SLIP AND FALL EVENTS DETECTION BY ANALYZING THE INTEGRATED SPATIOTEMPORAL ENERGY MAP

Tim Liao and Chung-Lin Huang
Electrical Engineering Department,National Tsing-Hua University, Hsin-Chu, Taiwan, ROC
Email: ytliao0423@gmail.com and clhuang@ee.nthu.edu.tw

Abstract

This paper presents a new method to detect slip and fall events by analyzing the *integrated spatiotemporal energy* (*ISTE*) map. *ISTE* map includes motion and time of motion occurrence as our motion feature. The extracted human shape is represented by an ellipse that provides crucial information of human motion activities. We use this features to detect the events in the video with *non-fixed frame rate*. This work assumes that the person lies on the ground with very little motion after the fall accident. Experimental results show that our method is effective for fall and slip detection.

Index Terms—Slip Event Detection, Fall Event Detection, I*ntegrated Spatiotemporal Energy* (*ISTE*) map.

1. Introduction

With the growing population of the elderly and increasing number of people living alone, we need a smart home care system to monitor abnormal activity. The existing wrist communicators and motion detector may easily produce false alarms. Many methods have been proposed using camera sensors and computer vision technology to detect the fall events. Toreyin *et al.* [1] suggested a method for fall detection by make use of an HMM using both video and audio. They used the aspect ratio of the moving object's bounding box. NaitCharif *et al.* [4] proposed a method to track the person using an ellipse and analyze the trajectory to detect inactivity.

In [5, 6], they used the combination of motion history image (MHI) and human shape variation for fall detection. Meng *et al*. [12] also chose the MHI as feature vector and applied the support vector machine (SVM) to classify the motion activity. Foroughi *et al.* [7] modified the MHI and proposed the Integrated Time Motion Image (ITMI) as motion feature, and then applied neural network for detection. MHI and ITMI are not appropriate for the non-fixed frame rate surveillance video. In [11], Foroughi *et al.* proposed another fall detection method by combing the variations of the ellipse around the human silhouette and the head pose for a multi-class SVM to determine the fall event.

Multi-view-based fall detection methods [2, 3, 9, 10] have been proposed to avoid the detection failure of the people falling in a direction "too close" to the optical axis. They classified the silhouette between the standing and lying poses and used HMM to recognize the poses. However, they cannot differentiate the fall from the lying-down. The Ubisense project [8] was developed to classify the human poses by computing the orientation of each detected blob.

Our method can detect the slip events as well the fall events in surveillance video with non-fixed frame rate. Early identification of the slip events can prevent the falls and injuries. We do not use MHI which may indicate large motion energy when people are only walking. Our slip and fall detection methods are based on the motion activity measure and human silhouette shape variation. The former is obtained by analyzing the energy of motion active (MA) area in the so-called *integrated spatiotemporal energy* (*ISTE*) map, whereas the latter is calculated by measuring the shape parameter variation of the human silhouette. Our method consists of four steps: (1) Extract the foreground object silhouettes. (2) Calculate the *ISTE* map and the motion energy variation. (3) Detect a large motion event. (4) Analyze the shape variations to identify the real slip and fall events. The major contributions in this paper are: (a) detect the fall parallel to the optical axis; (b) be applied the non-fixed frame rate video; (c) detect the slip-only event. Early identification of the slip-only event can help prevent the falls and injuries.

2. Foreground Object Extraction

To segment the foreground object from a complex background, we use the background subtraction method. The background model (B) is adaptively updated as

$$B_t = \alpha \bullet B_{t-1} + (1-\alpha) \bullet (\sim I_{at}) \bullet D_t \qquad (1)$$

where the weighting coefficient $0 \le \alpha \le 1$, D_t is the input frame, and I_{at} is the binary people mask that is a binary representation of the object extracted from the image sequence at time *t*. The estimated foreground is defined as

$$I_{at}(i,j) = \begin{cases} 1, & M_t(i,j) \ge I_{th} \\ 0, & M_t(i,j) < I_{th} \end{cases} \qquad (2)$$

where $M_t(i, j) = | V_t(i, j) - B_t(i, j) |$, $V_t(i, j)$ and $B_t(i, j)$ are the input frame and the background image model, respectively. I_{th} is the threshold. After background subtraction, we need to remove noise and obtain a more clean-cut object silhouette by applying morphological filtering and labeling process. We apply closing and opening operations as $I = (I_a \circ B) \bullet B$, where " $\bullet$" and " $\circ$ " are the opening and closing operations. *I* is a foreground object. Figure 1 shows the extracted foreground objects.

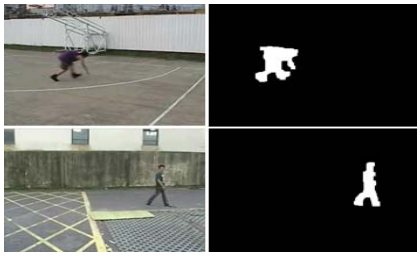


Figure 1. The extracted foreground objects.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.425

3. Motion Information Extraction

A falling event and a slipping event actually cause massive motion. To get the motion information, we quantize the shape variation of the segmented human object. To quantify the motion of the human object, we develop a method to compute a motion activity coefficient C_{ISTE} by analyzing the *integrated spatiotemporal energy* (*ISTE*) map. Before we detect a fall or a split, we detect a large motion based on the coefficient C_{ISTE}.

A. Normalization

To generate the *ISTE* map, we have to normalize the foreground image. Here, we use the bounding box that is described by its centroid, and the length w and h of its width and height as shown in the Figure 2. The bounding box is used to represent human object which is also applied to normalize the human object size for building an *ISTE* map. The normalization process is developed to generate the normalized human object image (*i.e.*, $I_{normalized}(t)$)).

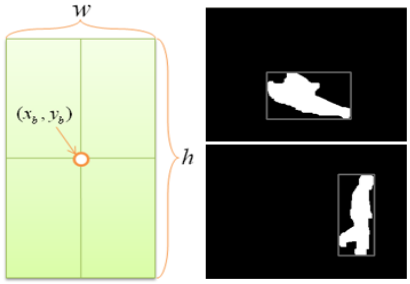


Figure 2. The corresponding bounding box.

B. Integrated Spatiotemporal Energy Map

After normalization, we compute the motion activity coefficient C_{ISTE} by analyzing the *ISTE* map. To build the *ISTE* map, we record T binary frames that are normalized. Each binary normalized frame has been assigned a weight to emphasize the influence of that specific frame. The weight is defined as

$$w(t) = A\exp\left[\frac{-\left(t-\hat{T}\right)^2}{2\sigma^2}\right],\ \hat{T}-T \le t \le \hat{T} \quad (3)$$

The weight function w(t) is an increasing function. As time t increasing, the binary image provides more influential information. We compute the *ISTE* map by summing the weighted normalized energy $E(t)$ as

$$E(t) = I_{normalized}(t) \bullet w(t) \quad (4)$$

where $I_{normalized}(t)$ the normalized human object image. Then we can define $ISTE(\hat{T})$ as

$$ISTE(\hat{T}) = \sum_{t=\hat{T}-T}^{\hat{T}} E(t) \quad (5)$$

As shown in Figure 3, the *ISTE* map illustrates the motion energy. To verify if it is a fall event, we check if the person is immobile on the ground by analyzing the *ISTE* map. In Figure 3(d), we can clearly see that the 3-D contours are similar. When a person is falling, the motion energy of the *ISTE* map may have large changes. In Figure 3, we can see that the 3-D contours are so different that the *MA* area is very small. A large motion activity occurs, the energy of the *ISTE* map suddenly changes, and the *MA* area is small. The motion activity can be analyzed by finding the variations of the *ISTE* coefficients (C_{ISTE}).

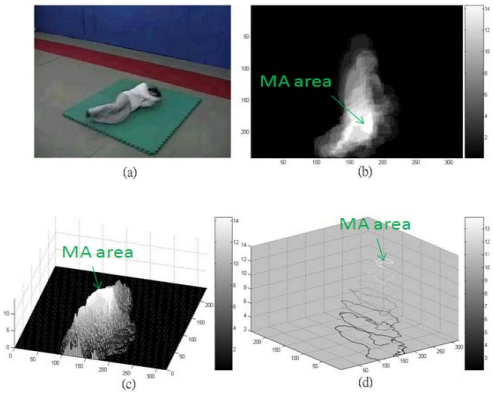


Figure 3. A person falling. (a) Input frame. (b) 2-D and (c) 3-D *ISTE* map. (d) The 3-D contours of the *ISTE* map.

The *MA* area is the overlapped region of the energy images $E(t)$ in the *ISTE* map, and it has the highest energy. The degree of motion activeness is reciprocally proportional to the *MA* area of the *ISTE* map. When people are running, a smaller *MA* area will be found which accumulates large *ISTE* energy. It is because the human posture changes a lot during the active motion so that there is a little overlapped region of $E(t)$. When people are walking, we may find a larger *MA* area, as shown in Figure 4(b). For normal walking or lying on the ground, the shape of the *MA* area is similar to the shape of human object. A falling event and a slipping event crate massive motion. Based on the *MA* area, we can differentiate the *active* motion from the *non-active* one.

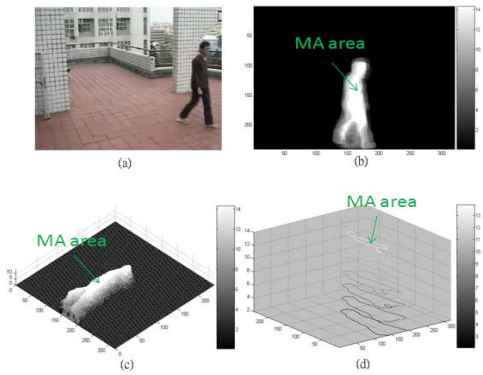


Figure 4. A person walking. (a) Input frame. (b) 2-D and (c) 3-D *ISTE* map. (d) The 3-D contours of the *ISTE* map.

For a person sitting down, the *MA* area of *ISTE* map will be large enough for us to differentiate between the real fall event and a simple lying or sitting down action. To quantify the motion, we define the *ISTE* motion activity coefficient C_{ISTE} at time $t=T$ as

$$C_{ISTE}(t)\big|_{t=T} = \frac{\sum_{(x,y)\in blob} ISTE(x,y,t)\,|_{t=T}}{\sum_{(x,y)\in blob} ISTE(x,y,t)\,|_{t=T}} \times 100\% \quad (6)$$

where the *blob* indicates the *MA* area with the highest energy. We may detect *large motion* using the *ISTE* motion activity coefficient C_{ISTE} as

$$Large\_Motion = \begin{cases} true, & C_{ISTE}(t)|_{t=T} - C_{ISTE}(t-1)|_{t=T} > \varepsilon \\ false & Otherwise \end{cases} \quad (7)$$

where ε is the best threshold which is determined by the Otsu thresholding method. Having detected a massive motion, we continue analyzing the shape.

C. Comparison with Other Motion Measures

However, optical flow is not well-suited for real-time application. Therefore, several motion features have been proposed to represent the motion activity, such as Motion History Image (MHI) [5, 6] as

$$H_\tau(x,y,t) = \begin{cases} \tau, & if \quad I(x,y,t) = 1 \\ Max\{0, H_\tau(x,y,t-1) - 1\}, & otherwise \end{cases} \quad (8)$$

where $I(x, y, t)$ is a binary image obtained from background subtraction, and τ is the maximum duration of accumulating motion activity.

In [5, 6], the motion activity of the object is described by analyzing the *MHI* within the blob. When the object makes a massive motion, $H_\tau(x, y, t)\bullet I(x, y, t)$ is more brightly. In surveillance system, we need to deal with input the video with non-fixed frame rate. The *MHI* suffers from the problem of generating too many false positives. It may detect a large motion when people are walking if the frame rate is 3 *FPS*. To overcome the problem of *non-fixed frame rate*, we use the *ISTE* map to quantify the motion of the human object in the video sequence. *ISTE* maps can quantify the motion of the human object correctly by using the video sequence with low frame rate.

4. Human Object Shape Analysis

Having detected a massive motion by using *ISTE* map, we continue analyzing the segmented human shape as shown in Figure 5. The shape parameters can be used to detect a fall and a slip event. A bounding box can be described the human object with two parameters: width and height. To represent a human object, we can also use an ellipse that is defined by its center, its orientation θ and the length a and b of its major and minor semi-axis. The orientation θ of the ellipse provides more information than other features of the ellipse. The approximated ellipse is more useful than a bounding box for analyzing the human shape.

5. Slip and Fall Event Detection

For moving object, we record the time-varying motion parameter as θ_{mean} and $ratio_{mean}=b_{mean}/a_{mean}$ during building *ISTE* map. To detect a slip in the video, we compute the variation of the orientation σ_θ and the ratio of axes $\sigma_{b/a}$ as

$$\sigma_\theta = \sqrt{(\theta_{mean} - \theta(t))^2}, \quad \sigma_{b/a} = \sqrt{\left(ratio_{mean} - \frac{b(t)}{a(t)}\right)^2} \quad (9)$$

where $\theta_{mean} = \sum_{t=\hat{T}-T}^{\hat{T}} \theta(t) \Big/ \hat{T}$, and $ratio_{mean} = \sum_{t=\hat{T}-T}^{\hat{T}} ratio(t) \Big/ \hat{T}$.

$\hat{T}$ is a period of time of determination. If the following condition is satisfied, then a slip will be detected.

$$Slip = \begin{cases} true, & \sigma_\theta > 45^\circ \; or \; \sigma_{b/a} > 0.3 \\ false, & otherwise \end{cases} \quad (10)$$

After a slip is detected, the human shape is further analyzed for fall detection. The orientation standard deviation v_θ and the ratio standard deviation $v_{b/a}$ of the ellipse as

$$v_\theta = \sqrt{(\theta(t) - \theta(t-1))^2}, \quad v_{b/a} = \sqrt{(ratio(t) - ratio(t-1))^2} \quad (11)$$

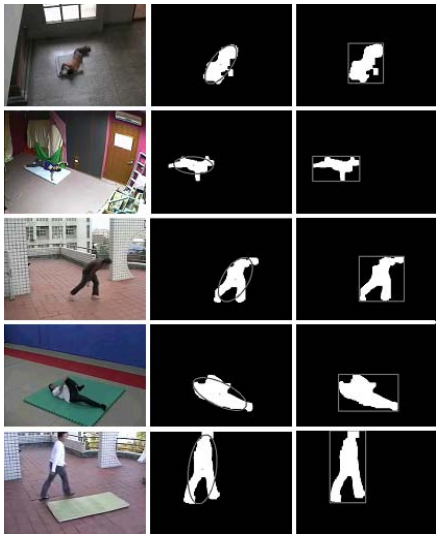
where $ratio=b/a$, is the proportion of the major semi-axis a and the minor semi-axis b. The displacement of the ellipse center $(\bar{x}, \bar{y})$ is

$$v_{(x,y)} = \sqrt{(\bar{x}(t) - \bar{x}(t-1))^2 + (\bar{y}(t) - \bar{y}(t-1))^2} \quad (12)$$

When the human object falls perpendicularly to the camera optical axis, the orientation θ changes significantly and v_θ is large. On the other hand, if the ratio b/a changes and $v_{b/a}$ is large, then human object falls parallel to the camera optical axis.

A possible fall is detected if $v_\theta > 15^\circ$, or $v_{b/a} > 0.3$, or $v_{(x, y)} > 20$ pixels. Once a possible fall is detected, the following verification is required to make sure if the human object is immobile on the ground. We confirm the fall when a stationary object lasts for more than 5 seconds after the possible fall detection occurs. If the object still continues to move during these 5 seconds, then it cannot be a fall. There are three criteria to detect a stationary ellipse:

(1) Few motions in the *blob* of the human object: $C_{ISTE}(t)\big|_{t=T} - C_{ISTE}(t-1)\big|_{t=T} \le \varepsilon$.
(2) An unmoving centroid: $v_{(x, y)} \le 5$.
(3) An unmoving shape: $v_{b/a} \le 0.1$ or $v_\theta \le 5^\circ$.



(a) (b) (c)

Figure 5. (a) Input images. (b) The approximated bounding ellipses. (c) The bounding boxes.

6. Experimental Results

The testing videos are captured by an ordinary surveillance camera with the image resolution of 320×240×24 bits. Here, we define the measurement of accuracy of the system as

$$CT_{slip} = \frac{\| Gth_{n\_slip} - Det_{n\_slip} \|}{Gth_{n\_slip}} \times 100\% \qquad (13)$$

$$CT_{fall} = \frac{\| Gth_{n\_fall} - Det_{n\_fall} \|}{Gth_{n\_fall}} \times 100\%$$

where Gth_{n_slipl} is the ground truth of the number of human object slipping in the video sequence, Det_{n_slip} is the detected number of human object slipping in the video sequence, Gth_{n_fall} is the ground truth of the number of falling events in the video sequence, Det_{n_fall} is the detected number of falling events in the video sequence. Our experiments results are demonstrated in the indoor and the outdoor scenes. For each scene, the illuminate condition varies, so the threshold of the background subtraction and the parameters of the smoothing filter are adjusted. The total number of testing frames is about 2.74×10^6, and the videos are in four main different testing scenarios such as

a) Object moving inside the scene continuously. The motion activity coefficient C_{ISTE} and the status can be calculated. A large motion is detected when the peak of the variation of C_{ISTE} is larger than a threshold. After we detect a slip event, we continue to detect a fall event. So a slip event is included in a fall event.

b) Object moving out of the scene. The foreground region decreases, which indicates the object moving out off the view. In this case, C_{ISTE} has large variation. Large motion does not generate the false alarm because there is no complete human shape to be analyzed. There are no false alarms of miss-identifying the slip events or fall events.

c) Camera is not stationary. The foreground object is not well segmented when the camera is moved by a blast wind that makes the foreground subtraction fail and segmented foreground object is not complete. In this case, C_{ISTE} curve has a sharp variety. However, it may detect a false alarm as the second slip event. After the background model is updated, the accuracy rate can be improved.

d) Object sitting down. Our system can differentiate between the real fall event and the other event of lying or sitting down. The motion activity coefficient C_{ISTE} changes little when a person is sitting down. The *ISTE* map can used to quantify the motion of the human object, so the false alarm can be avoided for a simple sitting down motion.

In our experiments, once the illumination of scene change severely, the segmented foreground object may not be accurate, and the event detection may not work correctly. Table 1 shows the overall event detection accuracy rate. We use 24 videos with 135'27" at many different scenes and different testing scenarios.

7. Conclusions

We have proposed a method to detect a slip event and a fall event by computing *integrated spatiotemporal energy* (*ISTE*) *map* as our motion feature and analyzing the shape of human object. The human shape that is fitted by an ellipse providing crucial information of human motion activities. We use this features to detect the slip and fall events for a surveillance system with *non-fixed* frame rate video in four different real scenarios.

Table 1. The experimental results.

	Slip Event #	Fall Event #
Ground Truth	145	90
Detected Events	139	82
Accuracy Rate	95.86%	91.11%
CT_{slip} & CT_{fall}	CT_{slip}=4.14%	CT_{fall} =8.89%

References

[1] B. Toreyin, Y. Dedeoglu, and A. Cetin. "HMM based falling person detection using both audio and video" *IEEE Int. Workshop on Human-Computer Interaction, Beijing, China, 2005.*

[2] R. Cucchiara, A. Prati, and R. Vezzani "An Intelligent Survillance System for Dangerous Situation Detection in home Environments" *Intelligenza artificable, vol. 1, n. 1, pp. 11-15, 2004.*

[3] R. R. Cucchiara, H. Rita, A. Prati, O. Andrea, R. Vezzani, and C. Roberto,"A Multi-camera vision system for fall detection and alarm generation," *Expert System*, vol. 24, no. 5, pp. 334-345, Nov. 2007.

[4] H. Nait-Charif and S. McKenna "Activity summarization and fall detection in a supportive home environment," *Proc. of the 17th ICPR, vol. 4, pp. 323-326, 2004.*

[5] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, " Fall Detection from Human Shape and Motion History Using Video Surveillance**"** 21st Int. Conf. on *Advanced Information Networking and Applications Workshops, Vol. 2, pp.875 – 880, May 2007.*

[6] T. Ogata, J. Tan, and S. Ishikawa "High-Speed Human Motion Recognition Based on a Motion History Image and an Eigenspace" *IEICE Transactions on Information and Systems, vol. E89-D, Issue 1, pp. 281- 289, January 2006.*

[7] H. Foroughi, N. Aabed, A. Saberi, and H. S. Yazdi "An Eigenspace-Based Approach for Human Fall Detection using Integrated Time Motion Image and Neural Networks" *IEEE Int. Conf. on Signal Processing* (*ICSP*)*, April 2008.*

[8] B. Lo, J. Wang, and G. Yang, "From imaging networks to behavior profiling: Ubiquitous sensing for managed homecare of the elderly," *Pervasive*, 2005.

[9] N. Thome, S. Miguet, and S. Ambellouis, "A Real time Multiview Fall Detection System: A LHMM-Based Approach," *IEEE Trans. on CAS for VT*, Vol. 18, No.11, Nov. 2008.

[10] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, "Probabilistic Posture Classification for Human-Behavior Analysis," *IEEE Trans. on SMC*, Part A, vol. 35, no.1, 2005.

[11] H. Foroughi, A. Rezvanian, and A. Paziraee, "Robust Fall Detection using Human Shape and Multi-class Support Vector Machine," *Sixth Indian Conf. on Computer Vision, Graphics and Image Processing*, 2008.

[12] H. Meng, N. Pears, and C. Bailey, "Recognizing Human Actions based on Motion Information and SVM," *2nd IET Int. Conf. on Intelligent Environments*, Athens, Greece, 2006.