View-Invariant Fall Detection System Based on Silhouette Area and Orientation

Behzad Mirmahboub, Shadrokh Samavi, Nader Karimi
Department of Electrical and Computer Engineering
Isfahan University of Technology
Isfahan, Iran

Shahram Shirani
Department of Electrical and Computer Engineering
McMaster University
Hamilton, Canada

Abstract—**Population of old generation that live alone is growing in most countries. Surveillance systems help them stay home and reduce the burden on the healthcare system. Automatic visual surveillance systems have advantages over wearable devices. They extract features from video sequences and use them for event classification. But these features are dependent on the position of cameras relative to the person. Therefore they need multi-camera for more accuracy that increases cost and complexity. In this paper we propose using silhouette area combined with inclination angle as robust features that can be measured using only one camera with an arbitrary direction. Through rigorous simulations on a publicly available dataset the error rate of the system is found to be less than 1%.**

Keywords-Video surveillance; fall detection; feature extraction; silhouette area; monocular system

I. INTRODUCTION

Number of elderly in developed countries is growing. Many of these people live alone. According to statistics, falling on the ground is the most important danger for this population that causes irreversible bone fracture and even death [1-6]. Immediate medical care is vital after the fall. But in many cases patients are unconscious and cannot call for help. Therefore we need automatic surveillance systems to detect the fall and produce an alert.

Commercial systems mainly consist of motion detection sensors, such as accelerometers and gyroscopes, which are needed to be worn by patients. But these devices are annoying and most people do not like or forget to wear them. Also, the devices error rates are high [1]. Visual surveillance systems have the advantage that they do not disturb the normal life of people and also provide more information about the environment. Monitoring the original videos by nurses is not acceptable because of personal privacy of patients. Hence, we need to extract some features from video sequences and decide based on them.

Shaou-Gang *et al.* [2] used the image sequence captured from an omni-camera mounted on ceiling. If aspect ratio of the bounding box of person is larger than a threshold in consecutive frames, it decides that fall has happened. This threshold should be adjusted for different persons. Toreyin *et al.* [3] used periodic nature of aspect ratio of bounding box in walking. They calculate wavelet transform of this signal to remove its stationary component. The high frequency part of the signal is used to train two Hidden Markov Models (HMM), one for walking and another for falling. In classification stage, each model that produces higher probability is selected. They also use audio signals to differentiate between falling and sitting down. Qian *et al.* [4] used two rectangles fitted on a person's image. One large box is for the whole body and a smaller one is fitted on the lower part of the body. Variations of the big box can differentiate standing from sitting, while the small box can separate walking from running. Features are given to Support Vector Machines (SVM) for classification.

The methods mentioned above need the person's motion to be parallel to the image plane. If he/she moves towards the camera, no change will be discovered form the bounding box. This requirement is not acceptable in real life. To solve this problem several research works proposed to use a multi-camera system. Rougier *et al.* [5] used shape matching and calculated a cost between consecutive frames. These costs are criteria for deformations and are fed to a Gaussian Mixture Model (GMM) for classification. Thereafter, the final result is obtained from voting between four cameras. Auvinet *et al.* [6] used camera calibration to reconstruct 3-D volume of the person from eight cameras. If a big portion of the body volume is near the ground for a period of time they classify it as a fall. Multi-camera methods have reached high accuracy (error rates of near zero) at the cost of high complexity. This could be a drawback if we want to build a low cost commercial system. We intend to propose a simple, yet accurate method in this paper.

In most of the existing fall detection systems we can distinct three major steps as shown in Fig. 1. Algorithm starts with separation of stationary background from moving foreground that is called motion segmentation. Then some features are extracted from foregrounds sequence which classification is done based on them. Event classifier can be a simple threshold or a complicated structure that needs to be trained beforehand [7]. In this paper we consider a single camera system with a fixed point of view. In our method

978-0-7695-4711-4/12 $26.00 © 2012 IEEE
DOI 10.1109/ICME.2012.193

there is no need for the person to be viewed at a specific angle to be able to distinguish his/her fall. Furthermore, the room under surveillance does not have to be marked by icons. We obtain silhouette of the person. Then we extract area and orientation of the silhouette as two features. Classification is done using Least Square Support Vector Machines based on these two features.

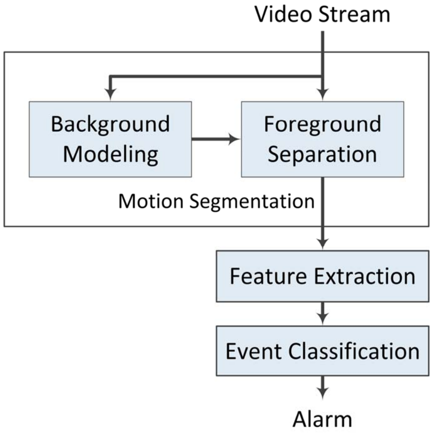


Figure 1. Overview of a visual fall detection system

The rest of paper is organized as follows. Section II describes proposed system and our motivation for selecting silhouette area. Section III reports experimental result. Eventually section IV concludes the paper.

II. Proposed Method

In this section we represent detailed explanation of our proposed method. At first we separate moving foreground from background. Then we extract silhouette area and orientation and show that silhouette area is a robust feature to view point. At last we classify fall and walk based on these two features.

A. Background separation

Separating moving objects in the scene (or segmentation of the scene) is the first step in automatic visual surveillance. Many methods exist for segmentation [8]. Each method is suitable for a special application. In this paper it is assumed that the indoor environment does not change rapidly. Therefore we use Running Gaussian Average because it has low computational requirement and we do not need very high accuracy. We keep one intensity value for each pixel in the background and update that as in (1):

$$b_t = \gamma I_t + (1-\gamma) b_{t-1}, \tag{1}$$

where b_t is the background value and I_t is pixel's current value for frame *t*. Also, γ determines the updating speed. At each frame *t* if $|I_t - b_t|$ is more than a threshold T_B, then the pixel belong to foreground. After subtraction we perform an erosion step, using a disk structuring element with radius of 1, to remove the noise.

A "silhouette" is a black and white image of a person or an object consisting of the outline and filled with solid color. Output of the background separation algorithm is a black image including only white silhouette of a moving object. Equation (1) has a drawback that it unnecessarily updates all pixels. As a result an unwanted "shadow" will appear in the silhouette that follows the moving object. Therefore, a modified version of (1) is usually used that prevents updating of foreground [8]. But we found out that the size of this shadow is a measure of motion speed. The faster an object moves, the bigger shadow will follow it. In the case of falling, sudden change in body position produces big shadow in the silhouette that can be used for fall detection. Furthermore, Equation (1) lets the silhouette of a stationary person to be faded out and hence the shrinking area could be used as a feature.

B. Feature extraction

Features, such as aspect ratio of bounding box or inclination angle of person's silhouette, are widely used for fall detection. But if the person moves towards the camera, these features will not give much information. Fig. 2 shows one frame before and another frame after fall event when motion of the person is parallel to image plane, while Fig. 3 shows the same situations when motion of the person is perpendicular to image plane.

As we see in Fig. 4(a) when the person moves parallel to the image plane, inclination angle changes from about 90 degrees to about 0 in the case of falling. But in Fig. 4(b) when person moves towards the camera, inclination angle decreases to about 60 degrees. If we use this angle as a measure for fall, we cannot know what threshold to use.

Process of falling is distinguished from normal activities by abrupt change in body position, impact to the ground and stationary period after that [1]. A good feature must take these characteristics into consideration.

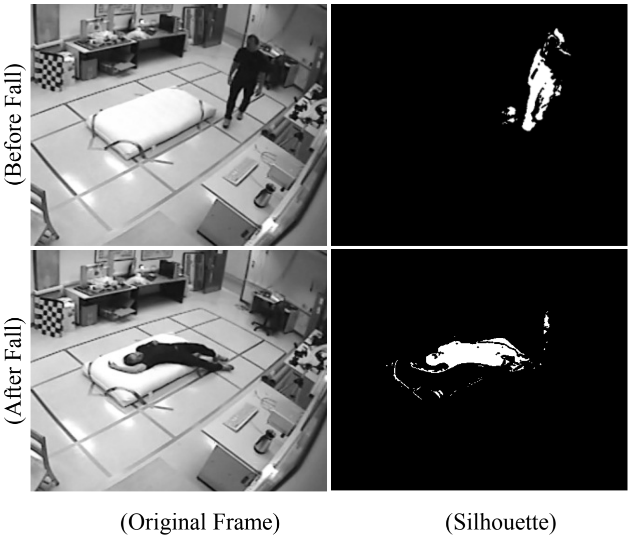


Figure 2. Original frame and silhouette before and after fall when falling is parallel to the image plane.

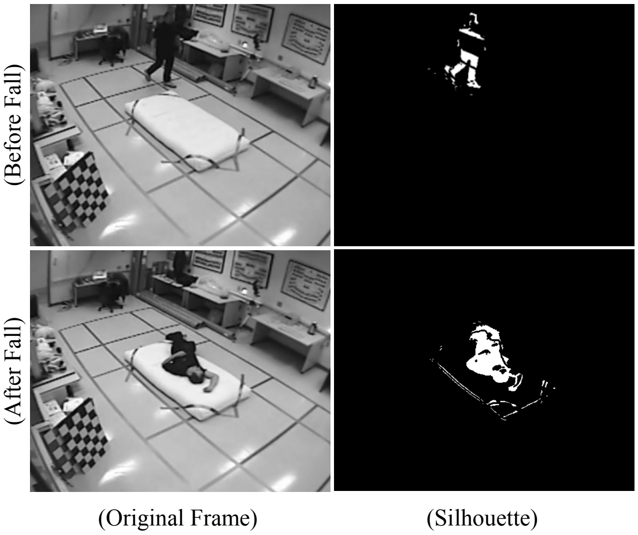


Figure 3. Original frame and silhouette before and after fall when falling is perpendicular to the image plane.

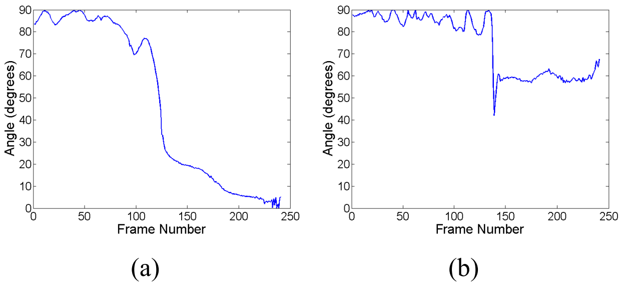


Figure 4. Variations of silhouette orientation corresponding to falling occurred in (a) Fig. 2 and (b) Fig. 3.

As we stated in subsection A, silhouette area computed based on (1) is a measure of abrupt movement. Fast motion produces large area of silhouette. Also impact of falling person shakes the environment that is detected as foreground and increases silhouette area. If the person does not move for a while, area gradually decreases because background is updating and the stationary person becomes a part of the background. This process is shown in Fig. 5.

These observations motivate us to use variations of silhouette area as a feature for fall detection that is robust to view point. Fig. 6 shows that unlike inclination angle, area variations produce almost the same pattern in (a) and (b). At the beginning of fall, area increases rapidly and there is a peak value in area variations. Then area decreases until it eventually becomes zero.

We compute orientation and area of silhouette in each frame and form two feature vectors with length of p; one for angle variations and another for area variations. These feature vectors are used in the classification step.

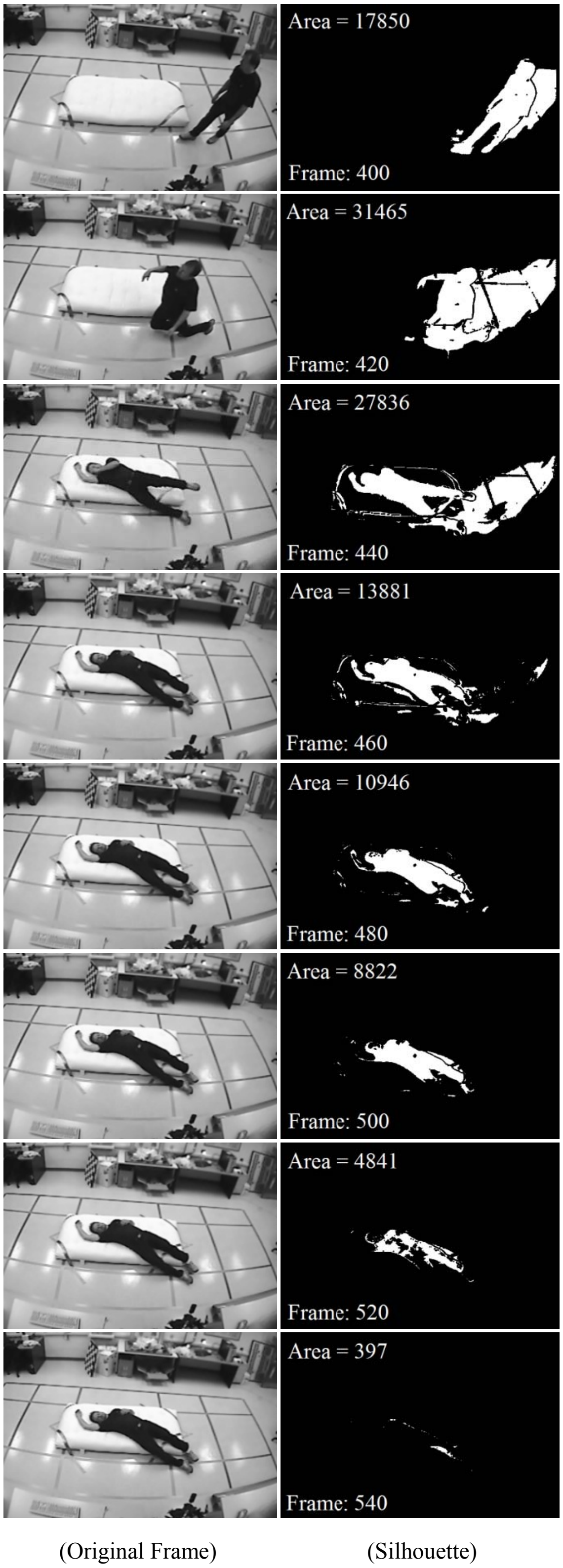
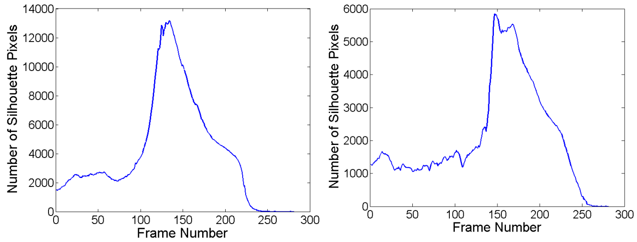


Figure 5. Variations in silhouette in the case of falling



(a) (b)

Figure 6. Variations of silhouette area corresponding to the falling events in (a) Fig. 2 and (b) Fig. 3.

C. Classification

Support Vector Machine (SVM) is widely used in classification problems. We use Least Square Support Vector Machines (LS-SVM) as explained in [9]. LS-SVM is computationally more efficient than standard SVM. In training phase it requires solving a linear set of equations rather than quadratic programming. Suppose we have N training samples of the form $\{x_i, d_i\}_{i=1}^{N}$ where $x_i \in \Re^p$ is p-dimensional input vector and d_i is labelled with 1 and -1. We want to find the optimal hyperplane to separate these samples. The solution is obtained by solving matrix equation that is shown in (2):

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & \Omega + C^{-1}I \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ d \end{bmatrix}, \tag{2}$$

where α is a vector of length N and β is the offset of hyperplane. Also, d is vector of labels of length N and $\vec{1}$ is a vector of 1's of length N. Here C is margin parameter and $\Omega_{ij} = K(x_i, x_j)$ is the kernel function that maps input training samples to higher dimensions. Some well known kernel functions are listed in TABLE I [9]. The optimal hyperplane is obtained by (3):

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + \beta, \tag{3}$$

where x is a test sample vector. In the training phase we assign label 1 and -1 to fall and walk events respectively and find α and β from (2). In classification phase we use (3) to find a label for testing sample x. If $y(x) > 0$ we classify x as a fall event, otherwise it is classified as a walk.

TABLE I. FAMOUS KERNEL FUNCTIONS

Linear	$K(x_i, x_j) = x_i^T x_j$
Polynomial of degree *m*	$K(x_i, x_j) = (x_i^T x_j + 1)^m$
Radial Basis Function (RBF)	$K(x_i, x_j) = e^{\frac{-\|x_i - x_j\|^2}{\sigma^2}}$

III. EXPERIMENTAL RESULTS

We used the fall dataset that is mentioned in [10,11]. In this dataset each action is captured with eight cameras from different directions. We selected 130 samples including walking and falling from different viewpoints. Each sample is characterised with two feature vectors. One vector consists of angle variations and another of area variations. Each fall event along with stationary part after that, takes place in about 200 frames. Orientation and area of the silhouette are computed in every frame. Hence, each of orientation and area feature vectors will have 200 elements. To reduce complexity and also to remove the noise, we choose one out of every five consecutive elements. The median of every five elements is chosen. Also we normalize data to compensate for variations in distances of person from the camera and variations in person's height. Therefore each feature vector has 40 normalized elements.

In the background separation step we select updating speed of $\gamma = 0.02$ and constant foreground threshold of $T_B = 30$. Also in the classification step we select margin parameter of $C = 100$.

The following parameters are used to analyse the recognition results of the algorithm [12].

- true positives (TP): number of falls detected correctly.
- true negatives (TN): number of walks detected correctly.
- false positives (FP): number of walks detected as fall.
- false negatives (FN): number of falls detected as walk.
- sensitivity: capability to detect a fall

$$Se = TP/(TP + FN) \tag{4}$$

- specificity: capability to detect a walk

$$Sp = TN/(TN + FP) \tag{5}$$

- accuracy: correct classification rate

$$Ac = (TP + TN)/(TP + TN + FP + FN) \tag{6}$$

- error rate: incorrect classification rate

$$Er = (FP + FN)/(TP + TN + FP + FN) \tag{7}$$

Simulations are implemented with MATLAB 7.10.0 on a Dell XPS L502X system with Intel® Core™ i7-2630QM CPU @ 2.00GHz, 8 GB of RAM, and Windows 7 Home Premium 64-bit operating system. We used 10-fold cross validation. Therefore we divide 130 samples to 10 equal parts. Each time we train classifier with 117 training samples and obtain TP, TN, FP, and FN for 13 remaining testing samples. This algorithm is repeated 10 times until all samples are tested. Simulation is performed separately based on orientation and area feature vectors with three kinds of kernels. Final results are shown in TABLE II and TABLE III.

For each feature vector, polynomial kernel of 2nd degree produces the best result. As we expected, variation of silhouette orientation (TABLE II) performs poorly in terms of accuracy and error rate; especially because the background separation method is not accurate and does not obtain exact silhouette. In fact, orientation has low sensitivity and cannot detect the fall events when the motion is towards camera. In contrast, variation of silhouette area (TABLE III) has good accuracy and error rate. Especially, sensitivity of area is high. Although it produced three false alarms with polynomial kernel, but missed only one fall event. In the special application of fall detection, sensitivity is more important than specificity. False alarms of the system can be deactivated by user. But if an unconscious person is not detected, it is devastating and is not tolerable.

TABLE II. CLASSIFICATION RESULTS BASED ON ORIENTATION

SVM Kernel	Linear	Polynomial (m=2)	RBF (σ = π)
TP	45	49	48
TN	64	63	63
FP	5	6	6
FN	16	12	13
Sensitivity (%)	73.77	80.33	78.69
Specificity (%)	92.75	91.30	91.30
Accuracy (%)	83.85	86.15	85.38
Error Rate (%)	16.15	13.85	14.62

TABLE III. CLASSIFICATION RESULTS BASED ON AREA

SVM Kernel	Linear	Polynomial (m=2)	RBF (σ = π)
TP	59	60	59
TN	61	66	61
FP	8	3	8
FN	2	1	2
Sensitivity (%)	96.72	98.36	96.72
Specificity (%)	88.41	95.65	88.41
Accuracy (%)	92.31	96.92	92.31
Error Rate (%)	7.69	3.08	7.69

Simulation results in TABLE II and TABLE III show that area produces less error than orientation. But they do not show how much overlap these two features have. Therefore, we counted numbers of correct and incorrect classifications based on orientation and area with polynomial kernel. Among 130 test samples, 108 ones are classified correctly by both classifiers. Also, 18 samples are classified correctly by the area based classifier but incorrectly by the orientation based classifier. The remaining 4 cases are classified correctly by the orientation based classifier and incorrectly by the area based classifier. There is not a single sample that is not correctly classified by at least one classifier.

We form new 2-dimensional samples. First dimension is the label produced by classifier based on area and second dimension is the label produced by classifier based on orientation. Fig. 7 plots these new samples in the new 2-dimensional feature space for three different kinds of kernels. For linear kernel in Fig. 7(a) a few samples are not separable. RBF kernel in Fig. 7(b) gives better results. But samples from polynomial kernel of 2nd degree in Fig. 7(c) are completely separable.

This observation motivates us to combine aforementioned features. Hence, we propose a system as shown in Fig. 8. Classification is done separately based on orientation and area feature vectors. Labels produced by these two classifiers will form new samples with 2-dimensional feature vectors that are fed to a final classifier that decides about fall or walk.

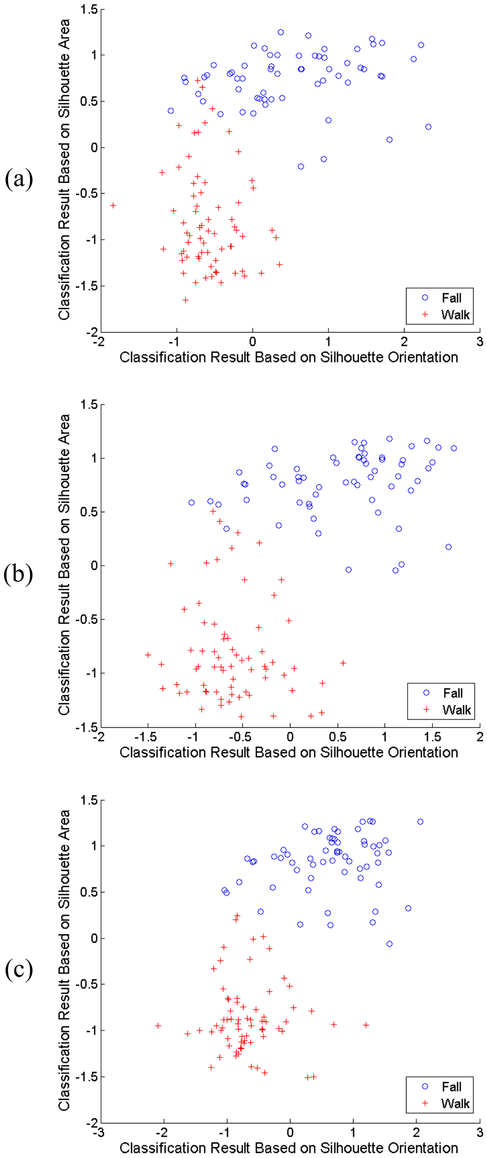


Figure 7. New samples formed from the labels produced by classifiers based on area and orientation for (a) linear kernel (b) RBF kernel (c) polynomial kernel

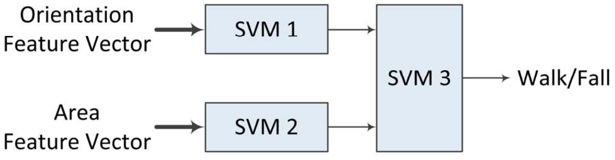


Figure 8. Proposed combined system for classification

We trained and tested our proposed system with 10-fold cross validation as mentioned before. Final results are shown in TABLE IV. As we see, classification results improved considerably. Especially with polynomial kernel we achieved error rates of less than 1%. In this case sensitivity of the system is 100% that means no single fall event is missed. Among 130 test samples only one false alarm was produced.

TABLE IV. CLASSIFICATION RESULTS USING COMBINED SYSTEM

SVM Kernel	Linear	Polynomial (m=2)	RBF ($\sigma = \pi$)
TP	60	61	59
TN	66	68	66
FP	3	1	3
FN	1	0	2
Sensitivity (%)	98.36	100.00	96.72
Specificity (%)	95.65	98.55	95.65
Accuracy (%)	96.92	99.23	96.15
Error Rate (%)	3.08	0.77	3.85

IV. CONCLUSION

Visual surveillance systems have advantages in automatic fall detection of elderly at home. These kinds of systems distinguish falls from other activities based on features extracted from image sequences. In this paper we exploited a drawback in background separation method and proposed variations in the silhouette area generated only by one camera as a simple feature that is not sensitive to the view point. Variations in the area can be a measure of rapid change of body posture, impact to a surface and inactivity after fall. Experimental results show that area has more sensitivity and specificity than orientation. Using area alone produced error rate of about 3% for the dataset that was experimented with, along with the sensitivity of more than 98%. We observed that classification results based on silhouette area and orientation can be nicely combined to reach the error rate of less than 1% and resulting in a sensitivity of about 100%.

REFERENCES

[1] Xinguo Yu, “Approaches and principles of fall detection for elderly and patient,” 10th International Conference on e-health Networking, Applications and Services, pp.42-47, 7-9 July 2008.

[2] Shaou-Gang Miaou, Pei-Hsu Sung, Chia-Yuan Huang, “A Customized Human Fall Detection System Using Omni-Camera Images and Personal Information,” 1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare, pp.39-42, 2-4 April 2006.

[3] B. U. Toreyin, Y. Dedeoglu, A.E. Cetin, “HMM Based Falling Person Detection Using Both Audio and Video,” Signal Processing and Communications Applications, pp.1-4, 17-19 April 2006.

[4] Huimin Qian, Yaobin Mao, Wenbo Xiang, and Zhiquan Wang, “Home environment fall detection system based on a cascaded multi-SVM classifier,” 10th International Conference on Control, Automation, Robotics and Vision, pp.1567-1572, 17-20 Dec. 2008.

[5] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, “Robust Video Surveillance for Fall Detection Based on Human Shape Deformation,” IEEE Transactions on Circuits and Systems on Video Technology, vol.21, no.5, pp.611-622, May 2011.

[6] E. Auvinet, F. Multon, A. St-Arnaud, J. Rousseau, and J. Meunier, “Fall Detection With Multiple Cameras: An Occlusion-Resistant Method Based on 3-D Silhouette Vertical Distribution,” IEEE Transactions on Information Technology in Biomedicine, vol.15, no.2, pp.290-300, Mar. 2011.

[7] Weiming Hu, Tieniu Tan, Liang Wang, S. Maybank, “A survey on visual surveillance of object motion and behaviors,” IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 34, no.3, pp.334-352, Aug. 2004.

[8] M. Piccardi, “Background subtraction techniques: a review,” IEEE International Conference on Systems, Man and Cybernetics, vol.4, pp. 3099- 3104, 10-13 Oct. 2004.

[9] Jozsef Valyon, “Extended LS-SVM for System Modeling,” Ph.D. dissertation, Department of Measurement and Information Systems, Budapest University of Technology and Economics, 2007.

[10] E. Auvinet, C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, “Multiple cameras fall dataset”, Technical report 1350, DIRO - Université de Montréal, July 2010.

[11] Multi camera fall dataset. (2010). [Online]. Available: http://vision3d.iro.umontreal.ca/fall-dataset/.

[12] N. Noury, A. Fleury, P. Rumeau, A.K. Bourke, G.O. Laighin, V. Rialle, and J.E. Lundy, “Fall detection - Principles and methods,” 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pp. 1663–1666, 22-26 Aug. 2007.