Automatic Fall Incident Detection in Compressed Video for Intelligent Homecare

Chia-Wen Lin
Department of Electrical Engineering
National Tsing Hua University
Hsinchu 30010, Taiwan
cwlin@ee.nthu.edu.tw

Zhi-Hong Ling
Department of Computer Science
National Chiao Tung University
Hsinchu 30010, Taiwan

Abstract—**This paper presents a compressed-domain fall incident detection scheme for intelligent homecare applications. First, a compressed-domain object segmentation scheme is performed to extract moving objects based on global motion estimation and local motion clustering. After detecting the moving objects, three compressed-domain features of each object are then extracted for identifying and locating fall incidents. The proposed system can differentiate fall-down from squatting by taking into account the event duration. Our experiments show that the proposed method can correctly detect fall incidents in real time.**

Keywords-homecare; compressed-domain processing; video surveillance; fall detetcion

I. INTRODUCTION

Electronic visual surveillance systems are an emerging application field involving multidisciplinary technologies spanning from image/video processing to communication, pattern recognition, and computer vision [1]. The ever-increasing demands on public area monitoring, transportation facilities (subways, highways, tunnels, etc.) monitoring, and indoor monitoring (homecare, home/office security, etc.) have been urging the development and deployment of new-generation visual surveillance systems. New-generation visual surveillance systems can benefit from new advances in digital video communication (video compression, bandwidth reduction, and convenient networking), digital video processing, and broadband access network infrastructures [1][2]. For example, digital video compression allows efficient transmission and recording of video events. Video enhancement algorithms can be used to enhance the quality of video under poor illumination conditions or low-resolution video captured by a low-cost camera. Video streaming and real-time video networking can provide flexible and ubiquitous video monitoring from remote locations. Automatic alarms can be generated and sent through networks or pagers to notify the users of abnormal situations. Research work on advanced video processing techniques for robust video transmission, color-video processing, event-based attention focusing, model-based sequence understanding in surveillance applications has been providing more and more interesting and useful features, thanks to the availability of low-cost high-performance computers, and mobile and fixed multimedia communications. In an intelligent visual surveillance system, it would be very helpful to provide features of automatically detecting and locating unusual events, such as, fall incident detection, intruder detection and tracking, and fire/smoke detection.

Automatically monitoring abnormal activities of the elderly and children using video cameras at home is an important issue for homecare. In the case of elderly people living on their own, there is a particular need for monitoring their behavior, such as a fall, unusual squatting, or a long period of inactivity. Falls amongst the elderly are particularly serious and often lead to injury, restricted activities, fear, or death. It was shown in [3] that 28-34% elderly people in the community experience at least one fall every year, and 40-60% of the falls lead to injury. The main reasons elderly people become bedridden are apoplectic ictus, decrepitude, falls, and fractures [4]. Fall-related injuries have also been among the five most common causes of death amongst the elderly population [5]. The early detections and recording of fall incidents can help the elderly to obtain in-time medical treatments as well as help identify reasons of incidents while sustaining a fall.

Most of the existing fall detection schemes described in [4]-[6] propose to use specially designed sensors and circuitry, which may not be convenient for the elderly to wear or bring all the time. Recently several computer vision based techniques, such as object tracking, behavior understanding and description, personal identification, and event detection, have been developed for visual surveillance and homecare applications due to the wide deployment of low-cost video cameras. A few computer vision based methods [7]-[10] have been proposed for detecting falls or other events at home. In [8], a method was proposed to detect portions of a video which are likely to contain a dynamic event from a compressed video. The events are assumed to happen in discontinuities in a motion field, or nonlinear changes of sizes in a moving region. The method presented in [9] uses an omni-directional camera to track a video object modeled with an ellipse contour using a particle filter. The tracked object trajectory within different regions of a living room is analyzed by temporal segmentation so as to train and annotate the models of different activities using Gaussian mixture models (GMMs). Abnormal events such as falls and unusual inactivity can be classified using the trained GMMs.

Many networked video cameras currently deployed are equipped with a video encoder in order to achieve efficient bandwidth consumption. Their computing power and storage

1-4244-1251-X/07/$25.00 2007 IEEE.

capacity are, however, usually still rather limited due to the cost consideration. Because detecting events such as fall incidents in a video clip usually needs to process a sustained period of video data (e.g., 1~2 seconds), pixel-domain processing would require a large size of frame buffer, leading to prohibitively expensive memory cost and high power consumption. As a result, event detection often needs to be done by using video post-processing in a surveillance control center, in which relatively powerful computers are equipped and videos are stored/received in compressed formats. Compressed-domain processing techniques are efficient in terms of computational complexity and storage cost because they can take advantage of the information already carried in a compressed video bitstream without the need of decoding the compressed video into pixel values, thereby drastically reducing the amount of data to be processed. Should the event detection be performed in a video camera, the camera can also use the information available in the compressed video bitstream (e.g., motion information and coding modes of macroblocks) to reduce the computation for event detection significantly.

In this work, we focus on compressed-domain fall incident detection schemes. The first task for vision-based fall incident detection is to detect human objects. We propose a compressed-domain vision-based fall detection scheme for intelligent homecare applications. The proposed scheme can detect and track moving objects from a compressed video in the compressed domain [10]. After detecting the moving objects, compressed-domain features of each object are then extracted for identifying and locating fall incident.

II. Proposed System Architecture

Fig. 1 shows the conceptual diagram of the proposed intelligent networked visual surveillance system. The control center contains a server which is responsible for receiving compressed video bitstreams from mobile surveillance cameras, recording video data on the storage device, and managing the video access from remote users. The video captured by a camera is compressed using an MPEG-4 encoder, and the compressed video is subsequently sent to the server via UDP packets. Remote users can access the surveillance video data ubiquitously using different multimedia terminal devices through the Internet. An automatic fall-incident detection scheme is implemented in the system for intelligent homecare applications.

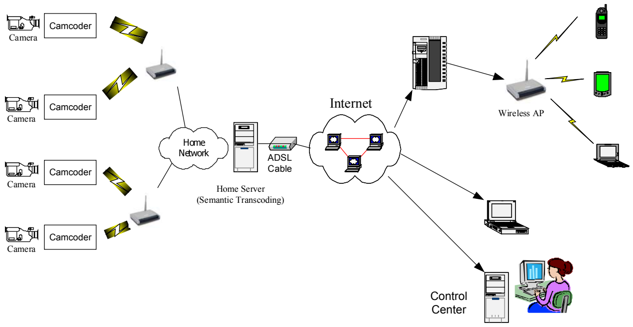


Fig. 1. Block diagram of the proposed compressed-domain fall-down detection scheme.

The flowchart of the proposed compressed-domain fall incident detection scheme is given in Fig. 2. The proposed scheme involves two steps: compressed-domain object extraction and fall-down detection. At first, the MVs and the DC+2AC image [11] of each video frame are extracted from the incoming bitstream for subsequent processing. The MVs extracted from the incoming bitstream are fed into the Global Motion Estimation (GME) module [12] to estimate the global motion (GM) parameters. As a result, the global motion and local object motion(s) are separated, and then those macroblocks with significant local motions are grouped together to obtain a rough object mask.

If the video shot contains global motion, the GM-compensated Change Detection operation is performed to refine the object mask. Otherwise, the Change Detection module is used to refine the object mask. For frames that contain more than a single object, the object clustering operation is performed to separate the object mask into multiple individual object sub-masks. The detailed procedure of compressed-domain object segmentation scheme is described in [10].

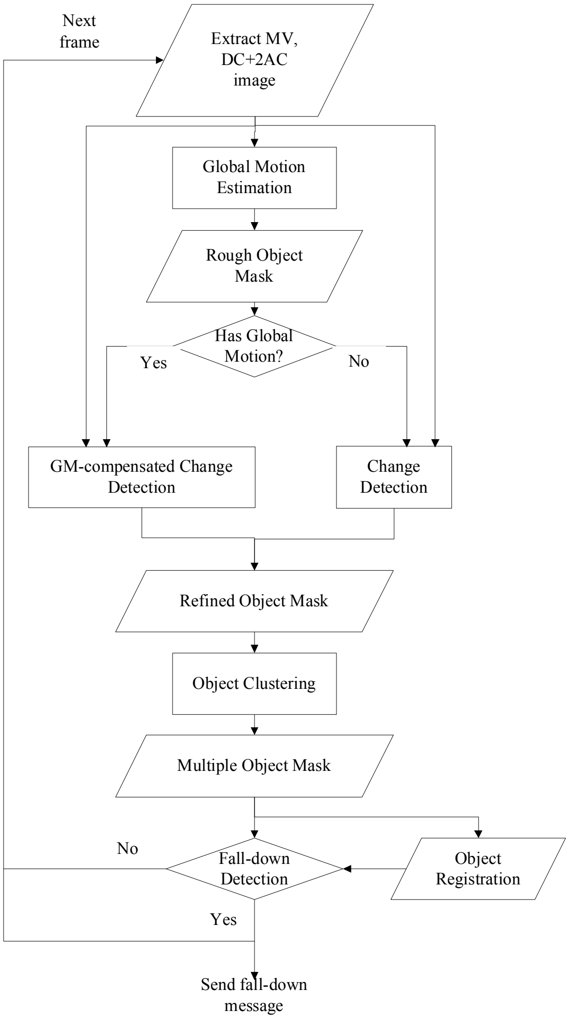
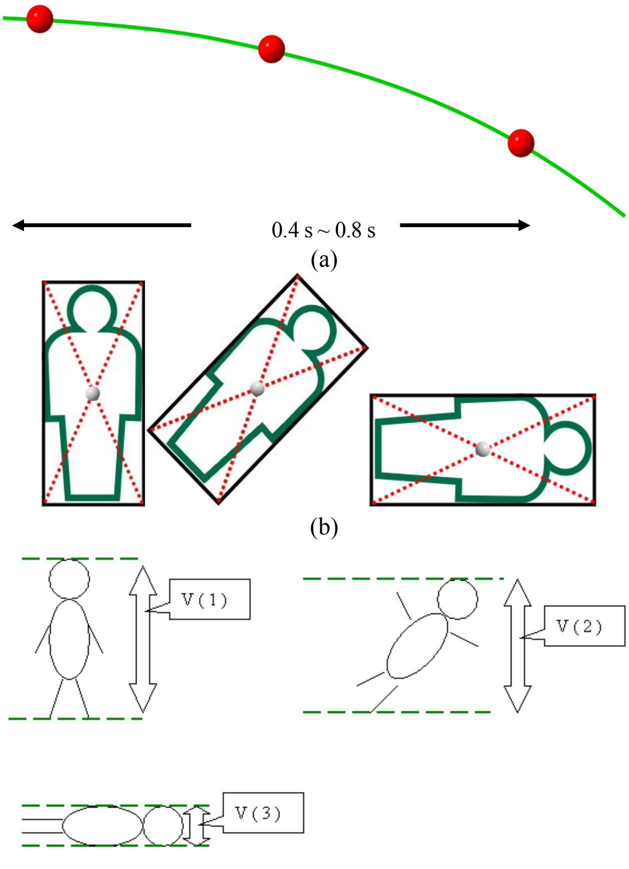


Fig. 2. Block diagram of the proposed compressed-domain fall-down detection scheme.

After extracting the video object, the fall-down detection module uses three feature parameters: the centroid of a human object, the maximum vertical projection histogram value, and the duration of an event to identify and locate fall-down events. Object tracking is activated in our method when the video has more than one object. The Object Labeling module is used to find the correspondence of video objects between two consecutive frames so as to obtain the associated feature parameters of each object.

III. Feature-based Fall Incident Detection

To identify and locate a fall incident of a person, we found that three features can be used to effectively capture fall-down events according to our experiments. First, as illustrated in Fig. 3(a), a fall incident usually occurs in a short time period with a typical range of 0.4s~0.8s. Second, Fig. 3(b) depicts that a falling person's centroid changes significantly and rapidly during the falling period. Third, the vertical projection histogram is also a useful feature for detecting a fall-down event because the vertical projection histogram of a falling person also changes significantly during the falling period as shown in Fig. 3(c).



(a)

(b)

(c)

Fig. 3. Three features used for detecting a fall incident: (a) the duration of an event; (b) The location and change rate of the centroid of the human object; (c) the vertical projection histogram of the human object.

In order to obtain the three feature values: the centroid of a human object, the vertical projection histogram, and the duration of an event detected, the human objects need to be extracted using the proposed compressed-domain segmentation method. After extracting a foreground object, the vertical projection histogram of the object is computed as follows.

$$H(x,y)=\begin{cases}1 & \text{if } (x,y) \text{ is an object pixel}\\ 0 & \text{otherwise}\end{cases} \tag{1}$$

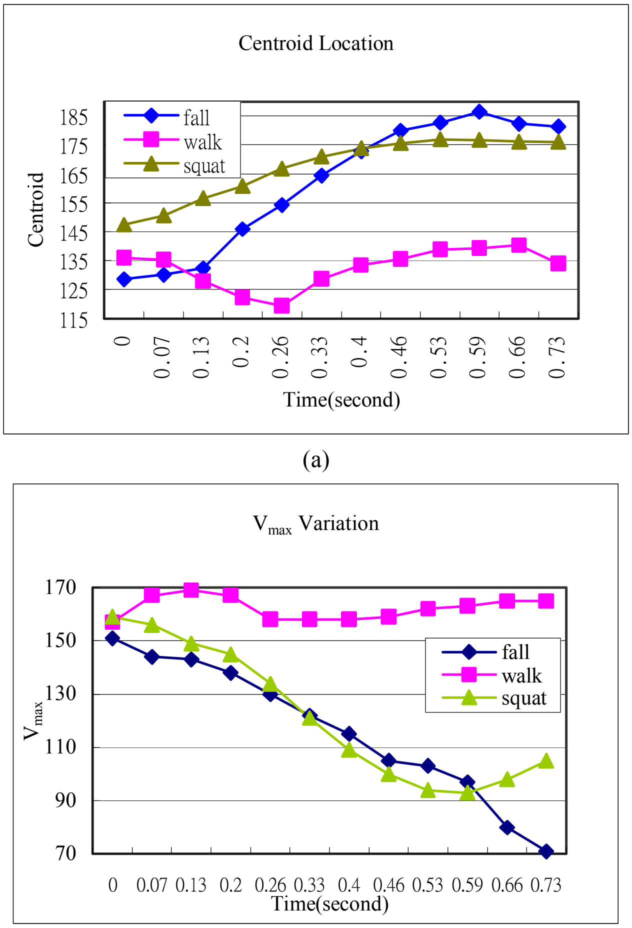
$$V(x)=\sum_{y} H(x,y) \tag{2}$$

Since $V(x)$ in (2) is an one-dimensional distribution, we can use some distance metrics, such as the Bhattacharyya distance [13] in (3), to measure the similarity of two vertical projection histograms (e.g., $V_1(x)$ and $V_2(x)$) of video frames within a sliding time window so as to identify significant changes of vertical projection histogram in contiguous frames due to fall incidents.

$$d(V_1,V_2)=\sum_{x}\sqrt{\frac{V_1(x)}{\sum_{u}V_1(u)}\frac{V_2(x)}{\sum_{v}V_2(v)}} \tag{3}$$

However, the computational complexity of computing (3) is high. To reduce the computation, we propose to use the maximum of a vertical projection histogram defined in (4), which maps the vertical project histogram into a single value.

$$V_{\max}=\max_{x} V(x) \tag{4}$$



(a)

(b)

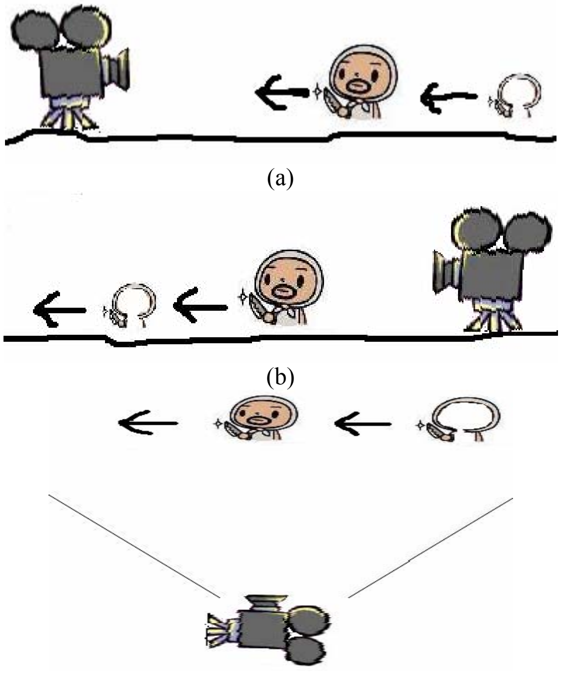
Fig. 4. Comparison of two feature values for a normal-walking person and a falling down person and a squatting person: (a) the centroid locations of objects versus time; (b) the vertical projection histogram values of objects versus time.

Fig. 4 compares the centroid locations and the $V_{\max}$ values of three different cases: walking, squatting, and falling. We can see that both feature values change significantly and repidly during the falling period. In this example, the centroid locations before and after falling down are 128 and 186, respectively. The $V_{\max}$ values before and after falling down are 151 and 70, respectively. The duration of the event is about 0.59 s which is within the typical time range of a fall-down event.

Because the above two feature values may vary with different object locations and object sizes, adopting fixed threshold values are not appropriate. We propose to use the following feature vector consisting of two normalized feature values for fall incident detection. The two feature values also take into account the effect of event duration.

$$\mathbf{x}(n) = \left[\frac{\left| f_{\text{cent}}(n-SW) - f_{\text{cen}}(n) \right|}{f_{\text{cent}}(n-SW)} \quad \frac{\left| V_{\max}(n-SW) - V_{\max}(n) \right|}{V_{\max}(n-SW)} \right]^{T} \quad (5)$$

where $f_{\text{cent}}(n)$ represents the location of the object centroid in the nth frame; $V_{\max}(n)$ denotes the maximum of vertical projection histogram of the object in the nth frame; SW stands for the length of sliding window, which is in the typical range of a fall incident's duration.



(a)

(b)

(c)

Fig. 5. Three different motion types: (a) a person going toward the camera; (b) a person going away from the camera; (c) a person walking horizontally in front of the camera.

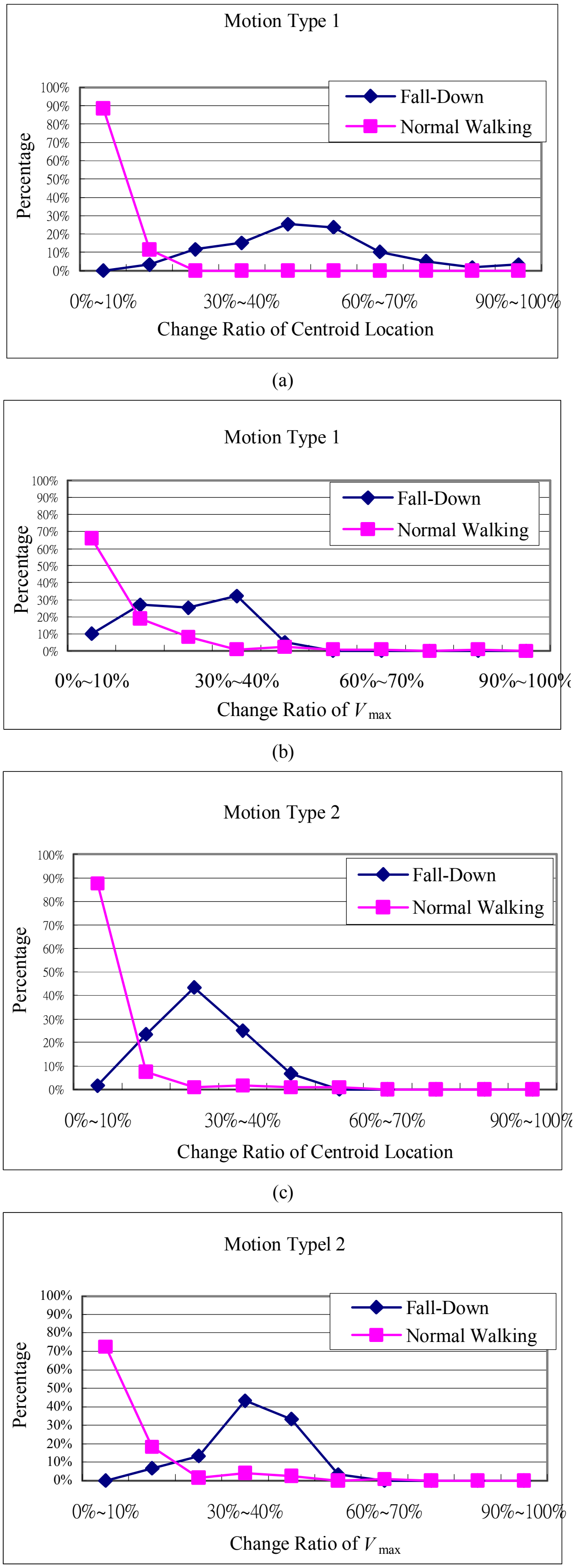
The relation between the direction of a moving object and the camera would affect the distribution of feature values extracted for fall incident detection. Fig. 5 illustrates three types of object motions: the first type is a human object going toward the camera; the second is the human object going away from the camera; whereas the third is the object walking horizontally in front of the camera. Other types of motions can be represented as the combinations of Type 1 and Type 3 or the combinations of Type 2 and Type 3. Because the distributions of feature vectors with different motion types are different as will be shown later, we use different threshold values for the three motion types, respectively.

Squatting has similar behavior with falling in terms of the centroid location. However, the change rates of the centroid of squatting is much slower than those of fall incidents as illustrated in Fig. 4. The characteristics can be used to differentiate normal squatting events (slow change rate) from fall-down events (fast change rate) by choosing appropriate thresholds. Typically, falling and squatting have significantly different centroid changes (128→186 for falling and 147→176 s for squatting, respectively). Using appropriate threshold can detect these two events as well as achieve good differentiation accuracy.

IV. Experimental Results

For fall incident detection, totally 78 sequences including 48 training sequences and 30 test sequences were used in our experiments. The 48 training sequences containing three different motion types (16 sequences for each motion type) were used to determine the thresholds for the three motion types, respectively. Among every 16 sequences for each motion type, eight sequences consist of fall incident events, whereas the other eight sequences contain no fall incidents. Fig. 6 depicts the statistical distributions of the centroid location and $V_{\max}$ change ratios collected from the training sequences for the three motion types, respectively. The change ratios are calculated by (5) with a sliding interval of 0.6 second between two frames. As shown in Fig. 6, the two normalized feature values (i.e., the horizontal axes) are both divided into 10 bins, each containing 10% of the whole range. For each object motion type, we choose a threshold for each feature value. Each threshold is chosen to minimize the error rate of event detection according to the associated distribution in Fig. 6. The thresholds for the three motion types are listed in Table I. Since the motion behavior of a human may be a combination of two of the three motion types, we use a linear combination of the two corresponding thresholds to calculate the threshold according to the motion types determined by using the trajectory of centroid and the change rate of object height.

We used 30 test sequences with different motion types of fall incidents to evaluate the performance of the proposed fall-incident detection algorithm. These sequences consist of 15 sequences with fall incidents and 15 normal walking sequences. Our system can correctly detect 28 events including 13 fall incidents and 15 normal ones from the 30 sequences; whereas two fall incidents were missed. The correctness ratio is about 93%. The miss ratio is 13% and the ratio of false positives is 0%. The reason of unsuccessfully detecting the two fall incidents was that the human objects in the two sequences has small $V_{\max}$ values, which was due to some false-segmentation caused in part by show noise. Because a small $V_{\max}$ value of an object leads to a small change amount of the object's centroid location and $V_{\max}$, it would become relatively difficult to detect a fall in such sequences correctly.



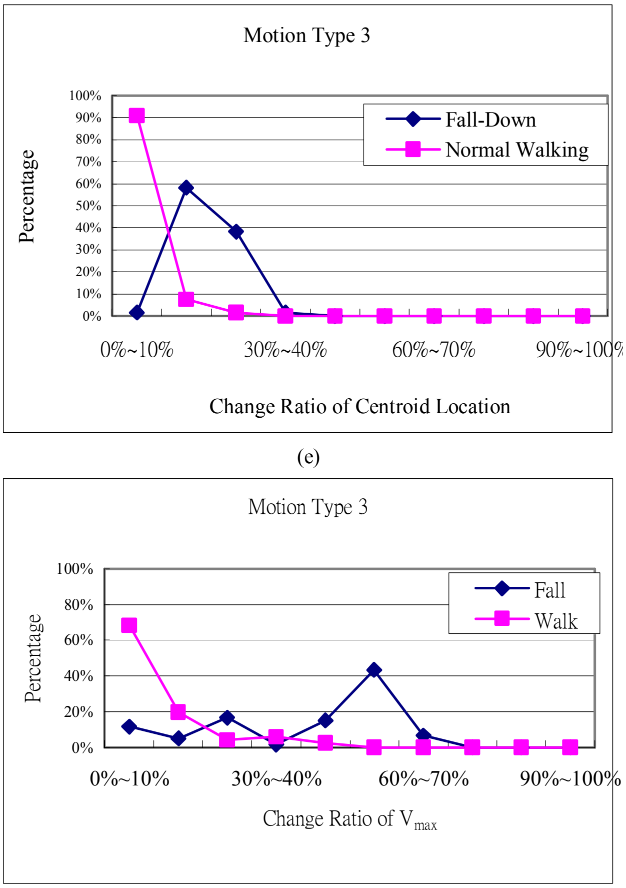


Fig. 6. Histograms of the centroid location and $V_{\max}$ change ratios between fall incidents and normal walking for the three motion types: (a)-(b) Motion Type 1; (c)-(d) Motion Type 2; (e)-(f) Motion Type 3.

TABLE I. THRESHOLDS FOR EACH MOTION TYPE

Motion type	Change ratio of centroid location	Change ratio of $V_{\max}$
Type 1	24%	25%
Type 2	25%	30%
Type 3	12%	20%

V. CONCLUSION

We have presented a feature-based compressed-domain fall-down detection scheme for intelligent surveillance applications. The proposed scheme involves two steps: compressed-domain object extraction and fall incident detection. After extracting moving human objects, three feature values: the change ratio of the centroid of a human object, the change ratio of the maximum of vertical projection histogram, and the duration of an event detected are used to identify and locate fall-down events. The proposed system can also differentiate fall-down from squatting by taking into account the event duration. The proposed object segmentation method can extract moving objects with or without cameral motions, thereby being useful for video surveillance applications equipped with still or pan-tilt-room cameras. Our experimental results show that the proposed method can detect fall incidents with high accuracy in real-time.

REFERENCES

[1] C. S. Regazzoni, V. Ramesh, and G. L. Foresti, "Scanning the issue/technology- Special issue on video communication, processing and understanding for third generation video surveillance systems," *Proc. IEEE,* vol. 89, no. 10, pp. 1355-1367, Oct. 2001.

[2] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. Systems, Man, and Cybernetics- Part C: Applications and Reviews*, vol. 34, no. 3, pp. 334-352, Aug. 2004.

[3] J. Teno, D. Kiel, and V. Mor, "Multiple strumbles: a risk factor for falls in community-dwelling elderly," *J. America Geriatrics Society,* vol. 38, no. 12, pp. 1321-1325, 1990.

[4] T. Tamura *et al*., "An ambulatory fall monitor for the elderly," in *Proc. IEEE Int. Conf. IEEE Int. Conf. Microtechnologies in Medicine and Biology*, pp. 2608-2610, July 2000., Chicago, IL.

[5] G. Williams *et al*., "A smart fall & activity monitor for telecare applications," in *Proc. Proc. IEEE Int. Conf. IEEE Int. Conf. Microtechnologies in Medicine and Biology*, vol. 20, no. 3, pp. 1151-1154, 1998.

[6] N. Noury, "A smart sensor for the remote follow up of activity and fall detection of the elderly," in *Proc. IEEE Int. Conf. Microtechnologies in Medicine and Biology*, pp. 314-317, May 2002, Lyon, France.

[7] I. Haritaoglu, D. Harwood, and L. S. Davis, "W^4: Real-time surveillance of people and their activities," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 809-830, Aug. 2000.

[8] K. Yoon, D.F. Dementhon, and D. Doermann, "Event detection from MPEG video in the compressed domain," in *Proc. IEEE Int. Conf. Pattern Rcognition*, 2000, Barcelona, Spain.

[9] H. Nait-Charif and S. J. McKenna, "Activity summarisation and fall detection in a supportive home environment," in *Proc. IEEE Int. Conf. Pattern Rcognition*, vol. 4, pp. 23-26, Aug. 2004, Cambridge UK.

[10] C.-W. Lin, Z.-H. Ling, Y.-C. Chang, and C. J. Kuo, "Compressed-domain fall incident detection for intelligent home surveillance," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 2005, Kobe, Japan.

[11] B. L. Yeo, *Efficient Processing of Compressed Images and Video*, Ph.D. thesis, Princeton University, Jan.1996.

[12] Y. Su, M.-T. Sun, and V. Hsu, "Global motion estimation from coarsely sampled motion vector field and the applications," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 2, pp. 628-631, Mar. 2003, Bangkok, Thailand.

[13] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Comm. Technology*, vol. 15, pp. 52-60, 1967.