking, Sensing and Control, London, UK, 15-17 April 2007

TuesM04

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

TABLE I
IMAGE PROCESSING METHODS USED TO RECOGNIZE RISK FACTORS

Fall Risk Factors	Image Processing Clutter Detection	Toddler Tracking (Motion)	(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.
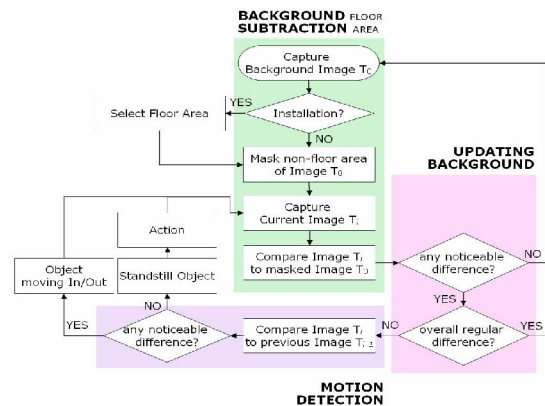

Fig. 1. Clutter detection consists of background subtraction, background image update, and motion detection.

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.
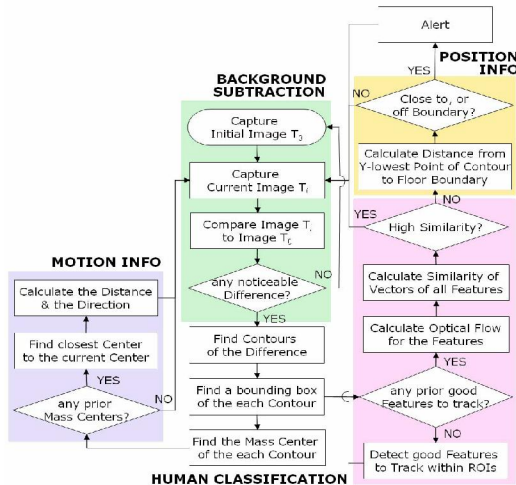

Fig. 2. *Toddler tracking* involves finding a toddler, matching identical regions in successive frames, and calculating speed, direction, and position of the toddler.

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.
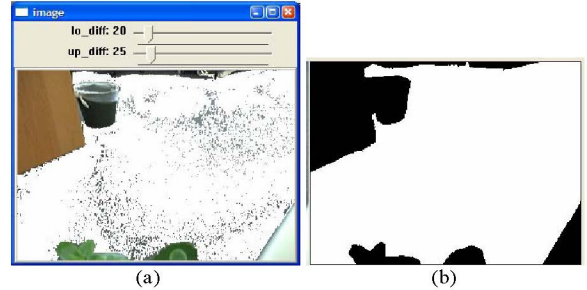

Fig. 3. A floor area is selected manually by filling neighboring pixels whose value is similar to the pixel clicked by the user.

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]. δ_{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.

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

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.
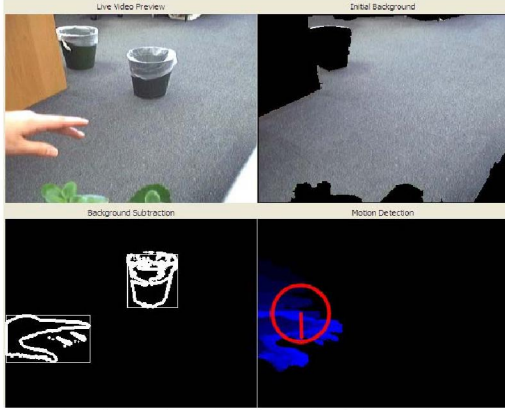

Fig. 4. (anticlockwise from the top-left) a) live video preview, b) background-subtracted regions, c) a moving region, and d) non-floor-marked background image

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).

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

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.

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

When all the vectors in one ROI are defined $\vec{a}^0, \vec{a}^1, ..., \vec{a}^n$, the average of *cosθ* 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 $\cos\theta$ 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 $\cos\theta$ value sometimes gets considerably close to 1 for a human motion and somewhat lower than 1 for an object motion, the average of the $\cos\theta$ values from the past frames is calculated. Based on the tests, we found that the average of every past frame's $\cos\theta$ 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.
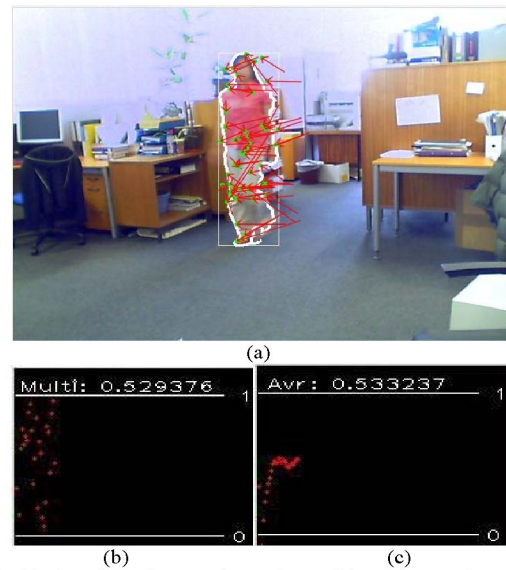

Fig. 5. (a) A walking human shows internal irregular motion vectors. (b) The graph shows the average of $\cos\theta$ between every two vectors from the each frame during the human walk and it is also irregular. c) The graph presents the average of all the past frames' $\cos\theta$ values.

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 $\cos\theta$ 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,
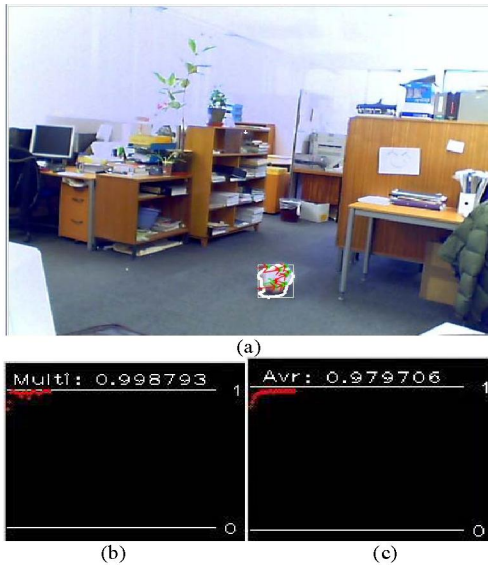
such as a nursery school.

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.
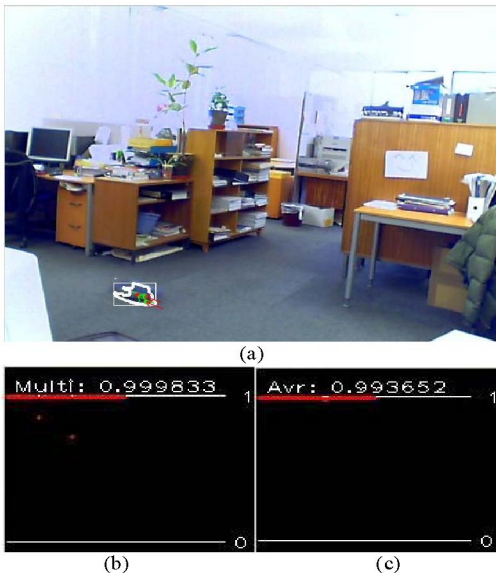

Fig. 7. (a) A radio controlled model car has almost parallel vectors. b) While the car is moving forwards and backwards, the averaged cosθ from each frame is fairly close to 1. c) The average of every frame's cosθ value is also considerably close to 1.

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