Fall Detection with Two Cameras based on Occupied Area

Dao Huu Hung*
Hideo Saito*

In the effort of supporting elderly people living alone, this paper describes a novel video-based system for detecting fall incidents. Widths of a same person are extracted from two cameras whose fields of view are relatively orthogonal, for estimating the occupied area. We divide the scene into many small patches. Sizes of a person moving through the scene are clustered and kept in the buffer of each patch in which the person is captured, so-called Local Empirical Templates (LET), for building spatial distributions of occupied areas of the person in walking or standing poses. We realize that occupied areas of lying-down and sitting person are proportional to that of LET, spotted in the same scene patch. Therefore, we normalize the height and the occupied area of a person estimated from the two cameras with respect to those of LET in the same scene patch, leading the generation of a promising feature space in which three human states of standing, sitting or bending, and lying down, are in separable regions. Fall incidents can be inferred from the time-series analysis of human state transition. The experimental results with 24 realistic video samples in Multiple cameras fall dataset [1] demonstrates high detection and low false alarm rates.

Keywords: fall detection, local empirical templates, normalized height, and normalized occupied area

1. Introduction

Recently, the population of elderly people has been surging, particularly in developed countries. Majority of them live alone at home. They might be suffered from a *fall accident* which is considered as the most common cause of injury for elderly people. The degree of injury is proportional to the delay time in detecting fall incidents and helping them immediately reach to health care services [2]. Unfortunately, a fall probably make elderly people experience unconscious states of mind and physical pains. They are, in turns, unable to call for emergency services by themselves. In this case, fall detection systems which can automatically detect fall incidents and generate an alarm to emergency centers are essential in health care for elderly people.

So far, many systems were developed by using wearable devices, ambient devices or camera sensors to detect fall incidents. Undoubtedly, wearable devices and ambient devices-based systems are able to achieve high accuracy. However, they often make inconvenient for users in performing daily activities. In particular, elderly people may easily forget to wear such kinds of devices everyday. Noury *et al.* [2] and Yu [3] provided comprehensive reviews of wearable sensors and ambient devices for fall detection. In contrast, camera sensors mounted on the walls, which are promising solutions are considered in this paper.

The main challenge of using video surveillance is to distinguish fall incidents from like-fall ones, i.e. losing balance, sitting down brutally, crouching on the floor, and lying down on a sofa, etc. which often happen when the person performs daily activities. In addition, the person is often occluded by the furniture in the room under the view of camera. Dynamic lighting conditions, and low contrast between the person's appearance and background also pose difficulties.

In order to handle these challenges, many methods were proposed and are summarized in the following text. By using a single uncalibrated camera, the early work of Anderson *et al.* [4] analyzed the size of human body silhouettes. When the person is standing and falling, the width-to-height ratios of silhouettes are small and large, respectively. The aspect ratio changes may not follow the rule due to the effect of human body upper limb activities. To eliminate this effect, Liu *et al.* [5] used a statistical scheme to remove peaks in vertical projection histograms of silhouette images. K-NN classifier is trained by the features of aspect ratio and the difference between height and width of human body silhouette. Both methods merely reported the experiments with the camera placed sideways. In practice of indoor surveillance, the camera is preferred to be mounted obliquely near to the ceiling for a wider field of view and occlusion avoidance. Shoaib *et al.* [6] presented a context model to learn the head and floor planes from the foregrounds of one person moving in the scene under an oblique setting of camera. Distance measures between locations of detected heads and reference heads, provided by the head plane, are used to classify fall incidents from other events. Obviously, analyzing 2D silhouettes has two weaknesses in common. It is unable to discriminate the fall in parallel to the optical axis of the camera from the sitting and bending poses. It is prone to poor performance when the person is occluded by other objects or

* Keio University
3-14-1, Hiyoshi, Kohoku-ku, Yokohama-shi, Kanagawa, Japan 223-8522,
hungdaohuu@hvrl.ics.keio.ac.jp
saito@hvrl.ics.keio.ac.jp

performs daily activities.

Motion History Image is adopted to quantify the motion of human body [7]. Large motion is more likely caused by a fall incident. Silhouettes are approximated by ellipse models whose orientation angle and ratio of major and minor semi-axes' length are used to discriminate a fall from other activities, including the like-fall ones, i.e. sitting down brutally. Similarly, integrated spatiotemporal energy map is used to calculate motion activity coefficients for detecting a large motion event [8]. Orientation angle, displacement, and major-to-minor-semi-axes ratio of human ellipse models are analyzed in the framework of Bayesian Belief Networks to recognize fall and slip-only events. Chen *et al.* [9] presented a combination of distance map of two sampling human skeletons and variation analysis of ellipse human models. Human shape is supposed to change progressively and slowly during usual activities, and drastically and rapidly during a fall [10]. Therefore, shape-matching costs during a fall and during a usual activity are high and low, respectively. The method is reported to work at the frame rate of 5 fps due to the expense of high computational cost.

Apparently, in single uncalibrated camera-based methods, 3D information that is important in detecting falls is not made use of. Therefore, Cucchiara *et al.* [11] used a sideways calibrated camera to train probabilistic projection maps for each posture, i.e. standing, crouching, sitting and lying. They suggested using a tracking algorithm with a state-transition graph to handle occlusions, in turn, leading reliable classification results. In their latter work [12], partial occlusion is detected and compensated by a wrapping method from multiple cameras. A Hidden Markov Model (HMM) is trained for obtaining more robust recognition results. 3D head trajectory is extracted from particle filter-based head tracking to discriminate fall incidents [13]. Thome *et al.* [14] applied the metric image rectification to derive the 3D angle between vertical line and principal axis of ellipse human models. Decisions made independently by multiple cameras are fused in a fuzzy context to classify postures. Layer HMM is hand designed to make event inference. Anderson *et al.* [15] introduced a framework of fall detection in the light of constructing voxel person. Linguistic aspect of the hierarchy of fuzzy logic used in this research for fall inference makes this framework extremely flexible, allowing for user customization based on their knowledge of cognition and physical ability. Recently, Auvinet *et al.* [16] discussed a method of reconstructing 3D human shape from a network of cameras. They proposed the idea of Vertical Volume Distribution Ratio since volumes of standing and lying-down person are vertically distributed significantly differently. The method is able to handle occlusion since the 3D reconstructed human shape is contributed from multiple cameras.

In this paper, we propose a novel method of fall detection by using two cameras whose fields of view are relatively orthogonal, facilitating the estimation of occupied areas of people. Instead of using calibrated cameras, our fall detection system learns the sizes of standing or walking people who appear in small local patches of the scene. The learning process is straightforward and can be implemented automatically. The purpose of this learning process is to build a so-called *local empirical template*, typically representing the size of a standing person for each local scene patch. Consequently, the spatial distribution of occupied area of a person in the standing pose can be constructed from LETs, learned from the two cameras. Since the height and occupied area of a sitting and lying-down person are proportional to those of a standing person in the same patch in the scene, we choose the feature space composed of height and occupied area normalized with respect to those of LETs. Interestingly, the novelty of this paper is the proposed feature space in which three human poses of standing, sitting and lying down are in three separable regions. Fall incidents are detected by time-series analysis of human state or pose transition.

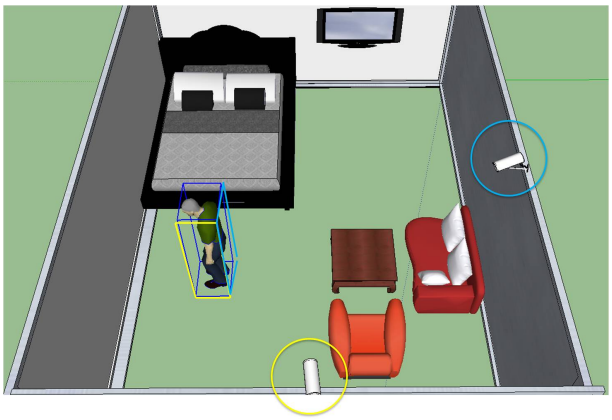


Fig. 1. Estimation of occupied area of a person from the views of two cameras

The rest of this paper is organized as follows. Section 2 is dedicated on our proposed fall detection system. Experimental results are reported in Section 3 with discussion. Conclusions are made in Section 4.

2. Our proposed fall detection system

2.1 Estimation of occupied area from two views Fig. 1 shows an example of home environments in which elderly people are assumed living alone. If there are more than one person presenting in the room, using a fall detection system seems to be not much meaningful. In this sense, our proposed fall detection system is activated upon the detection of only one person. Two cameras are mounted on the walls, near to the ceiling, and in the oblique settings so that their fields of view are relatively orthogonal. Foregrounds are extracted from the input videos by using Gaussian Mixture Model [17] (GMM). The person detected in the first camera is identical to the one detected in the other camera. Foreground of the person in each view is represented by a rectangle or a bounding box. Under the two orthographic views, the occupied area of the person is roughly estimated by

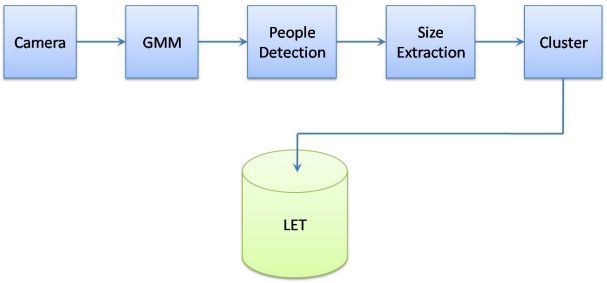


Fig. 2. Flowchart of extracting Local Empirical Templates

$$OA = W_{Cam1} \times W_{Cam2} \cdots\cdots\cdots\cdots\cdots\cdots (1)$$

where OA, occupied area of the person, appearing in the intersection of two fields of view, W_{Cam1}, width of the person extracted from the first camera, and W_{Cam2}, width of the person extracted from the other one.

2.2 Local empirical templates Apparently, the occupied area of a person estimated by Eq. 1 varies throughout the scene due to the perspective effect of cameras. To eliminate the perspective effect, we suggest using *Local Empirical Templates* (LET), which have demonstrated to be effective in dealing with this issue [18].

The cameras are assumed to be stationary and to be higher than people's head. Images captured from the two cameras are divided into many cells. The number of cells depends upon the resolution of the images and the camera settings. Foregrounds of people moving in the scene are captured along their trajectories and kept in the buffer of each cell. We cluster the sizes, width and height, of the foregrounds in each cell since people are different in size, leading the generation of an appropriate *empirical template* for each cell. By its nature, the sizes of templates near to and far away the camera are large and small, respectively. Therefore, LET provides a good reference information of perspective caused by the camera. In addition, LET is also able to handle the effect of image distortion. The procedure of obtaining LET in unknown scenes is straightforward and can be performed automatically as shown in Fig. 2

2.3 Feature selection and computation We assume that a person is detected at the cell (m, n) in the view of the first camera †. The person is also appeared at the cell (p, q) in the view of the other one. From the buffer of cell (m, n) in the view of the first camera, the template $T_{Cam1}(m,n) = \{W^{temp}_{Cam1}(m,n), H^{temp}_{Cam1}(m,n)\}$ is extracted in which $W^{temp}_{Cam1}(m,n)$ and $H^{temp}_{Cam1}(m,n)$ are width and height of the template $T_{Cam1}(m,n)$, respectively. Similarly, the template $T_{Cam2}(p,q) = \{W^{temp}_{Cam2}(p,q), H^{temp}_{Cam2}(p,q)\}$ is also extracted from the buffer of cell (p, q) in the view of the other one.

Firstly, we choose the person's height as a feature to distinguish the standing pose from sitting and lying-down ones. Since the height of person in standing pose is always well larger than that of person in sitting and lying-down poses, we normalize the height of detected person with respect to the height of LET, extracted in the same scene patch with the detected person. It is important to note that LET is the template of people in standing pose. If we denote widths and heights of the person detected at the cell (m, n) in the first camera and at the cell (p, q) in the other one by $W_{Cam1}(m,n), H_{Cam1}(m,n), W_{Cam2}(p,q)$, and $H_{Cam2}(p,q)$, respectively. The normalized height is calculated as the follows.

$$NH_{Cam1}(m,n) = \frac{H_{Cam1}(m,n)}{H^{temp}_{Cam1}(m,n)}$$
$$NH_{Cam2}(p,q) = \frac{H_{Cam2}(p,q)}{H^{temp}_{Cam2}(p,q)} \cdots\cdots\cdots\cdots (2)$$

In indoor environments, the person is likely occluded by the furniture under the view of camera, leading the inaccurate measures of height. Broken foreground also contributes to the inaccuracy of the height measures. However, it is rarely that the person is occluded by other objects and broken foreground happens in the both views of cameras. Therefore, the two measures in Eq. 2, extracted from two cameras should be fused to produce a more reliable feature, representing the height of the person. In this paper, we take their average.

$$NH(m,n,p,q) = \frac{NH_{Cam1}(m,n) + NH_{Cam2}(p,q)}{2} (3)$$

However, in two cases of the person falling in parallel to the optical axis of camera and the person sitting on a chair or bending the body, 2D measures of the person's height by Eq. 3 are quite similar. Fortunately, the occupied areas of the person in these two cases are significantly different. We estimate the occupied area of the person as the follow.

$$OA(m,n,p,q) = W_{Cam1}(m,n) \times W_{Cam2}(p,q) (4)$$

Similarly, occupied area of the LET in the same scene patch with the detected person is estimated by

$$OA_{temp}(m,n,p,q)$$
$$= W^{temp}_{Cam1}(m,n) \times W^{temp}_{Cam2}(p,q) \cdots\cdots\cdots (5)$$

We reveal that the occupied areas of the person in sitting and lying-down poses are proportional to that of the person standing in the same position in the scene. We normalize the occupied area of a person estimated by Eq. 4 with respect to the occupied area of LET, estimated by Eq. 5, leading the generation of a feature, so-called *Normalized Occupied Area*.

$$NOA(m,n,p,q) = \frac{OA(m,n,p,q)}{OA_{temp}(m,n,p,q)}$$
$$= \frac{W_{Cam1}(m,n) \times W_{Cam2}(p,q)}{W^{temp}_{Cam1}(m,n) \times W^{temp}_{Cam2}(p,q)} \cdots\cdots\cdots (6)$$

The normalization process cancels the perspective effect of cameras. Therefore, the normalized measures of

† a person is considered to be detected in a cell if the person's head appears in the cell

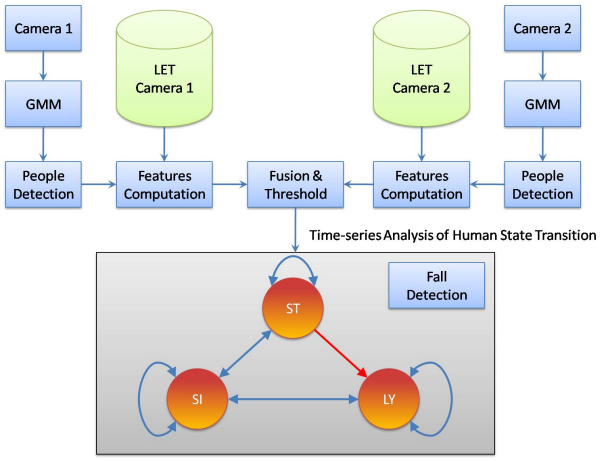


Fig. 3. Flowchart of our proposed fall detection system in which ST, SI, and LY are STanding, SIting, and LYing states, respectively

both person's height and occupied area are lower and upper bounded. They are also independent of positions of the person in the scene. The position notation in Eq. 3 and Eq. 6 can be simplified. In summary, our proposed feature space is composed of *normalized height*, NH and *normalized occupied area*, NOA of a person.

We reveal that three states of a person, standing, sitting and lying down, are in three separable regions in the proposed feature space. For a standing person, both features of normalized height and normalized occupied area vary around one, since LET is the template of a standing person. The height of sitting or lying-down person is smaller than that of standing person in the same patch in the scene. We find a threshold TH_{height} to distinguish sitting and lying-down people from standing people. Obviously, the value of threshold is smaller than one. It is the fact that occupied area of a lying-down person is larger than that of a sitting person, considering in the same scene patch. We also find a threshold TH_{area} to differentiate lying-down people from sitting people. In our implementation, we choose 1 out of 24 video samples in the dataset for training, in order to find the thresholds TH_{height} and TH_{area}. In this training video, the person must exhibit all three poses of standing, sitting on a chair, and lying on the floor. The other video samples are used for cross validation.

2.4 Fall detection The flowchart of our proposed fall detection system is shown in Fig. 3. Foregrounds of a moving person are extracted from both cameras for features computation and fusion. After thresholding the features, we are able to detect the states of the person since they lie in three separable regions of the feature space. In this paper, three human states, namely, standing, sitting and lying are taken into account. Bending bodies are considered as the state of sitting since people in either sitting or bending poses occupy similar areas.

It is noted that the human states are detected frame by frame. However, to detect fall incidents, we must keep eye on the history of human states, as shown in Fig. 3. Fall incidents are claimed to be detected when

Table 1. Actions can be inferred from the time-series analysis of human state transition

	Standing	Sitting	Lying
Standing	Standing or Walking	Sitting down	**Falling**
Sitting	Standing up	Sitting	Lying or Crouching
Lying	*Prohibited*	Getting up	Lying

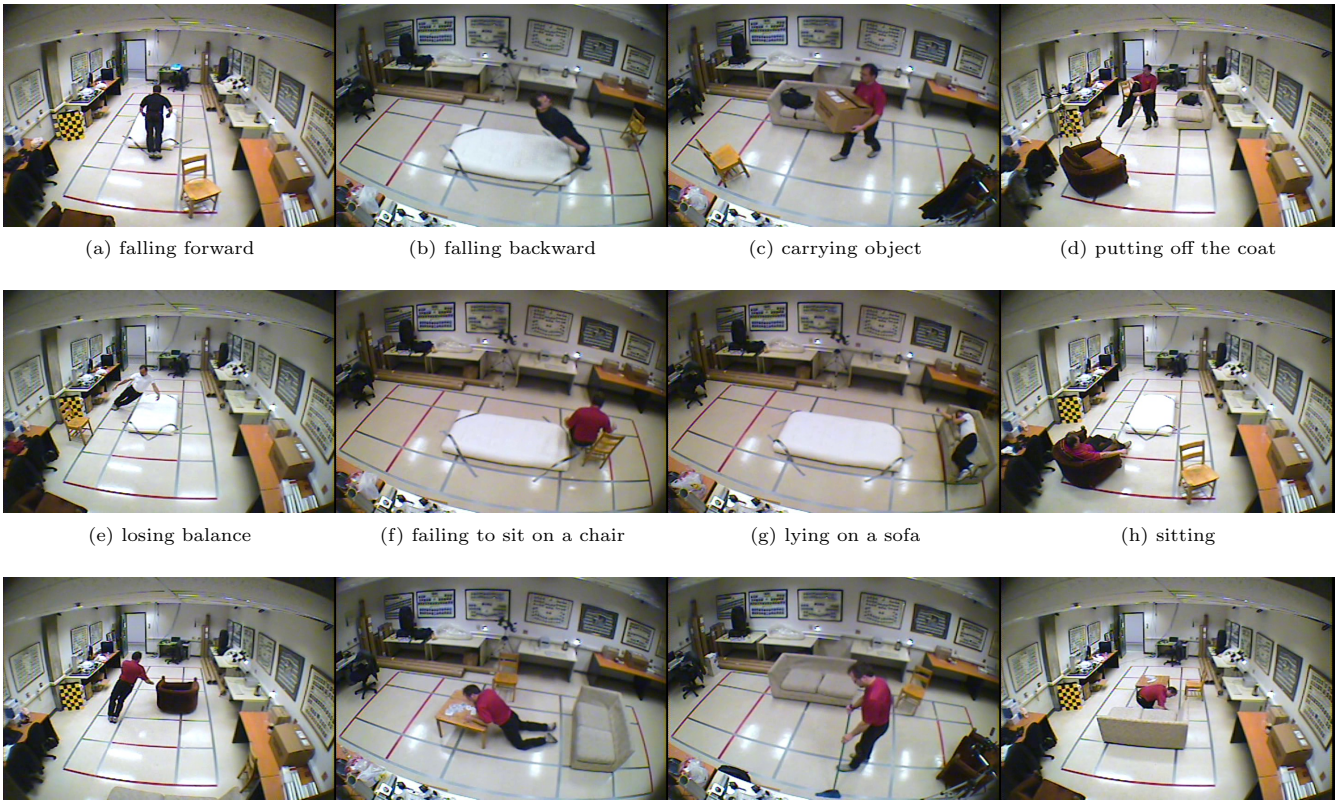
people changes their states directly from standing to lying [11]. For lying-down action of elderly people, for example, crouching on the floor or lying on the floor, the state transition should be from standing to sitting and subsequently to lying-down states. Changing states directly between lying and standing is prohibited. The transition must undergo the intermediate state of sitting. The summarization of actions that can be inferred by the time-series analysis of human state transition is given in the Table 1.

In this paper, we keep states of the person detected in N frames for analyzing the state transition in a probabilistic framework. We define a stable state in the N frames that appears with the highest probability. The other states are considered as unstable states with small probabilities. When the person makes a state transition, the probability of the stable state is gradually reduced. Meanwhile, the probability of an unstable state is progressively surged. The state transition is confirmed when the probability of the unstable state is higher than that of the stable state. In the implementation, N is set to 60.

3. Experimental results and discussion

3.1 Multi-view fall dataset For fair comparisons with existed methods of fall detection, the performance evaluation must be conducted in a same dataset. In this paper, we use the "Multi-view fall dataset" released by Auvinet *et al.* [1] and compare the performance of our proposed fall detection system with that of latest works [10] [16], tested on the same dataset. This dataset recorded simulated falls simultaneously from eight cameras, which are mounted on the walls and in the oblique settings. The dataset consists of 24 realistic scenarios, showing 24 fall incidents and 24 confounding events (11 crouching, 9 sitting and 4 lying on a sofa). An experienced clinician in the field of health care for elderly people tried to simulate different kinds of fall, for instance, falling forward, falling backward, and losing balance, etc. The ground truth for each frame is also provided along with the video samples. We use the video samples captured by the second and the fifth cameras for making experiments since they form two orthographic views. In the dataset, inexpensive IP cameras with a wide angle to cover all the room are employed. Consequently, the images are highly distorted. Fig. 4 illustrates some examples of simulated falls and daily activities.

3.2 Separable feature space In Section 2.3, we discuss about the separable characteristics of our proposed feature space, consisting of normalized height NH and normalized occupied area NOA. In this section, we will demonstrate the validity of our discussion. We use 1 out of 24 video samples in the dataset in the training



(a) falling forward (b) falling backward (c) carrying object (d) putting off the coat

(e) losing balance (f) failing to sit on a chair (g) lying on a sofa (h) sitting

(i) falling to the sofa (j) falling from the sofa (k) doing house work (l) crouching with occlusion

Fig. 4. Examples of simulated falls and daily activities

phase for finding the thresholds TH_{height} and TH_{area} to separate the feature space. This video sample must contain three actions of walking, sitting, and lying on the floor. Therefore, the ninth scenario seem to fit the training purpose. We also use another video sample showing a man walking through the scene for learning LETs. In the video sample, he does not walk through all corners of the scene. Therefore, interpolation and extrapolation with the prior knowledge of the scene perspective are performed for learning LETs. We compute the features by Eq. 3 and Eq. 6 for every frame of the ninth scenario and draw them in the feature space along with the ground truth, as shown in Fig. 5.

In this training scenario, the man enters the scene and approaches to the chair, that is, he is in the standing pose. The features of his standing pose in this duration vary around the point (1, 1) in the feature space. When he sits on a chair, his normalized height declines below 0.65; The red line in Fig. 5 shows the transition when he sits down and stands up. After standing up, he falls and lies on the floor. In the state of lying on the floor, his normalized height is also well smaller than 0.65; Therefore, we can set the threshold TH_{height} to 0.65 for discriminating the standing pose from sitting and lying poses. Fortunately, the normalized occupied areas of sitting and lying poses are significantly different. We use the threshold TH_{area} of 2 to separate the regions of sitting and lying poses in the feature space. We test our proposed fall detection system with these above thresholds on the other video samples of the dataset. Fig. 6

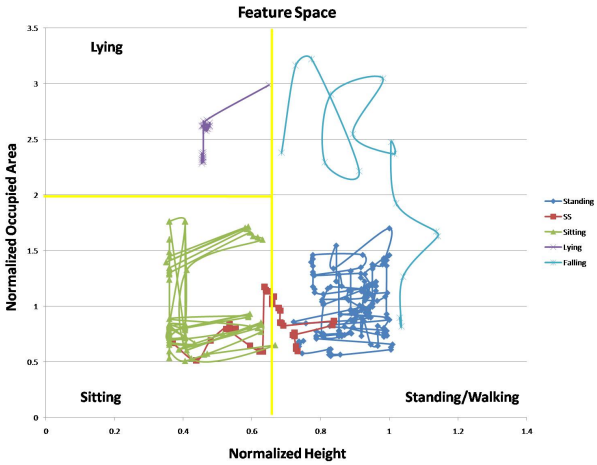


Fig. 5. The feature space of the ninth scenario, in which SS stands for the transition from Standing to Sitting

demonstrates the feature space of the eighth scenario, tested with the thresholds obtained in the training. The fall incident is detected by the time-series analysis of human state transition and are denoted by the red line in Fig. 6. It also confirms the validity of the important separability characteristics of our proposed feature space.

3.3 Performance evaluation, comparison and discussion In this section, we will evaluate the performance of our proposed fall detection system and com-

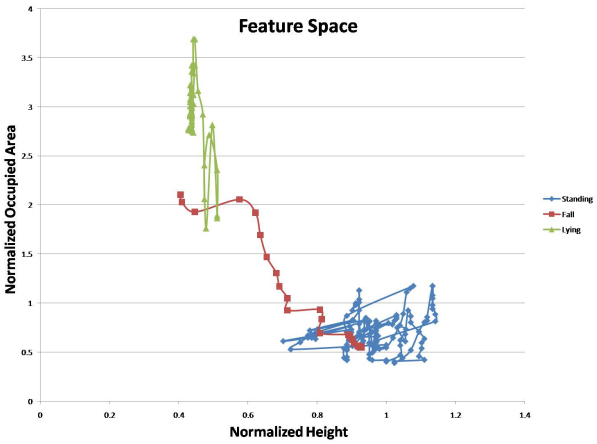


Fig. 6. The feature space of the eighth scenario, testing with thresholds obtained in the training

pare with other methods [10] [16], tested on the same dataset. To do that, we compute the sensitivity and the specificity, obtained from the time-series analysis of human state transition, as the follows.

(1) True Positive (TP): the number of falls correctly detected.

(2) False Negative (FN): the number of falls not detected.

(3) False Positive (FP): the number of normal activities detected as a fall.

(4) True Negative (TN): the number of normal activities not detected as a fall.

(5) Sensitivity: $Se = \frac{TP}{TP+FN}$

(6) Specificity: $Sp = \frac{TN}{TN+FP}$

High sensitivity means that most fall incidents are correctly detected. Similarly, high specificity implies that most normal activities are not detected as fall events. A good fall detection system has high values of sensitivity and specificity.

We have tested our proposed fall detection system on 24 realistic video samples of the dataset. The system is able to detect 23 out of 24 fall incidents. It fails to detect the fall event in the 22nd scenario since the person is sitting stably on a chair and subsequently he slips to the floor. Therefore, our system detects as a lie-down event. No normal activities detected as falls are reported. Our sensitivity and specificity of our fall detection system are 95.8% and 100%, respectively. Table 2 shows the performance comparison between our methods and two other methods [10] [16], tested on the same dataset.

Table 2 shows that our results are very competitive in comparison with those of other methods [10] [16] tested on the same dataset. It is noted that the results of the method proposed by Auvinet *et al.* [16] are reported with a network of three cameras. The sensitivity of their method can be boosted to 100 % if a network of more than four cameras is employed. Both of these methods are high computational costs. Rougier *et al.* [10] reports the implementation of 5 fps and argues that it is sufficient for detecting fall events. Auvinet *et al.* [16] presents the GPU implementation to realize their method in real-time. However, our proposed fall detection system composing of low cost modules is implemented in real-time in a common desktop PC.

Table 2. Performance comparison between our method and two other works [10] [16], tested on the same dataset

	Sensitivity (Se)	Specificity (Sp)
Our method	95.8 %	100 %
Auvinet *et al.* [16] †	80.6 %	100 %
Rougier *et al.* [10]	95.4 %	95.8 %

Our proposed fall detection system is user-friendly. After the cameras are mounted on the wall, users can run the system in the learning mode. The learning process is performed automatically, capturing the foregrounds of a person walking through the scene. It is better for the system to capture the size of monitored person. When the learning process is completed, the system is ready for detecting falls. It is noted that the initial settings can be done by the users without knowledge of computer vision. In addition, our fall detection system is able to preserve the privacy of the users.

4. Conclusions

We have presented a novel method of fall detection for aiding elderly people living alone. Local empirical templates are used to handle the effects of distorted images and the perspective. We have proposed a novel feature space, composing of normalized height and normalized occupied area. The most important characteristics of the feature space is that three human states of standing, sitting and lying on the floor lie in three separable regions. Fall incidents are detected by the time-series analysis of human state transition. Our proposed method achieves the very competitive performance in comparison with other methods [10] [16], tested on the same dataset. However, it has two other advantages: (1) the real-time implementation in a common PC since it is composed of low computational cost modules, and (2) the user-friendly system. In addition, it is able to preserve the privacy of the users.

References

(1) E. Auvinet, C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Multiple cameras fall data set", Technical Report 1350, DIRO Université de Montréal (2010)

(2) N. Noury, A. Fleury, P. Rumeau, A. K. Bourke, G. Ó. Laighin, V. Rialle, and J. E. Lundy, "Fall detection: principles and methods", Proc. of 29th IEEE Int'l Conf. on EMBS, pp. 1663-1666 (2007)

(3) X. G. Yu, "Approaches and principles of fall detection for elderly and patient," Proc. of 10th IEEE Int'l Conf. on e-Health Networking, Applications and Services, pp. 42-47 (2008)

(4) D. Anderson, J. M. Keller, M. Skubic, X. Chen, and Z. H. He, "Recognizing falls from silhouettes," Proc. of 28th IEEE Int'l Conf. on EMBS, pp. 6388-6391 (2006)

(5) C. L. Liu, C. H. Lee, and P. M. Lin, "A fall detection system using k-nearest neighbor classifier," Expert Systems with Applications, Vol. 37, pp. 7174-7181 (2010)

(6) M. Shoaib, R. Dragon, and J. Ostermann, "View-invariant fall detection for elderly in real home environment," Proc. of the 4th Pacific-Rim Symposium on Image and Video Technology, pp. 52-57 (2010)

(7) C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Fall detection from human shape and motion history using video surveillance," Proc. of 21st Int't Workshops on Advanced Information Networking and Applications, pp. 875-880 (2007)
(8) Y. T. Liao, C. L. Huang, and S. C. Hsu, "Slip and fall event detection using Bayesian Belief Network," Pattern Recognition, Vol. 45, pp.24-32 (2012)
(9) Y. T. Chen, Y. C. Lin, and W. H. Fang, "A hybrid human fall detection scheme," Proc. of IEEE Conf. on Image Processing, pp. 3485-3488 (2010)
(10) C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Robust video surveillance for fall detection based on human shape deformation," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 21, No. 5, pp. 611-622 (2011)
(11) R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, "Probabilistic posture classification for human-behavior analysis," IEEE Trans. on Systems, Man, and Cybernetics, Vol. 35, No. 1, pp. 42-54 (2005)
(12) R. Cucchiara, A. Prati, and R. Vezzani, "A multi-camera vision system for fall detection and alarm generation," Expert Systems, Vol. 24, No. 5, pp. 334-345 (2007)
(13) C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Monocular 3D head tracking to detect falls of elderly people," Proc. of 28th IEEE Int'l Conf. on EMBS, pp. 6384-6387 (2006)
(14) N. Thome, M. Serge, and S. Ambellouis, "A real-time multiview fall detection system: A LHMM-based approach," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 18, No. 11, pp. 1522-1532 (2008)
(15) D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, and M. Aud, "Linguistic summarization of video for fall detection using voxel person and fuzzy logic," Computer Vision and Image Understanding, Vol. 113, pp. 80-89 (2009)
(16) E. Auvinet, F. Multon, A. St-Arnaud, J. Rousseau, and J. Meunier, "Fall detection with multiple cameras: An occlusion-resistant method based on 3D silhouette vertical distribution," IEEE Trans. on Information Technology in Biomedicine, Vol. 15, No. 2, pp. 290-300 (2011)
(17) C. Stauffer and W. L. R. Grimson, "Learning patterns of activities using real-time tracking," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, pp. 747-757 (2000)
(18) D. H. Hung, S. L. Chung, and G. S. Hsu, "Local empirical templates and density ratios for people counting," Proc. of 10th Asian Conf. on Computer Vision, Vol. 4, pp. 90-101 (2010)

Dao Huu Hung received B.Sc. from Hanoi University of Technology, Vietnam and M.Sc. from National Taiwan University of Science and Technology, Taiwan, both in Electrical Engineering, in 2007 and 2010, respectively. In 2011, he joined Panasonic R&D Center Vietnam and spent one month at Panasonic Advanced Technology Development Center, Nagoya, as a visiting engineer. Currently, he is a PhD student at Hyper Vision Research Laboratory, Keio University at Yagami, Japan. His research interests include image processing, computer vision, pattern recognition and applications to surveillance systems.

Hideo Saito received his B.E., M.E., and PhD degrees in Electrical Engineering from Keio University, Japan, in 1987, 1989, and 1992, respectively. He has been on the faculty of Department of Electrical Engineering, Keio University since 1992. From 1997 to 1999, he stayed at Robotics Institute, Carnegie Mellon University as a visiting researcher. Since 2006, he has been a Professor of Department of Information and Computer Science, Keio University. He is currently the leader of the research project "Technology to Display 3D Contents into Free Space", supported by CREST, JST, Japan. His research interests include computer vision, mixed reality, virtual reality, and 3D video analysis and synthesis. He has been a program committee member of ICCV' 03, ICCV' 05, ICCV'09, area chair of ACCV' 09, ACCV' 10, general co-chair of ICAT' 06, ICAT' 08, and general chair of MVA' 09. He is a senior member of IEEE and IEICE, Japan.