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 our 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 a 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.

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.

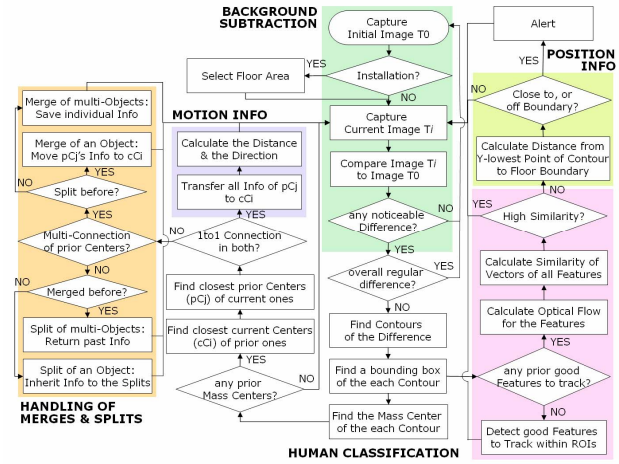


Fig. 1. System work flow

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.

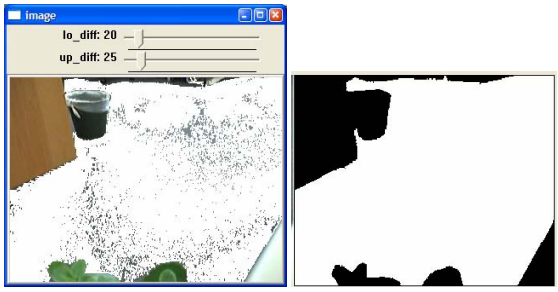
A. Installation (Floor Selection)

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:

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

where v_0 is the value of one of the pixels in the repainted domain that begins with the selected pixels [29]. 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. 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.

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

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, $\vec{a} \bullet \vec{b} = a_x \times b_x + a_y \times b_y = ab\cos\theta$, the similarity of the motion vectors in one ROI is calculated over two adjacent frames.

$$avg(\cos\theta) = [\sum_{i=0}^{n-1}\sum_{j=i+1}^{n}\{(a_x^i \times a_x^j)+(a_y^i \times a_y^j)\}/a^i a^j]/\sum_{k=0}^{n} k \quad (3)$$

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).

$$x_c = \sum_x\sum_y xI(x,y)/\sum_x\sum_y I(x,y)$$
$$y_c = \sum_x\sum_y yI(x,y)/\sum_x\sum_y I(x,y) \quad (4)$$

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

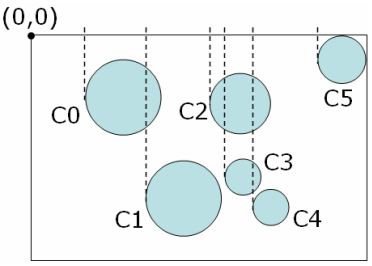


Fig. 3. Contour indexing

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.

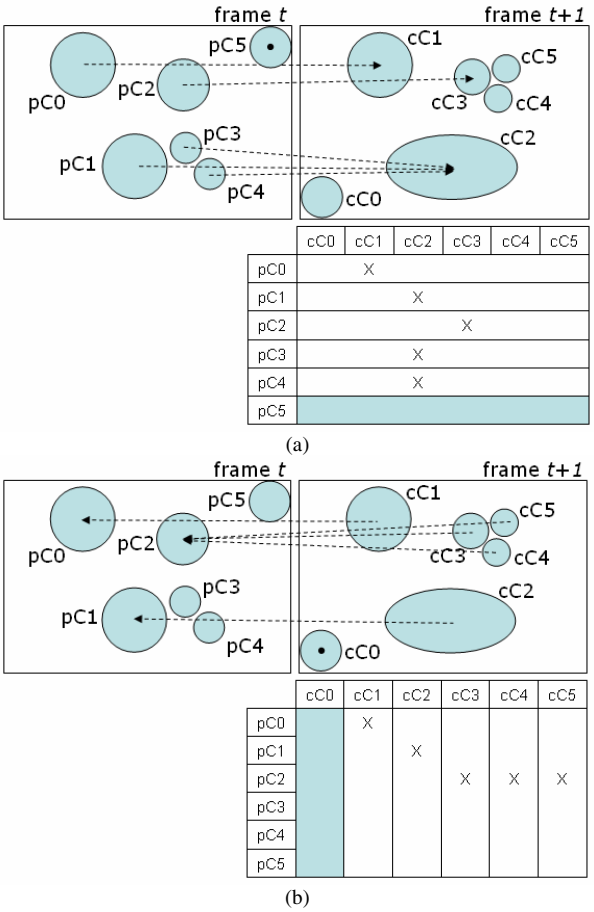


Fig. 4. (a) Detections of closest centers from the previous frame and (b) from the current frame

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 detection of the closest centers over two frames is limited within the regions. 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, 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.

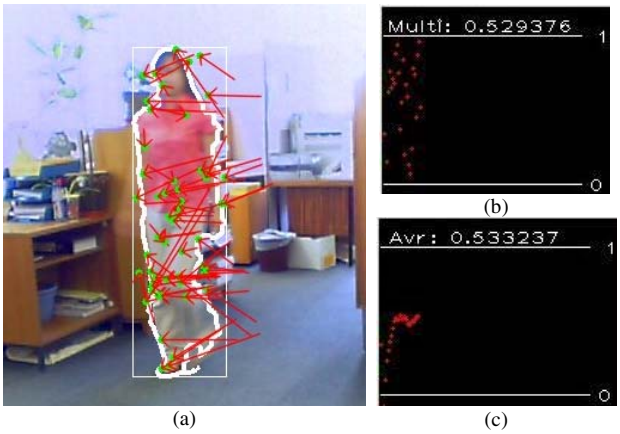
B. Background Subtraction

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, discarding is conducted in each ROI and new features are detected in any ROI with less than five features.



(a) (b) (c)

Fig. 5. (a) A walking human region with internal motion vectors (b) its graph showing the average of cos between every two vectors from each frame during the human walk and (c) graph presenting the average of all the past frames' cos values.

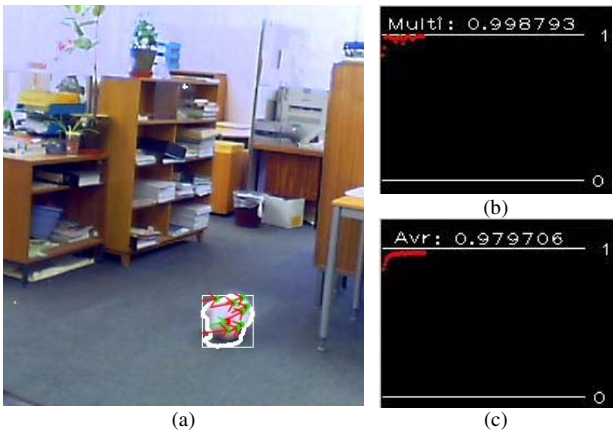


Fig. 6. (a) A rolling ball region and (b) its graph of each frame's cos , and (c) graph of the averages of past frames' cos values

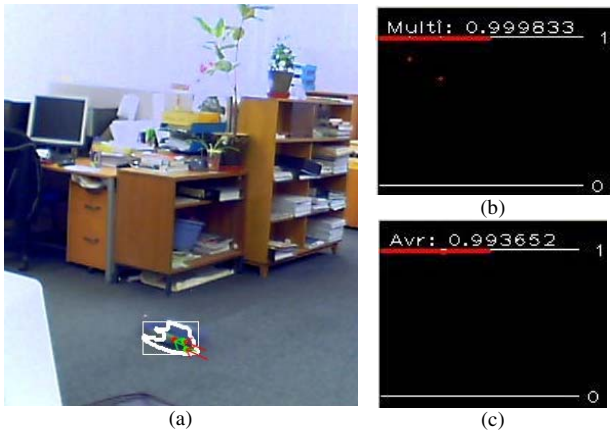


Fig. 7. (a) A region of a radio controlled model car and (b) its graph of each frame's cos while the car is moving forwards and backwards and (c) graph of the averages of past frame's cos values

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 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.

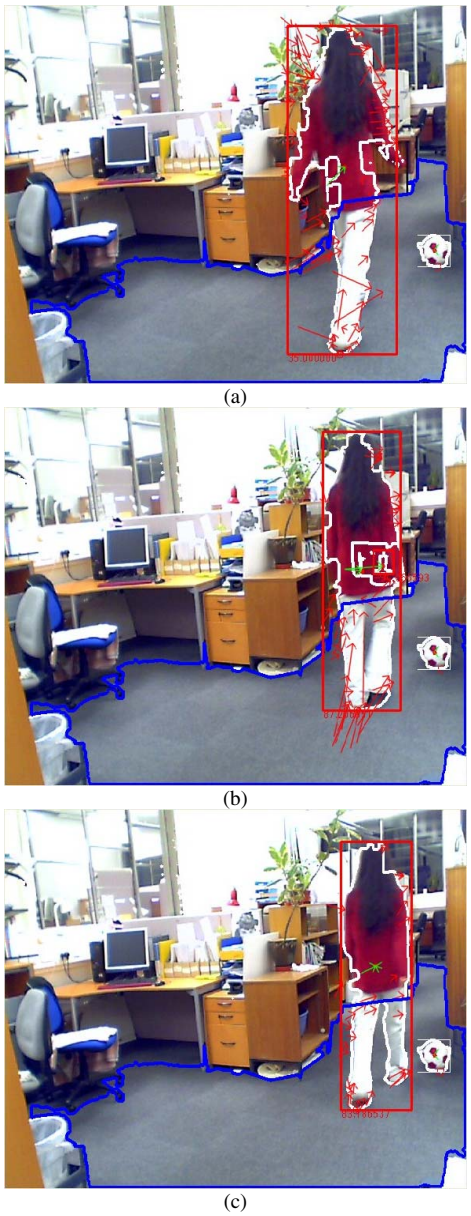


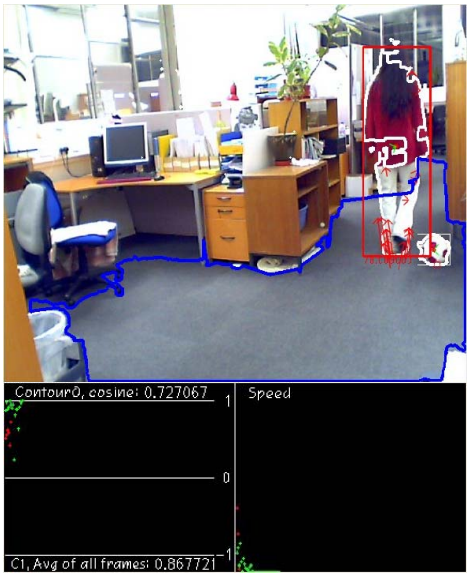
Fig. 8. A split and a merge of a single object: (a) a region of a walking person, (b) its split regions both bounded by red boxes due to information inheritance, and (c) a merge of the split regions

E. *Handling of Merges and Splits*

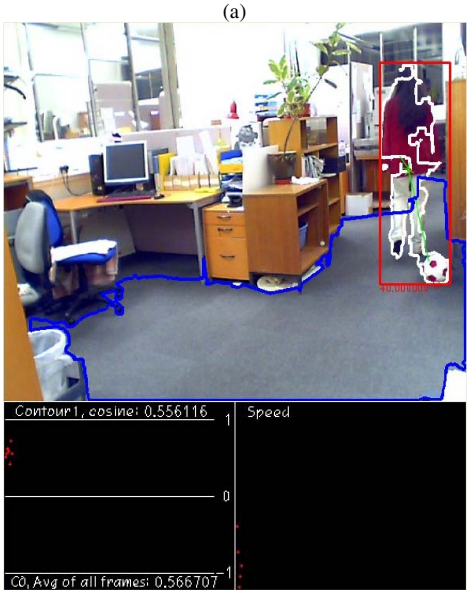
Capturing regional merges and splits works well by detecting the closest contour center in the current frame for each contour center in the previous frame and again in the inverse way. Fig. 8 and Fig. 9 present successful cases to recognize a merge and a split of a human and multiple objects respectively.

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.

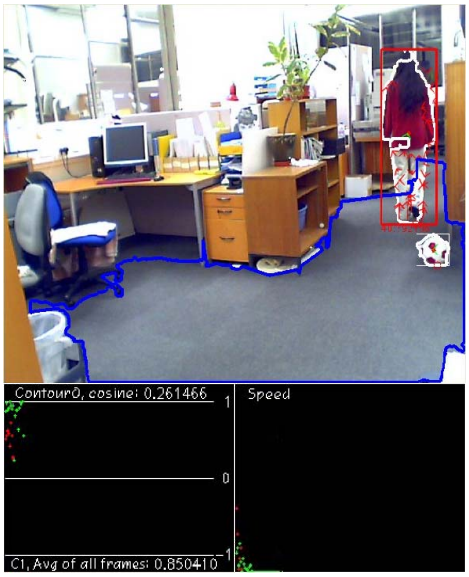
In Fig. 9, a person is passing by a ball and their region's past information, which is averaged cosθ values and speeds, is recorded in graphs in red and green respectively (Fig. 9a). When the regions merge, only the person's region data are kept (Fig. 9b), and when they split, the ball's region data are returned in the graphs (Fig. 9c).



(a)



(b)



(c)

Fig. 9. A merge and a split of multiple objects: (a) regions of a person and a ball and their averaged cos 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. 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/