Improvement of Fall Detection Using Consecutive-frame Voting

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*— The Centers for Disease Control and Prevention (CDC) reported the older adult statistics that in every second there is an older adult fall down, 25% of elderly reported a fall in 2014, and it is the first cause of hip fracture in the USA. A fall accident detection system, which can automatically detect the fall accident and call for help, is essential for elderly. This paper proposes Improvement of Fall Detection Using Consecutive-frame Voting. 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. The moving object can be classified from the human centroid distance in human centroid tracking section. Each posture will be classified in event classification. Finally, majority voting of the results from consecutive is finally performed. The experimental results show improvement of the accuracy of the proposed method with our previous work which tested on the Le2i dataset.**

Keywords—Elderly care; Fall Detection; Principal Component Analysis; Consecutive-frame Voting

I. Introduction

Falls in the older adults are risky especially in whose living alone. From The Centers for Disease Control and Prevention (CDC) reported the statistics of fall in elderly. In 2014, there is 25% of elderly in the United States of America were reported a fall accident. CDC also reported that an older adult falls in every second every day [1]. Nowadays, there are many ways to prevent a fall. Fall detection system is crucial for elderly especially for who lives alone.

The existing fall detection system can be classified roughly into two based approaches including Wearable device in which sensors are connecting the body to find a location or motion, and determine human activities. The other type is a non-wearable device-based approach that eliminating compliance issues; there is no need to charge devices or to wear something on. The vision-device based takes advantage of cameras to monitor as well as characterize a person's movement and test the incidence associated with falls.

Recently, the popular research topic in computer vision is vision-based fall detection which is using single or multiple RGB cameras and camera combined with another sensor. The reviews of conventional methods are presented as follows.

Yixiao et al. [2] 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, the performance tested with three datasets and discussed the method of which can use images showing the whole body only.

Chamle et al. [3] 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.

Kumar et al. [4] 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.

Gnouma et al. [8] tested a fall detection in single surveillance camera. A Block Matching motion estimation, acceleration, and changes of the human body silhouette area were proposed to distinguish the fall events in normal daily activities. This method is able to detect falls even with a realistic and challenging videos, but requires an information relating to the light intensity. Another thing is difficult to distinguish between falls and jumping.

The problem summary of previous work divides into two groups by results, including accuracy and processing time. The problem caused decreasing accuracy is human detection or posture classification. Some methods cannot completely describe human postures. 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.

Our previous work [5] proposed the Fall Detection Using Gaussian Mixture Model and Principle Component Analysis to increase an accuracy of detection in Le2i dataset, the human detection technique was sensitive to detect fall in every frames because of the changing of aspect ratio and angle of major axis caused a false detection. In this work, we propose an

978-1-5386-2615-3/18/$31.00 ©2018 IEEE

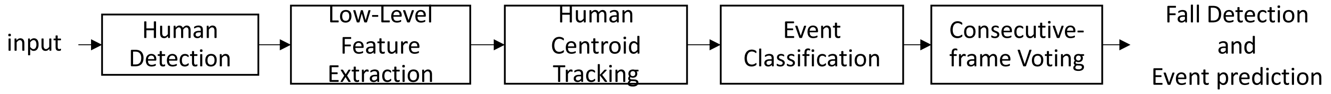

Fig.1. Fall detection block diagram

improvement of fall detection using consecutive-frame voting to improve our previous work accuracy. Prediction of the fall and events, the method consists of five stages: 1) Human Detection, 2) Low-Level Feature Extraction, 3) Human Centroid Tracking 4) Event Classification, and 5) Consecutive-frame voting. 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 are calculated from the Principal Component Analysis (PCA) of a human silhouette. The movement tracking can be classified from the dynamic properties that is the range of human centroid distance between frames. Each posture will be classified in event classification section. Last step, the consecutive-frame voting is proposed to increase an accuracy of prediction. The proposed method block diagram is shown in Fig.1.

II. Proposed Method

A. Dataset

The videos from a database Le2i [6] 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 general 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 frame is then 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 background [7].

A mixture of Gaussian model (MoG), which 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:

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

Where N (μi,s,t, ∑i,s,t) is the ith 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 as follows [8]:

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

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

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

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 :

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

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. Improvement of the human detection is to remove the noise 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 and merge the results of MoG and mean filter altogether. Then the noise removal is done by morphological operation *imclose*. The example results of merging is shown in fig. 2.

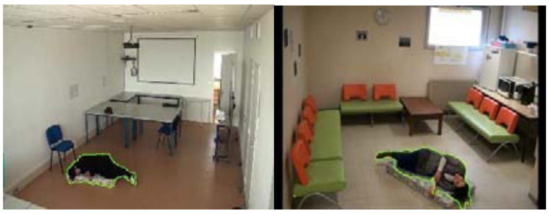

Fig.2. the example results of merging in human detection

C. Low-Level Feature Extraction

This process, we calculate features from human silhouette on the central 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 more straightforward representation of the data, reduction in memory, and faster classification. The major axis of the central 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. (6), 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:

$$\Sigma = \begin{bmatrix} E((X - \mu_x)(X - \mu_x)) & E((X - \mu_x)(Y - \mu_y)) \\ E((Y - \mu_y)(X - \mu_x)) & E((Y - \mu_y)(Y - \mu_y)) \end{bmatrix} \tag{6}$$

The orientation of the main 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) is calculated as:

$$\varphi = \tan^{-1}\left(\frac{V(1,1)}{V(1,2)}\right) \tag{7}$$

To describe the proportional relationship between width and height for different postures, after we rotated the human silhouette into major axis in φ degree by using function *imrotate* in MATLAB, an aspect ratio is calculated.

$$\text{Aspect ratio} = \frac{\text{Number of pixels on major axis}}{\text{Number of pixels on minor axis}} \tag{8}$$

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.

D. *Human Centroid Tracking*

For the moving object can be classified from the dynamic properties using the difference of human centroid distance between frames. Which is calculated by:

$$Distance = \sqrt{(X2 - X1)^2\,(Y2 - Y1)^2} \tag{9}$$

Where X1, and Y1 are the position of human centroid at the previous frame. X2, and Y2 are the position of human centroid at the recent frame.

E. *Event Classification*

In this step, there are three features for classifying so, we will classify each posture from rule-based classification. The conditions of rule-based for statics and dynamics in video surveillance are shown in Table I. 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. The dynamic group includes walking, getting 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 145 and 190 degrees from major axis to the x-axis. The aspect ratio should be in the range of 0.8 to 2.6 from height per width. The distance of centroid was set to classify a dynamic group, the distance should be more than 1.

Table I the condition of rule-based for statics and dynamics in video surveillance

Features	Standing	Sitting	Lying	Walking	Getting Up	Falling
Orientation (degree)	70-110	70-110, 190-270	145-190	10-90	70-110, 190-270	145-190
Aspect Ratio (height/width)	1.2-3	0.7-1.6	0.8-2.6	1.2-3	0.7-1.6	0.8-2.6
Distance (in 25 frames)	-	-	-	>1	>1	>1

E. *Consecutive-frame Voting*

Because of the wrong prediction in some sequence decreased the detection rate, the consecutive-frame voting is used for increasing an accuracy of prediction. We consider from prediction results from the previous section, if the frame rate is 25 frame per a second in this dataset, we will find the maximum number of prediction in 25 frames or one second. To make it clearer understanding we will describe in (10) and (11) below.

$$Prediction_{new} = MAX\,(X_{Prediction}) \tag{10}$$

$$X_{Prediction} = \{N_{stand}, N_{sit}, N_{lying}, N_{walk}, N_{get\,up}, N_{fall}\} \tag{11}$$

Which $X_{Prediction}$ is a set of numbers of the prediction with $N_{stand,\,sit,\,lying,\,\ldots}$ is a number of prediction which shown “standing, sitting, lying, etc.” in the sequence from n to 25n, with n is a number of sequence which is defined from sequence= loop/25.

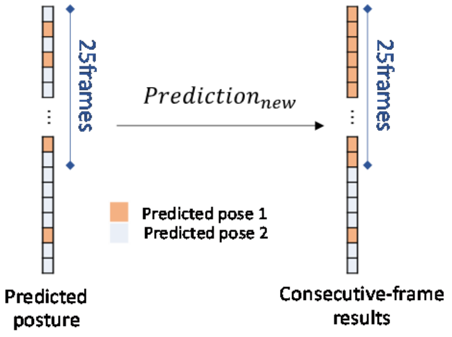

Fig.3. Description of Consecutive-frame Voting

In this step, in every 25 previous frames will be changed as a new prediction ($Prediction_{new}$) which is a maximum numbers of predicted-posture in each sequence. For easily understanding, the description is shown in fig.3.

III. Experimental Results

To evaluate the efficiency of our proposed method, 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 below.

- 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.

Our previous work [5] compared to Chamle et al. [3] proposed method on 58 selected videos from *Le2i* database, the selected videos consist of 20 video from *Lecture room*, 15 videos from *Coffee_Room_01*, 15 videos from *Coffee_Room_02*, and 8 videos from *Office*. We gave an accuracy of 86.21% in the previous experiment which is more than Chamle et al [3]. whose gave 79.31%. By the way, there was still below 90%. Comparing with the previous work with the same datasets our proposed method has been improved the validation as shown in Table II.

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

Method	TP	TN	FP	FN	Se (%)	Sp (%)	Acc (%)	DR (%)	FR (%)
Chamle et al. [3]	27	19	7	5	83.47	73.07	79.31	84.37	26.92
Our Previous work [5]	41	9	5	3	93.18	64.29	86.21	93.18	35.71
Proposed Method	**47**	**4**	**6**	**1**	**97.92**	**60.00**	**91.38**	**97.92**	**40**

From the results, our method achieved the highest accuracy up to 91.38% compared with our previous work [5] and Chamle et al [3] method. But it still has 8.62% of false detection. The false detection occurred from the aspect ratio is out of range from the realistic poses because of variation of the environment in this dataset, for example, the user is dragging a chair caused the wrong detection that is human and a chair are detected. The moving shadow on the wall which appeared in *Lecture Room* videos from Le2i dataset because of the light outside the room is brighter than inside. So, an algorithm detects the darken shadow as a human silhouette which is caused the wrong detection.

IV. Conclusion

In this paper, we propose an improvement of fall detection using consecutive-frame voting to improve the accuracy of our previous work. 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 low-level feature extraction section, the orientation, aspect ratio and area ratio are calculated from the Principal Component Analysis (PCA) of a human silhouette. The human centroid tacking can be classified from the dynamic properties using the difference of centroid distance. Each posture will be classified from rule-based classification since there are three features using inter-frame human centroid distance feature. The last step the consecutive-frame voting is proposed to increase an accuracy of prediction. The validation shows that the proposed method gives a good results up to 91.38% on the 58 videos from the Le2i dataset. The future work is to test with another dataset with different viewpoint.

References

[1] National Center for Injury Prevention and Control. Preventing Falls: A Guide to Implementing Effective Community-based Fall Prevention Programs. 2nd ed. Atlanta, GA: Centers for Disease Control and Prevention, 2015.

[2] Yixiao Yun, Irene Yu-Hua Gu, “Human fall detection in videos by fusing statistical features of shape and motion dynamics on Riemannian manifolds,” Neurocomputing, Volume 207, pp: 726-734

[3] 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.

[4] 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 (ICASSP), Shanghai, pp. 1337-1341.

[5] A. Poonsri, and W. Chiracharit,” Fall Detection Using Gaussian Mixture Model and Principle Component Analysis,” 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE), Thailand, 2017, pp. 1-4.

[6] T. Bouwmans, F. El Baf, B. Vachon, 2008, “Background Modeling using Mixture of Gaussians for Foreground Detection,” Journal, Recent Patents on Computer Science, Bentham Science Publishers, 1 (3), pp.219-237.

[7] C. Stauffer and W.E.L. Grimson, “Adaptive Background Mixture Models for Real-Time Tracking,” Conference Paper in Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2:252, Vol. 2, February 1

[8] Mariem Gnouma et al., 2017, “Human fall detection based on block matching and silhouette area,” Conference Paper, Ninth International Conference on Machine Vision (ICMV 2016), Tunisia, Proceedings Volume 10341