Slip and Fall Detection using Spatiotemporal Characteristics of Human Object for Video Surveillance System

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 method to detect a slip event and a fall event by computing *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 no or little motion after the fall accident. For indoor and the outdoor scenes with different illuminate conditions, we need to adjust the threshold of the background subtraction and the parameters of the smoothing filter independently. Experimental results show that our method is effective for fall and slip detection. The total number of testing frames is about 2.7×10^6, and we use an Intel Core2 Duo 1.8GHz CPU on a Microsoft Windows XP operating system.**

***Index Terms*—Slip Event Detection, Fall Event Detection, *Integrated Spatiotemporal Energy* (*ISTE*) map, Shape Analysis .**

I. INTRODUCTION

Nowadays, there are the growing populations of seniors who require special aid in case of emergency. For home alone people, they need a secure environment and improvement of their quality of life. Visual surveillance has a wide range of potential applications such as security surveillance system, traffic monitoring, and home care system. Surveillance companies have been considering developing an automatic system to monitor the abnormal activities. This work develops an automatic security surveillance system to detect a fall and a slip event in the video. Our method can deal with input video with non-fixed frame rate.

The human object is firstly extracted by background subtraction with an adaptive background model. To quantify the motion of the human object, the motion history image (MHI) has been used; however it generates too many false positives. It may indicate large motion when people are only walking. We define the usual events as the high recurrences of postures that are similar, and the unusual events possess large motion energy. The *integrated spatiotemporal energy* (*ISTE*) map is developed to detect inflected postures as our motion feature factor. The shape of human object is enclosed by an ellipse after large motion is detected in the video. Finally, the shape parameters of an ellipse are analyzed by fall detector and slip detector to identify the abnormal events.

Some systems [1, 2] use a direct approach based on the analysis of a detailed human body model. Most of these systems are not robust, once losing tracking; they often need to retrieve information from every human body parts. In [3], Brand *et al*. has demonstrated how to monitor the office activity with hidden Markov models (HMM). They used entropic based HMMs to acquire models and analyze human activities.

In [4], Zhong *et al.* presented an unsupervised method for detecting the unusual activity by using a combination of many simple features, e.g., analysis of the co-occurrence between the video clips and motion of the moving objects, without building the models of usual activities. Derek *et al*. [5] extract features from a silhouettes and outline a procedure for training and identifying the activity. Using a method similar to the document-keyword analysis, they divided the video into equal length segments and classified the extracted features into prototypes, from which a prototype-segment co-occurrence matrix was calculated and used to determine the

abnormal activity.

Another method [6, 7] analyzes the aspect ratio of the moving object's bounding box representing the person in a single image. The works in [8] used the normalized vertical and horizontal projection of segmented object as feature vectors. Sixsimith et al. [16] use infrared sensor and classify falls by using a neural network and the 2D vertical velocity of the person. Their systems can not differentiate between real fall incident and an event when the person is simply lying or sitting down abruptly. This mistake is made due to either inappropriate feature extraction or poor classification.

NaitCharif *et al*.[9] track the person using an ellipse, and analyze the resulting trajectory to detect inactivity outside the normal zones like chairs and sofas. In [12], Homa etal. used the Integrated Time Motion Image (ITMI) as motion feature. They used the eigenspace technique for motion feature reduction. The main supposition behind this procedure is that the behavior space has a lower dimension than the image space and that the recognition of the behaviors can be performed in this reduced space [13]. For each activity, which has to be recognized, a solitary neural network has been used. So for recognition of *n* activities, *n* neural networks exit. However, the ITMI is an inappropriate feature in a realistic surveillance system. When human people move from left to right or from right to left in the video frame horizontally, the ITMI can be used to record exactly the motion information.

This work presents a method for tracking human objects and measuring the posture-energy by using *ISTE* map which is different from [10]. Our method detects a fall and a slip by assembling the observed information. There are four main steps in our method illustrated as follows:

(1) Extract the foreground object from frames in video.

(2) Calculate *ISTE* map from the foreground image.

(3) Approximate the human object by an ellipse using moments [11].

(4) Analyze the parameters of the ellipse to detect a slip and a fall in the frame.

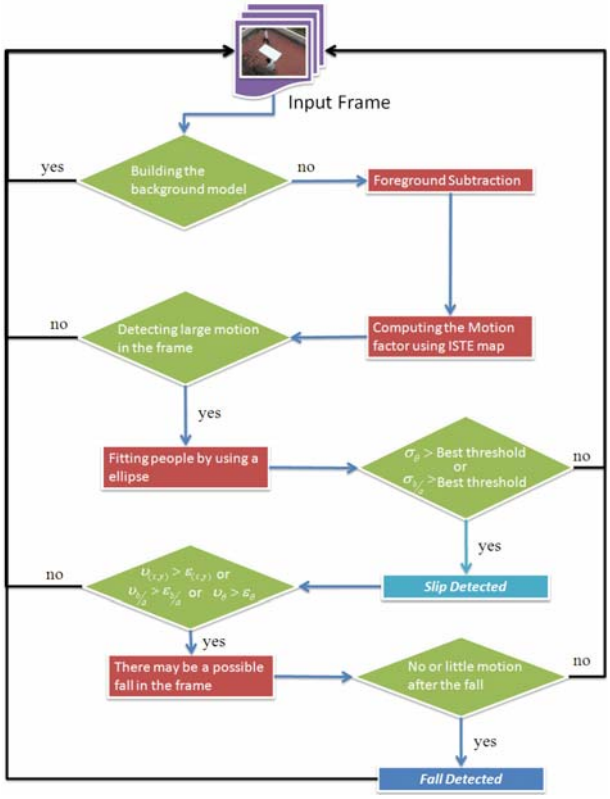

Figure 1. System overview.

II. FOREGROUND EXTRACTION

First, we introduce how to extract the foreground objects in each frame. We assume that the region of interesting (ROI) has sufficient illumination, so that we can use RGB color feature of each pixel for background models training.

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 (α) is a real number between 0 and 1, D_t is the input frame at time t, and I_{at} is the binary people mask that is a binary representation of the people extracted from the image sequence at time t. It is defined by subtracting the '*brightness'* within input image sequences. The estimated stationary background model is defined as

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

$$M_t(i,j) = \| B_t(i,j) - V_t(i,j) \| \quad (3)$$

where $V_t(i,j)$ and $B(i,j)$ are the '*brightness*' at pixel (i,j) in the input frame and the background model at time t, and M_t is the difference value between a input frame and the estimated stationary background.

After background subtraction, we may find some noise of the extracted foreground image (I_a). Therefore, it is very troublesome for objects extraction, and object tracking. To remove noise and obtain a more clean-cut object silhouette, we apply morphological filtering and labeling process.

To remove noises from $I_a(t)$, closing and opening operation are used as

$$I = (I_a \circ B) \bullet B \quad (4)$$

where $\bullet$ and $\circ$ are the opening and closing operations. I is a foreground image indicating the human object. Figure 2 shows some foregrounds after removing noises.

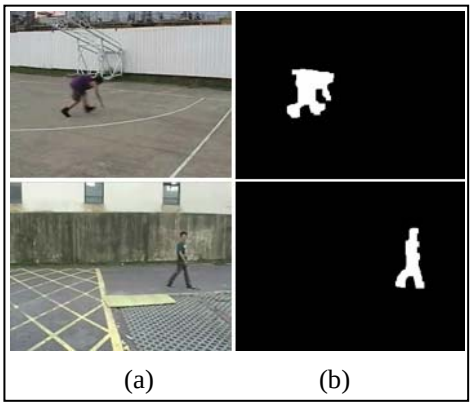

Figure 2. The results of foregrounds after removing noises. (a) input frame (b) results

Once the foreground regions are separated from the background, we need to label the disconnect regions for different objects. The purpose of labeling is to assign the same value to connected component, no matter the component is large or small. After labeling, all the connected components are assigned the same label. We can calculate the component size and sorting these components according to their sizes.

III. MOTION INFORMATION

A falling event and a slipping event actually cause massive motion in video. Therefore, we may take advantage of the motion information of the video sequence. Here, we describe how to obtain the features based on detecting a large motion. To get the motion information, we compute the quantity of shape variation of the posture 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} using the *integrated spatiotemporal energy* (*ISTE*) map. Before we detect a fall or a split, we detect a large motion by analyzing the coefficient C_{ISTE} .

A. Normalization

To compute the *ISTE* map, we have to normalize the foreground image $I(x,y)$. Here, we use the bounding box that is defined by its center (x_b, y_b), and the length w and h of its width and height as shown in the Figure 3(a). The bounding box represents the human object as shown in the Figure 3(b). Then, we normalize the bounding box and build an *ISTE* map.

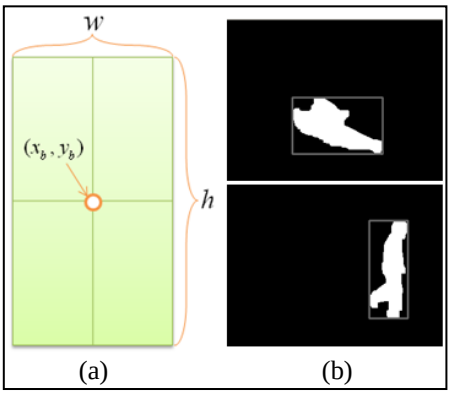

Figure 3. (a) The bounding box is defined (b) The results of the corresponding bounding box.

There are two main steps in the normalization process that can be used to generate the normalized human object $I_{normalized}(t)$ image at time t:

(1) Set x_b as the same coordinate x_{new} .

(2) Normalize the size of the bounding box of each frame, when the width is smaller than one third of the height.

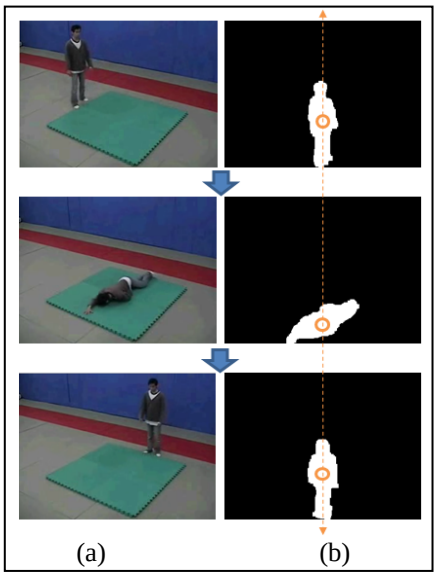

Figure 4. (a) An input frame (b) Put the center of the bounding box to the new coordinate. The circles represent the new coordinates of the centers of the bounding boxes.

B. Integrated Spatiotemporal Energy Map

After normalization, we compute the motion activity coefficient C_{ISTE} by using the *ISTE* map. To build the *ISTE* map, we record T binary frames that are normalized as shown in Figure 4(b). Each binary normalized frame has been assigned a weight to emphasize the influence of that specific frame. The weight is defined as

$$w(t) = Ae^{\frac{-(t-\hat{T})^2}{2\sigma^2}}, \quad \hat{T}-T \le t \le \hat{T} \qquad (5)$$

Because the weighting function $w(t)$ is an increasing function, we can see that as time increasing, the binary images provide more influential information. We can define the *ISTE* map obtained by summing the energy images E as shown in the following:

$$E(t) = I_{normalized}(t) \bullet w(t) \qquad (6)$$

$$ISTE(\hat{T}) = \sum_{t=\hat{T}-T}^{\hat{T}} E(t) \qquad (7)$$

where $I_{normalized}(t)$ and $ISTE(\hat{T})$ is the normalized human object image at time t and the *ISTE* map at time $\hat{T}$ respectively. Besides, the most active (MA) area has the highest energy in an *ISTE* map shown in Figure 5:

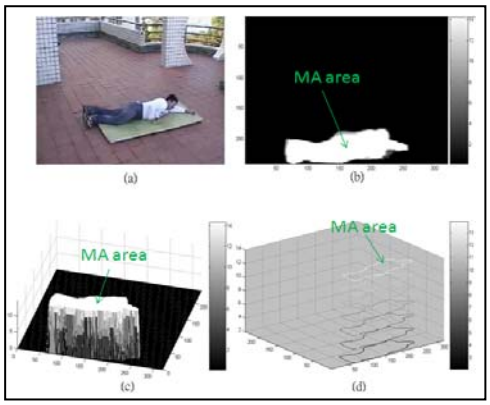

Figure 5. A person is on the ground without moving. (a) An input frame. (b), (c) The *ISTE* map. (d) The contour line of the ISTE map.

As shown in Figure 5, the *ISTE* map indicates a person laying on the ground. To verify a fall event, we may check if the person is immobile on the ground. We detect little motion if C_{ISTE} changes slowly. The energy of an *ISTE* map is high when people do not move as shown in Figure 5(c), and there are a few low-energy regions in an ISTE map.

When people suddenly change their postures, the motion energy of the *ISTE* map may also change as shown in Figure 6. In Figure 6(d), we can clearly see that the contour lines are different and the MA area is very small. A large amount of motion activity occurs when the energy of the *ISTE* map suddenly changes. Besides, the MA area is smaller than the other regions.

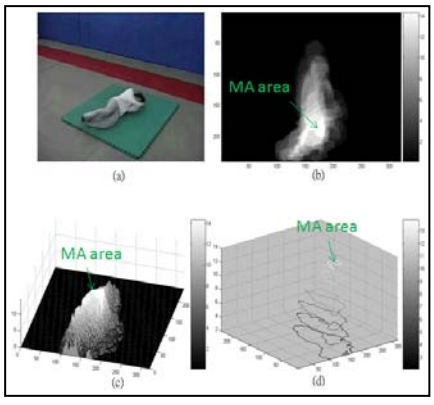

Figure 6. A person is falling. (a) An input frame. (b), (c) The *ISTE* map. (d) The contour line of the *ISTE* map.

The MA area of the *ISTE* map is the overlapped region of the energy images $E(t)$, so it has the highest energy. For people with normal walking or on the ground, the shape of the MA area is similar

to the shape of human object. The above results show that motion activity of the human object is reciprocally proportional to the MA area of the *ISTE* map.

When people are running, the MA area will accumulate large *ISTE* energy as shown in Figure 7(b) and Figure 8(b). The part of the legs in the *ISTE* map has small energy because the posture of the legs changes a lot as people are running. In our system, we assume that a falling event and a slipping event crate massive motion. To avoid the false alarm, we should quantify the motion activity as running or small motion. However, the *ISTE* map can be used to quantify motion accurately such as people running as shown in Figure 7 and Figure 8.

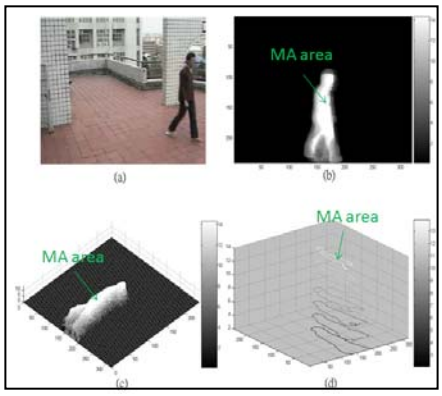

Figure 7. A person is walking. (a) An input frame. (b), (c) The *ISTE* map. (d) The contour line of the *ISTE* map.

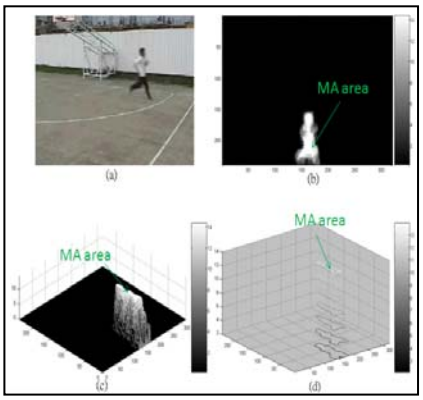

Figure 8. A person is running. (a) An input frame. (b), (c) The *ISTE* map. (d) The contour line of the *ISTE* map.

In Figure 9, a person is sitting down. The MA area of *ISTE* map is large enough to be used to identify the people with small motion. So our system can differentiate between the real fall incident and an event when the person is simply lying or sitting down abruptly.

To quantify the motion, we define the *ISTE* motion activity coefficient C_{ISTE} at time $t = T$ consisting of the characteristic of the *ISTE map* as

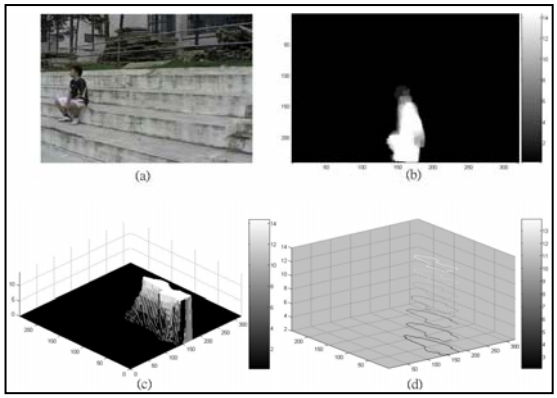

Figure 9. A person is sitting down. (a) An input frame. (b), (c) The *ISTE* map. (d) The contour line of the *ISTE* map.

$$C_{ISTE}(t)\big|_{t=T} = \frac{\sum_{(x,y)\notin blob} ISTE(x,y,t)\,|_{t=T}}{\sum_{(x,y)\in blob} ISTE(x,y,t)\,|_{t=T}} \times 100\% \quad (8)$$

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

$$large \quad Motion = \begin{cases} true, & C_{ISTE}(t)\,|_{t=T} - C_{ISTE}(t-1)\,|_{t=T} > \varepsilon \\ false, & otherwise \end{cases} \quad (9)$$

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

Optical flow [15] is commonly used to detect motion in a video sequence. However, optical flow is not well-suited for real-time application. Therefore, several motion features have been proposed to represent the motion activity in one frame. Motion history image (MHI) as shown in Figure 10 has been proposed as

$$H_\tau(x,y,t) = \begin{cases} \tau, & if \quad I(x,y,t) = 1 \\ Max\{0, H_\tau(x,y,t-1) - 1\}, & otherwise \end{cases} \quad (10)$$

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

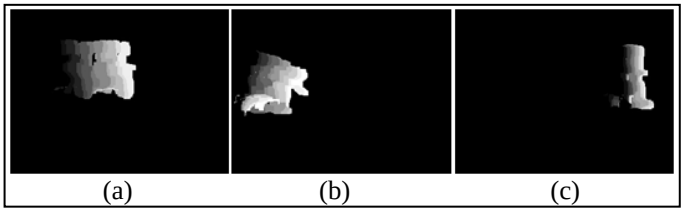

Figure 10. MHI (a) running (b) slipping (c) walking.

In [10], Rougier *et al.* quantify the motion activity of the object by analyzing the *MHI* within the blob as

$$motion \quad of \quad person = \frac{\sum_{x,y} H_\tau(x,y,t) \bullet I(x,y,t)}{\#(pixels \in blob)} \quad (11)$$

where $H_\tau(x,y,t) \bullet I(x,y,t)$ is shown in the bottom row of Figure 11. When the object makes a massive motion, $H_\tau(x,y,t) \bullet I(x,y,t)$ is more brightly. For a 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. For example, it may detect a large motion when people are walking as shown in Figure 12. The frame rate in the Figures 12 and 13 is 3 FPS.

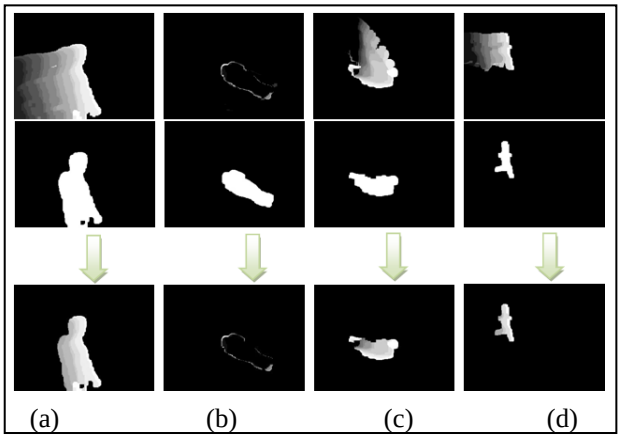

Figure 11. Top row images indicate MHI. Middle row images represent the blob. Bottom row images represent $H_\tau(x,y,t) \bullet I(x,y,t)$. Column (a) walking. Column (b) a person is on the ground without moving. Column (c) falling. Column (d) running.

To overcome the problem of floating frame rate, we use the *ISTE* map to quantify the motion of the human object in the video sequence. In Figure 13, *ISTE* maps can quantify the motion of the human object correctly by using the video sequence with low frame rate.

IV. HUMAN OBJECT SHAPE ANALYSIS

Having detected a massive motion by using *ISTE* map, we continue analyzing the segmented human shape. The human shape is approximated by an ellipse based on the moment analysis [11]. An ellipse is defined by five different shape parameters as its center $(\bar{x}, \bar{y})$, its orientation θ and the length a and b of its major and minor semi-axis. The shape parameters of ellipse can be used to detect a fall and a slip.

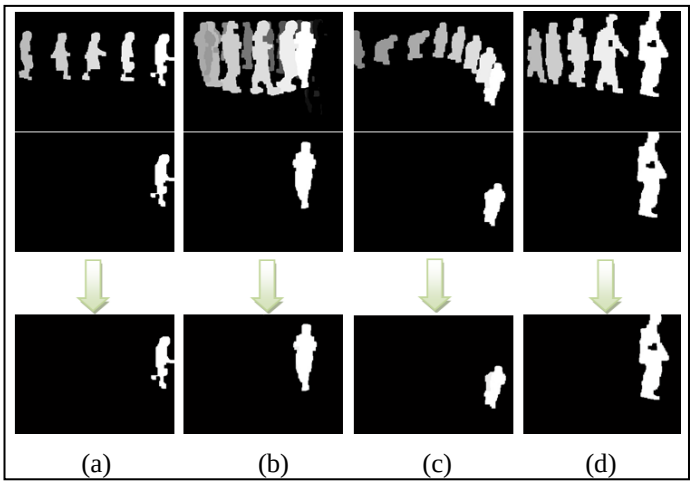

Figure 12. Motion of the three results are bigger than 90%. However, people are just walking. Video with 3FPS. (a) running ,100%. (b) walking, 93.33%. (c) walking, 89.86%. (d) walking, 90.2%.

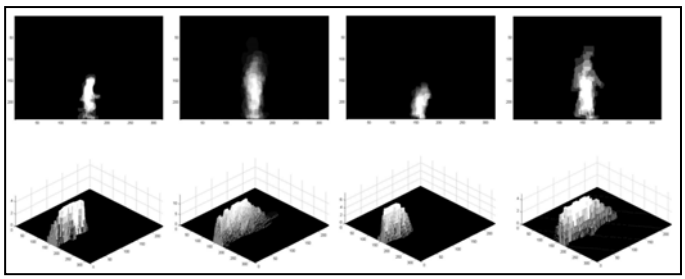

Figure 13. *ISTE* map results of the Figure 12.

A. The center of the ellipse

With a binary human object image $I(x,y)$, the moment of order $(p+q)$ is defined as

$$m_{pq} = \sum_x \sum_y x^p y^q I(x,y) \quad (12)$$

where $p, q = 0, 1, 2....$ The center of the ellipse is obtained by computing the coordinates of the centroid with the first and zero-order spatial moments as

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad , \quad \bar{y} = \frac{m_{01}}{m_{00}} \tag{13}$$

where $(\bar{x}, \bar{y})$ is the centroid of the human object. The centroid $(\bar{x}, \bar{y})$ is used to compute the central moments as follows:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x, y) \tag{14}$$

B. The angle of the ellipse

Here, we compute the angle between the major axis of the human object and the horizontal axis $y = 0$. The angle gives the orientation of the ellipse, which can be computed by calculating the central moments of the second order as follows:

$$\theta = \frac{1}{2} \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right) \tag{15}$$

C. The semi-axes of the ellipse

To compute the major semi-axis a and the minor semi-axis b of the ellipse, we firstly compute $I_{\min}$ and $I_{\max}$, that are the least and the greatest moments of inertia. However, they can be computed by evaluating the eigenvalues of the co-variance matrix in [11] as follows:

$$J = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix} \tag{16}$$

The eigenvalues $I_{\min}$ and $I_{\max}$ of the co-variance matrix are given as

$$I_{\min} = \frac{\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2} \tag{17}$$

$$I_{\max} = \frac{\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2} \tag{18}$$

In [14], they computed the major semi-axis a and the minor semi-axis b of the best fitting ellipse as

$$a = \left(\frac{4}{\pi}\right)^{\frac{1}{4}} \left[\frac{(I_{\max})^3}{I_{\min}} \right]^{\frac{1}{8}}, b = \left(\frac{4}{\pi}\right)^{\frac{1}{4}} \left[\frac{(I_{\min})^3}{I_{\max}} \right]^{\frac{1}{8}} \tag{19}$$

A bounding box can be described with two parameters: width and height, which is used to represent a human object as shown in Figure 14. To represent a human object, we can also use an ellipse that is defined by its center $(\bar{x}, \bar{y})$, 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.

Figure 14. (a) An input frame (b) approximated ellipse (c) bounding box.

V. SLIP AND FALL 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

$$\theta_{mean}=\frac{\sum_{t=\hat{T}-T}^{\hat{T}}\theta(t)}{\hat{T}},\ ratio_{mean}=\frac{\sum_{t=\hat{T}-T}^{\hat{T}}ratio(t)}{\hat{T}} \quad (20)$$

$$\sigma_\theta=\sqrt{\left(\theta_{mean}-\theta(t)\right)^2},\sigma_{b/a}=\sqrt{\left(ratio_{mean}-\frac{b(t)}{a(t)}\right)^2} \quad (21)$$

If the following condition is satisfied, then a slip will be detected.

$$Slip=\begin{cases}true, & \sigma_\theta>45^\circ\ or\ \sigma_{b/a}>0.3\\ false, & otherwise\end{cases} \quad (22)$$

After a slip is detected, the human shape is further analyzed for fall detection. The orientation standard deviation υ_θ and the ratio standard deviation $\upsilon_{b/a}$ of the ellipse that represent the human object are defined as

$$\begin{aligned}\upsilon_\theta&=\sqrt{\left(\theta(t)-\theta(t-1)\right)^2},\\ \upsilon_{b/a}&=\sqrt{\left(ratio(t)-ratio(t-1)\right)^2}\end{aligned} \quad (23)$$

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

$$\upsilon_{(x.y)}=\sqrt{\left(\bar{x}(t)-\bar{x}(t-1)\right)^2+\left(\bar{y}(t)-\bar{y}(t-1)\right)^2} \quad (24)$$

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

A possible fall is detected if υ_θ >15°, or if $\upsilon_{b/a}$ >0.3, or if $\upsilon_{(x,y)}$ >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 : $C_{ISTE}(t)|_{t=T}-C_{ISTE}(t-1)|_{t=T}\le\varepsilon$

(2) An unmoving centroid : $\upsilon_{(x.y)}\le 5 pixels$

(3) An unmoving shape : $\upsilon_{b/a}\le 0.1$ or $\upsilon_\theta\le 5^\circ$

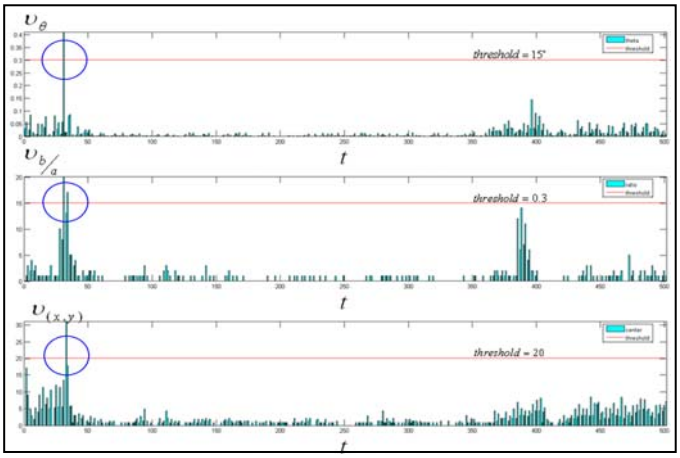

Figure 15. A possible fall.

VI. EXPERIMENTAL RESULTS

The testing videos are captured by a JVC camera. The resolution of color image frame is $320\times240\times24$ bits. We do the experiments using with an Intel Core2 Duo 1.8GHz CPU on a Microsoft Windows XP operating system.

In the experiments, we define the measurement of accuracy of the system as follows:

$$\begin{aligned}CT_{slip}&=\frac{\|Gth_{n\_slip}-Det_{n\_slip}\|}{Gth_{n\_slip}}\times100\%\\ CT_{fall}&=\frac{\|Gth_{n\_fall}-Det_{n\_fall}\|}{Gth_{n\_fall}}\times100\%\\ CT_{avg}&=\frac{CT_{slip}+CT_{fall}}{2}\times100\%\end{aligned} \quad (25)$$

Where Gth_{n_slip} 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 human object falling in the video sequence, Det_{n_fall} is the detected number of human object falling in the video sequence.

Our experiments results are demonstrated in the

indoor and the outdoor scenes as shown in Figure 16. For each scene, the illuminate condition varies, so the threshold of the background subtraction and the parameters of the smoothing filter are adjusted independently. The total number of testing frames is about 2.74×10^6.

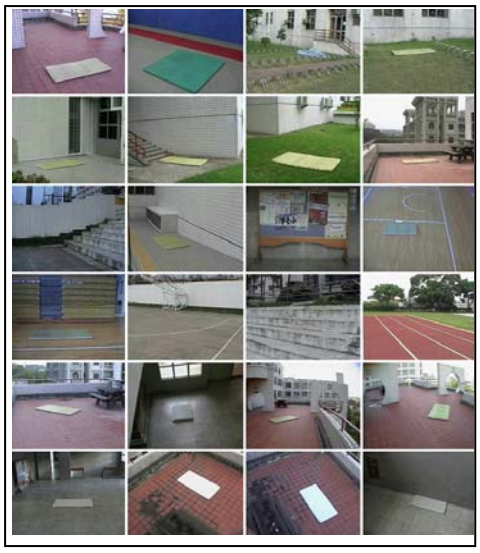

Figure 16. Test videos.

A. Object moving inside the scene continuously

The motion activity coefficient C_{ISTE} and the status are shown in Figure 17-1. A large motion is detected when the peak of the variation of C_{ISTE} is larger than a threshold (indicated by a red line). After we detect a slip event in the video, we continue to detect a fall event. So a slip event is included in a fall event. The variety of the parameters of the ellipse and some results are shown in Figure 17-2.

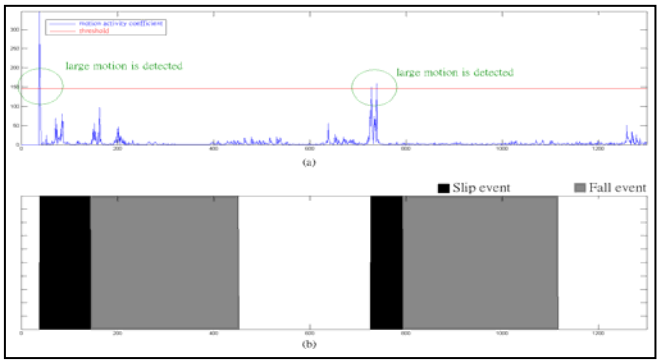

Figure 17-1. (a) C_{ISTE}. (b) The status.

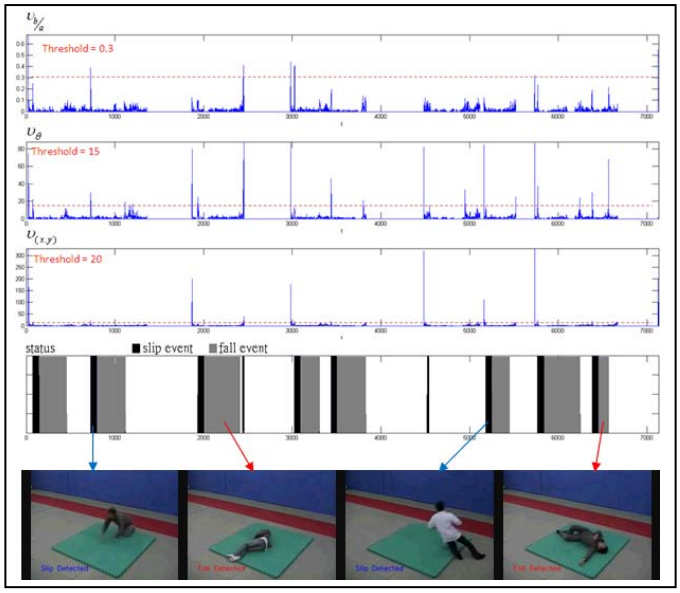

Figure 17-2. Results of a video in the gym.

B. Object moving out of the scene

The foreground region decreases, which indicates the object moving out off the view. So C_{ISTE} has large variation as shown in the Figure 18-1. Large motion in this case do not generate the false alarm in our system because there is no complete human shape to be analyzed. So, the slip detector and the fall detector are not activated. When a person is on the ground without moving, C_{ISTE} changes very little. The variety of the parameters of the ellipse and some results are shown in Figure 18-2.

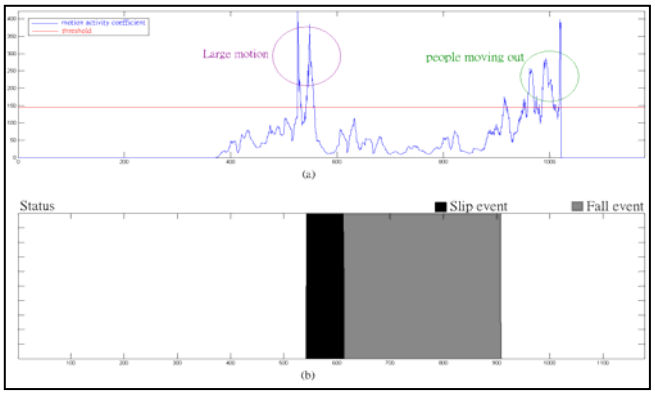

Figure 18-1. (a) C_{ISTE}. (b) The status.

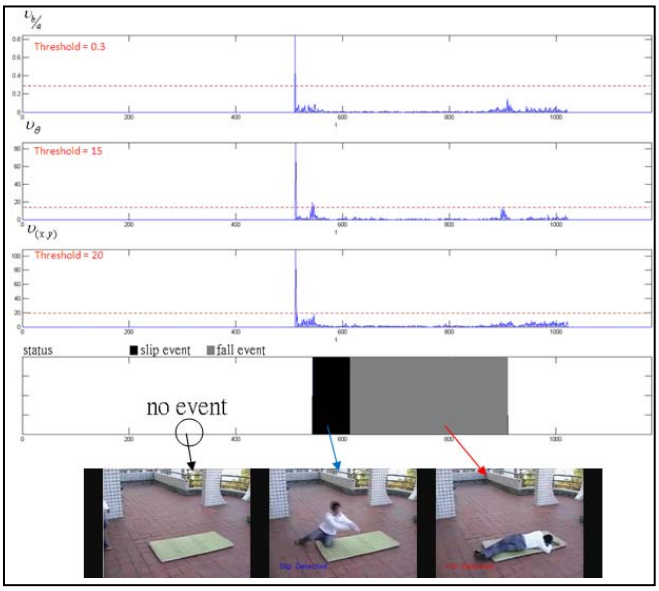

Figure 18-2. Results of a video at the piazza.

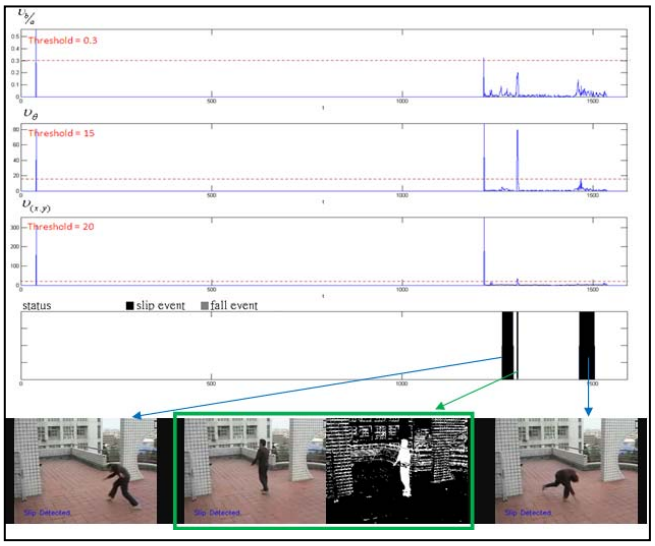

Figure 19-2. Results of a video at the piazza with one false alarm. The worse foreground that makes the false alarm is segmented as the second column.

C. Camera is not stationary

The foreground object is not well segmented when the video is noisy as shown in the Figure 19-2. It occurs when the camera is moved by a blast wind, that makes the foreground subtraction fail. So C_{ISTE} curve has a sharp variety as shown in the Figure 19-1. 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. Human object sitting down

Our system can differentiate between the real fall incident and the event when the person is simply lying or sitting down abruptly as shown in Figure 20-2. The motion activity coefficient C_{ISTE} changes little when a person is sitting down as shown in Figure 20-1. The *ISTE* map can exactly quantify the motion of the human object, so the false alarm can be avoided as people sitting down.

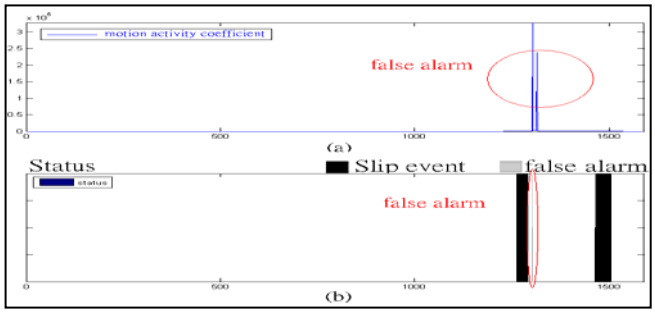

Figure 19-1. (a) C_{ISTE} . (b) The status.

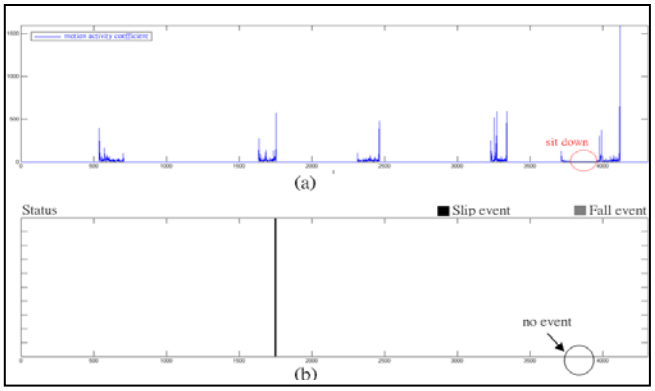

Figure 20-1. (a) C_{ISTE} . (b) The status.

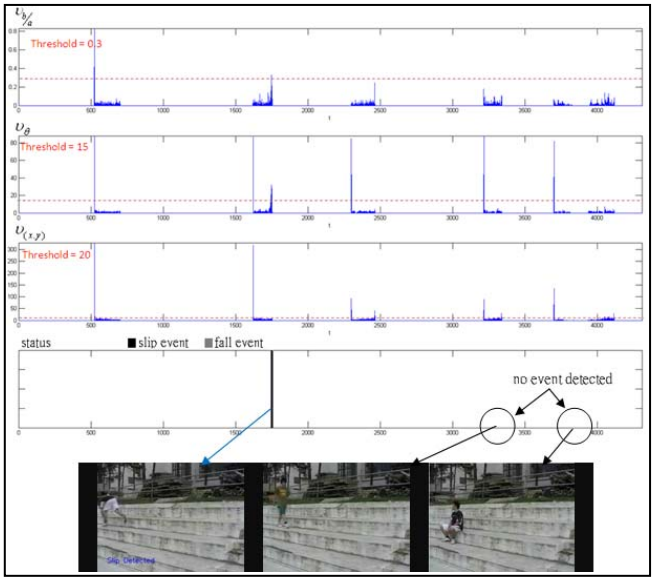

Figure 20-2. Sitting down will not be detected as falling incident.

In our experiment, the illumination of glass and automatic door blinks severely so that the foreground object segmentation may fail. Table 1 shows the overall counting accuracy rate. We use 24 videos with 135’27” at different places as our testing data as shown in Figure 21.

Figure 21. Experimental results .

Table 1. The experimental result of counting number for our testing database.

	Slip Event	Fall Event
Ground Truth	145	90
Counting Number	139	82
Accuracy Rate	95.86%	91.11%
CT_{slip} & CT_{fall}	$CT_{slip} = 4.14\%$	$CT_{fall} = 8.89\%$
CT_{avg}	$CT_{avg} = 6.51\%$	

VII. CONCLUSION AND FUTURE WORK

We 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 measure the events on a realistic surveillance system with non-fixed frame rate.

An unusual scene is further detected in our system. Error detections typically occur when the human object is not well segmented by background subtraction. The unfixed cameras will generate incorrect information of the background model, so the foreground object cannot be segmented accurately. We need to develop a algorithm that can adaptively update the foreground model once the camera is unfixed.

REFERENCE

[1] Haritaoglu, I., Harwood, D., Davis, L.S. “W^4: Real-Time Surveliance of People and Their Activies”. IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 22, no.8, pp. 809-803, August 2000.

[2] Wren, C.R., Azarbyejani, A., Darrel, T., Pentland, A.P. “Pfinder : Real-Time Tracking of the Human Body”. IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 780-785, July 1997.

[3] M. Brand, V. Kettnaker., “Discover and Segmentation of Activities in Video ”. IEEE Trans. On Pattern Analysis and Machine Intelligence, 22(8), p. 844-851, August 2000.

[4] H. Zhong, M. Visontai, J. Shi, “Detecting Unusual Activity in Video” IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 2, pp. 819-826, June 2004.

[5] Anderson, D., Keller, J.M., Skubic, M., Xi Chen, Zhihai He “Recognizing Falls from Silhouettes” IEEE Annual International Conf. on Engineering in Medicine and Biology Society, pp. 6388-6391, August 2006.

[6] B. T oreyin, Y. Dedeoglu, and A. C, etin. “Hmm based falling person detection using both audio and video” In IEEE International Workshop on Human-Computer Interaction, Beijing, China, 2005.

[7] S. -G. Miaou, P.-H. Sung, and C.-Y. Huang “A Customized Human Fall Detection System Using Omni-Camera Images and Personal In-

formation" Proc of Distributed Diagnosis and Home Healthcare (D2H2) Conference, 2006.

[8] R. Cucchiara, A. Pratti, and R. Vezani "An Intelligent Survillance System for Dangerous Situation Detection in home Environments" In Intelligenza artificable, vol. 1, n. 1, pp. 11-15, 2004.

[9] H. Nait-Charif and S. McKenna "Activity summarization and fall detection in a supportive home environment " In Proceedings of the 17th International Conference on Pattern Recognition (ICPR), vol. 4, pp. 323-326, 2004.

[10] Rougier, C., Meunier, J., St-Arnaud, A., and Rousseau, J. " Fall Detection from Human Shape and Motion History Using Video Surveillance" Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference, Vol. 2, pp.875 – 880, May 2007.

[11] W. Pratt "Digital Image Processing" 3rd edition, John Wiley & Sons, New York, 2001.

[12] Homa Foroughi, Aabed Naseri, Alireza Saberi, Hadi Sadoghi Yazdi "An Eigenspace-Based Approach for Human Fall Detection using Integrated Time Motion Image and Neural Networka" In IEEE International Conference on Signal Processing Proceeding, April 2008.

[13] Takehito OGATA, JooKooi TAN Seiji 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.

[14] A. Jain. "Fundamentals of digital image processing" Prentice Hall, Englewood Cliffs, New Jersey, 1989.

[15] J. Barron, D. Fleet, and S. "Beauchemin. Performance of optical flow techniques" International Journal of Computer Vision, 12(1):43-77, February 1994.

[16] A. Sixsmith and N. Johnson. "A smart sensor to detect the falls of the elderly" IEEE Pervasive Computing, 3(2):42-47, April-June 2004.