Fall Detection Using Gaussian Mixture Model and Principle Component Analysis

Arisa Poonsri and Werapon Chiracharit
Department of Electronic and Telecommunication Engineering
Faculty of Engineering
King Mongkut's University of Technology Thonburi
arisa.poonsri@mail.kmutt.ac.th

***Abstract*—Fall accident whose rates increase exponentially is the major risk for the elderly, especially those living alone. A fall accident detection system to detect the fall accident and call for an emergency is essential for elderly. This paper proposes to extract human from a video camera using a mixture of Gaussian model combined with average filter models. The proposed method extracts six postures of physically movements of human including lying, sitting, standing, getting up, walking, and falling. Unique features such as inter-frames information, shape description from a silhouette aspect ratio, and orientation of principal component are obtained. The method could automatically alarm when the fall is detected. The experimental results show the detection rate up to 86.21% of the 58 videos from the Le2i dataset.**

Keywords—Elderly Care; Fall Detection; Principal component analysis; Gaussian mixture model

I. INTRODUCTION

Falls in the elderly are a significant cause of morbidity and mortality. According to the World Health Organization global reports on Falls Prevention in Older Age, fatal falls rates exponentially increased with age for both genders, highest at the age of 85 years and over [1]. The fall prevention and a calling for emergency system, which can automatically detect and alarm human falls, is imperative especially those living alone. In the population aging, many researchers have made the topics of fall detection which are based on wearable devices and video surveillance. Some have attempted to detect the falls using a variety of sensors such as cameras, accelerometers, gyroscopes, microphones, and so forth. They use not the only standalone sensor but also a combination of them. However, users could feel uncomfortable after wearing the devices in a long time besides its battery consuming. The video information problem can cause the exposure of personal privacy as well. This issue can be avoided by not directly using the raw videos, but extract any numerical features where the personal identity is not revealed from it.

The existing methods can be classified roughly into two kinds of fall detection technique including Wearable device based approach, in which sensors are connecting the body of subject to find its location or motion for example accelerometer and posture sensor as well as estimated changing of blood pressure and pulse rates can determine human activities [2]. The other type is non-wearable device based approach that eliminating compliance issues because they are always on; there is no need to charge devices or to wear something on. The vision-device based approach takes advantage of cameras, positioned on overhead positions, to monitor as well as characterize a person's movement and test the incidence associated with falls [3].

Recently, the popular research topic in computer vision field is vision-based fall detection which doesn't need to wear any system on the human body. The most efficient and popular types of sensors in this area divide into two groups which are single/multiple RGB cameras [4-7] and camera combined with another sensor [8-10]. The reviews of conventional methods are presented as follows.

Yixiao et al. [11] proposed human fall detection in RGB videos by fusion of statistical shape features, velocity, and motion dynamics on Riemannian manifolds. This technique is more efficient and used instead bounding box, which is the most commonly used in other state-of-the-art research. This study tested its performance with three datasets and discussed the method of which can use images showing the whole.

Chamle et al. [12] presented automatic unusual event detection in video surveillance to classify fall and non-fall event. The rectangular and elliptical bounding box is created from marking objects. Gradient boosting classifier is deployed to classify the fall from the features: aspect ratio, fall angle, and silhouette height. This method gives 79.31% accuracy for fall detection testing in Le2i datasets.

Wang et al. [13] researched about automatic fall detection of human in RGB video using a combination of features. The method proposes HOG (Histogram of Oriented Gradient), LBP (Local Binary Pattern), and feature extracted by the Deep Learning Framework Caffe.

Liu et al. [14] detected fall by using the human body silhouette to improve privacy and vertical projection histograms and statistical scheme to reduce human body upper limbs activities. The kNN classification algorithm is used to classify postures using the ratio and difference of human body silhouette, bounding box height, and width. Even though they demonstrate their effectiveness, these works are based on the presumption that the lighting conditions stay relatively uniform which is not always true in daily life.

Kumar et al. [15] presented a method to classified fall and non-fall cases by combining features extracted from RGB-D video. Instead of using bounding box, they obtained shape, and motion features from target contours combining with HOG and HOGOF encoded features. The limitation of this work is about human extraction such as lying down activities can appear quite confusing in comparison with human falls causing overall performance degradation so, video segmentation was manually chosen instead of automatically done .

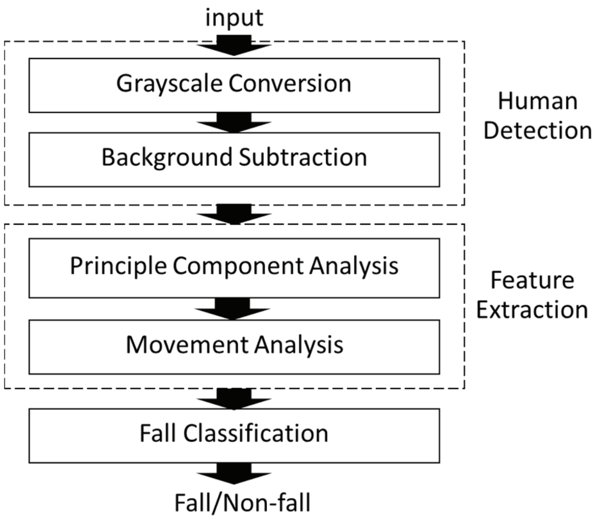

Fig. 1. Fall detection block diagram

The problem summary of previous works divides into two groups by results, including accuracy and processing time. The problem caused decreasing efficiency is human detection or posture classification. Some methods cannot completely describe human poses especially with falling and laying that decreased human detection efficiency. The robustness of detection depends on occlusion, camera location changing, environment, and the light condition is also causing several problems of many state-of-the-art methods.

In this paper, we purpose the fall detection using a mixture of Gaussian model and principal component analysis. This approach includes three main steps: 1) Human detection, 2) Feature extraction, and 3) Fall classification. The first step is human detection we propose background subtraction using a mixture of Gaussian models (MoG) combined with average filter model to implement the subtraction results. In feature extraction section, the orientation, aspect ratio and area ratio are calculated from the Principal Component Analysis (PCA) of a human silhouette. All selected elements are used in posture classification which predicts the falls and non-fall events.

II. Proposed Method

A. Dataset

The videos from a database Le2i [16] is chosen for the proposed method. The frame rate is 25 frames/s, and the resolution is 320x240 pixels. The video sequences include a variety of illumination and normal difficulties such as occlusions and textured background. The actors performed various normal activities and fall in different locations; coffee room, office, lecture room, and home.

B. Human Detection

After reading the input video file, the individual frames are extracted using MATLAB from VideoReader function. The video frame is converted to grayscale using the rgb2gray for background subtraction process.

We can detect a human from video by detecting a moving object. A commonly used technique is to use background subtraction, which is a method used in moving regions segmentation in image sequences taken from a static camera by comparing each new frame to a model of the scene background. A mixture of Gaussian model (MoG) is applied to create background subtraction process in which the background model is parametric and not an actual environment [17].

A mixture of Gaussian model (MoG) [18] is a probabilistic model for representing the presence of subpopulations within an overall population. We can create background model to detect moving texture from the mixture of the Gaussian model method, the probability of occurrence from a color of pixel *s* is given by equation (1):

$$P\left(I_{s,t}\right)=\sum_{i=1}^{K} \omega_{i,s,t} N\left(\mu_{i,s,t},\Sigma_{i,s,t}\right) \tag{1}$$

Where N (μi,s,t, $\sum$i,s,t) is the i^{th} Gaussian model and $\omega_{i,s,t}$ is weight. Stauffer and Grimson suggest about the computational methods, the covariance matrix $\sum_{i,s,t}$ can be assumed to be diagonal, $\sum = \sigma^2 Id$. In their method, parameters of the matched component are updated following equation (2), (3) and (4) :

$$\omega_{i,s,t}=(1-\alpha)\omega_{i,s,t-1}+\alpha \tag{2}$$

$$\mu_{i,s,t}=(1-\rho).\mu_{i,s,t-1}+\rho I_{i,s,t} \tag{3}$$

$$\sigma^2_{i,s,t}=(1-\rho).\sigma^2_{i,s,t-1}+\rho.d_2(I_{s,t},\mu_{i,s,t}) \tag{4}$$

Parameters μ and σ of unmatched distributions remain the same while their weight is reduced as follows: $\omega_i = (1-\alpha)\omega_{i,s,t-1}$ to achieve decay. To determine which components are part of the background model, once every Gaussian has been updated, the K weights $\omega_{i,s,t}$ are normalized so they sum up to 1. Then, the K distributions are ordered based on a fitness value $\omega_{i,s,t}/\sigma_{i,s,t}$ and only the H most reliable ones are chosen as part of the background shown in equation (5):

$$H = \text{argmin}\left(\sum_{i=1}^{h} \omega_{i} > \tau\right) \tag{5}$$

Where τ is a threshold.

Then, those pixels whose color $I_{s,t}$ is located at more than 2.5 standard deviations away from every H distributions are labeled "in motion." So, we can extract a human part from the frames. We can improve the quality of the detected objects and remove the false detection from morphological operations, but there still has some wrong detection so we can apply the means filter to support its effectiveness. We create the background model from the average of video frames for each environment so that it gives the closest background to extract human from each frame.

$$B(x,y,t) = \frac{1}{n} \sum_{i=0}^{n-1} I(x,y,t-i) \qquad (6)$$

We apply the human detection results from merging MoG and mean filter results altogether. An example result of human detection is shown in Fig.2 below.

Fig. 2. Human detection results

C. Feature Extraction

In this process, we calculate features from human silhouette on the main directional axis plane from the Principal component analysis (PCA), which obtains the information to create features such as orientation, aspect ratio, and inter-frames information will be used in the movement analysis.

PCA is the mathematical method used to reduce the number of features and used to represent data including the dimensionality reduction, which provides a simpler representation of the data, reduction in memory, and faster classification. The major axis of the main directional axis is used to identify human shape from the direction. In this step, the covariance matrix Σ is computed, according to the formula in the equation. (7), where μ denotes the average value, and X and Y are the distributions of the pixels in x and y directions, respectively. This matrix equals:

$$\Sigma = \begin{bmatrix} E((X - \mu_x)(X - \mu_x)) & E((X - \mu_x)(Y - \mu_y)) \\ E((Y - \mu_y)(X - \mu_x)) & E((Y - \mu_y)(Y - \mu_y)) \end{bmatrix} \qquad (7)$$

The orientation of the principal axis can be obtained by a singular value decomposition of the covariance matrix Σ, decomposing Σ into the matrix product Σ = USV'. The angle between the main axis and the Y-axis (φ) (The PCA axis is shown in fig.(3) is calculated as:

$$\varphi = \tan^{-1}\left(\frac{V(1,1)}{V(1,2)}\right) \qquad (8)$$

An aspect ratio is calculated to describe the proportional relationship between width and height for different postures.

$$\text{Aspect ratio} = \frac{\text{Number of pixels on major axis}}{\text{Number of pixels on minor axis}} \qquad (9)$$

The obtained features will be used to classify fall and non-fall. In the case of fall, an absolute angle should be large, as well as lying. The inter-frame information will be calculated to classify human movement.

D. Fall Classification

From the previous section, the calculated features are used to classify fall and non-fall events. We classify the possible events into two conditional groups of, static group and dynamic group. The condition of the static group includes standing sitting lying (or after fall). In this section, we consider from the orientation and aspect ratio if there are in the same range of activity at the same time (the frame rate is 25 frame per 1 second in this dataset). The dynamic group includes walking, getting up (both of getting up from lying to sit and from sit to stand up), and falling which is focused in our proposed method. We set the fall confirmation using the orientation range (In this case, we set the inclined threshold for 30 and 120 degrees). For the moving object can be classified from the dynamic properties using the difference of human centroid distance.

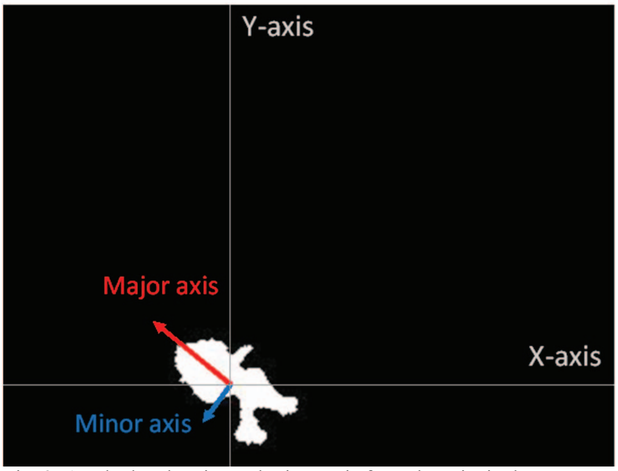

Fig. 3. A calculated major and minor axis from the principal component analysis (PCA)

III. Experimental Results

Evaluating of the method efficiency, we calculated the fall detection rate (DR: sensitivity: the calculation of the fall events related to all events) and false alarm rate (FR: 1-specificity: the calculation of the all events excepting the true negative (TN) events) on the Le2i dataset which includes 191 videos in the different environments. The qualitative analysis of the proposed system explains in the table I below.

Table I the qualitative analysis of the proposed system

	Condition: Fall	**Condition: No fall**	
Test outcome: Fall	TP	FP	Accuracy
Test outcome: No fall	FN	TN	Error
	Sensitivity $Se = \frac{TP}{TP+FN}$	Specificity $TN = \frac{TN}{TN+FP}$	

Where,

- True positive (TP) is a fall occurs, and the system detects it.
- False positive (FP) is the system announces a fall, but it did not happen.
- True negative (TN) is a non-fall event, and movement is performed, the system does not declare a fall.
- False negative (FN) is a fall has occurred, but the system does not detect it.

M.Chamle et al. [12] tested their proposed method on fifty-eight videos from Le2i database; it gave an accuracy of 79.31%. To compare with their method, we test on the same datasets; the validation is shown in Table II.

Table II the comparison of the qualitative between the previous work and the proposed method tested on 58 videos from Le2i dataset

Method	TP	TN	FP	FN	Se (%)	Sp (%)	Acc (%)	DR (%)	FR (%)
Proposed	41	9	5	3	93.18	64.29	86.21	93.18	35.71
M. Chamle [12]	27	19	7	5	83.47	73.07	79.31	84.37	26.92

The validation shows the proposed method provides good results up to 86.21% accuracy on 58 videos from the Le2i dataset. However, there is 13.79% of false detection occurred in some specific environments because of the MOG technique is very sensitive to detect all of the moving objects. According to the Le2i dataset, the user is dragging a chair, holding a large object makes human shape changed. So, the aspect ratio will be out of range from the realistic poses. We can improve this occurrence by using color recognition technique to recognize user's color identify users and remove the shadow which contains noises or make a human shape changing.

IV. Conclusion

In this paper, we proposed the fall detection using fall detection using a mixture of Gaussian model and principal component analysis. This method is proposed to extract human from a video camera. The algorithm extracts six postures of human physically movements include lying, sitting, standing, getting up, walking, and falling from unique features such as inter-frames information, shape description is shown by silhouette aspect ratio and orientation of principal component analysis, etc. The validation shows that the proposed method provides good results up to 86.21% on the 58 videos from the Le2i dataset. The future work is to improve the robustness of another activity by using another feature such as the velocity and acceleration of the object etc.

References

[1] World Health Organization, 2008, WHO global report on falls prevention in older age, pp.2

[2] M. Mubashir, L. Shao, and L. Seed, 2013, "A survey on fall detection: Principles and approaches," Neurocomputing, 100, 144–152.

[3] R. Kaur, and P. Deep Kaur, 2016, "Review of Fall Detection Techniques based on Elder People," Journal, Advanced Research in Computer Science, 8, 3, pp: 1062-1067

[4] N. Schräder, 2011, "Detecting Falls and Poses in Image Silhouettes," Master's Thesis, Chalmers University of Technology, 1

[5] D. Anderson, R.H. Luke et al., 2009,"Linguistic Summarization of Video For Fall Detection Using Voxel Person and Fuzzy Logic," Emergency System for the Elderly, Computer Vision and Image Understanding., 1, 113, pp: 80–89

[6] D. Anderson, R. H. Luke, Jet al, 2009, "Modeling Human-Activity from Voxel Person Using Fuzzy Logic," Fuzzy Systems, 17, 1, pp: 39–49

[7] Rougier C., St-Arnaud, A., Rousseau, J. and Meunier, J., "Video Surveillance for Fall Detection," Video Surveillance, pp. 357-383

[8] Bian, Zhen-Peng, et al., 2015, "Fall detection based on body part tracking using a depth camera," IEEE Journal of biomedical and health informatics, vol.19.2, pp: 430-439.

[9] X. Ma, H. Wang, B. Xue, M. Zhou, B. Ji, and Y. Li, 2014, "Depth-based Human Fall Detection via Shape Features and Improved Extreme Learning Machine," IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 6, pp. 1915-1922.

[10] S. Kido, T. Miyasaka, T. Tanaka, T. Shimizu and T. Saga, 2009, "Fall Detection in toilet rooms using thermal imaging sensors," IEEE/SICE International Symposium on System Integration (SII), Tokyo, pp. 83-88.

[11] Y. Yun, and I. Yu-Hua Gu, "Human fall detection in videos by using statistical features of shape and motion dynamics on Riemannian manifolds," Neurocomputing, Volume 207, pp: 726-734

[12] M. Chamle, K. G. Gunale, and K. K. Warhade, 2016, "Automated unusual event detection in video surveillance," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, pp. 1-4.

[13] K. Wang, G. Cao, D. Meng, W. Chen and W. Cao, 2016, "Automatic Fall Detection of human in the video using a combination of features," IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Shenzhen, pp. 1228-1233.

[14] H. Liu and C. Zuo, "An improved algorithm for automatic fall detection," AASRI Procedia, vol. 1, pp. 353–358, 2012.

[15] D. P. Kumar, Y. Yun and I. Y. H. Gu, "Fall Detection in RGB-D videos by combining shape and motion features," 2016, International Conference on Acoustics Speech and Signal Processing, Shanghai, pp. 1337-1341.

[16] I. Charfi, J. Mitran, J. Dubois, M. Atri, and R. Tourki, "Optimised spatiotemporal descriptors for real-time fall detection: comparison of SVM and AdaBoost based classification," Journal of Electronic Imaging (JEI), Vol.22. Issue.4, pp.17, October 2013.

[17] T. Bouwmans, F.El Baf, and B. Vachon, 2008, "Background Modeling using Mixture of Gaussians for Foreground Detection," Bentham Science Publishers, pp.219-237.

[18] C. Stauffer and W.E.L. Grimson, 2007, "Adaptive Background Mixture Models for Real Time Tracking," Journal, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2.