Research Journal of Applied Sciences, Engineering and Technology 4(24): 5361-5366, 2012
ISSN: 2040-7467

Submitted: March 18, 2012 Accepted: April 26, 2012 Published: December 15, 2012

RVM Based Human Fall Analysis for Video Surveillance Applications

B. Yogameena, G. Deepika and J. Mehjabeen
ECE Department, Thiagarajar College of Engineering and Technology Madurai,
Tamil Nadu, India

Abstract: For the safety of the elderly people, developed countries need to establish new healthcare systems to ensure their safety at home. Computer vision and video surveillance provides a promising solution to analyze personal behavior and detect certain unusual events such as falls. The main fall detection problem is to recognize a fall among all the daily life activities, especially sitting down and crouching down activities which have similar characteristics to falls (especially a large vertical velocity). In this study, a method is proposed to detect falls by analyzing human shape deformation during a video sequence. In this study, Relevance Vector Machine (RVM) is used to detect the fall of an individual based on the results obtained from torso angle through skeletonization. Experimental results on benchmark datasets demonstrate that the proposed algorithm is efficient. Further it is computationally inexpensive.

Keywords: Fall detection, Gaussian Mixture Model (GMM), Relevance Vector Machine (RVM), torso angle, video surveillance

INTRODUCTION

Many older persons fall and are not able to get up again. The lack of timely aid can even lead to more severe complications. Although not all falls lead to physical injuries, psychological consequences are also important, leading to fear of falling, losing self-confidence and fear of losing independence. The existing technological detectors are mostly based on wearable sensors. Most fall detection techniques are based on accelerometers or help buttons. But the major problem with these types of technology is that older people often forget to wear them and in the case of a help button, it is useless if the person is unconscious after the fall. Also during housekeeping tasks the sensors are removed, to prevent false alarms due to the needed sensitivity. The devices are using battery power, so no alarm will be generated if the batteries are depleted. They are also sometimes removed when the person finds them uncomfortable. If a fall occurs at these moments, it will not be detected.

LITERATURE REVIEW

A number of video surveillance systems for towards fall detection and abnormal detection have been reported. Among fall detection methods, one of the simplest and commonly used techniques is to analyze the bounding box representing the person in the image that was used by Toreyin *et al*. (2005) and Anderson *et al*. (2006). However, this method is efficient only if the camera is placed sideways and can fail because of occluding objects. For more realistic situations, the camera has to be placed higher in the room to avoid occluding objects and to have a larger field of view.

Lee and Mihailidis (2005) detected falls by analyzing the silhouette and the 2-D velocity of the person, with special thresholds for inactivity zones like the bed. Nait-Charif and McKenna (2004) tracked the person using an ellipse and analyzed the resulting trajectory to detect inactivity. However, (Lee and Mihailidis, 2005; Nait-Charif and McKenna, 2004) used a camera mounted on the ceiling and therefore did not have access to the vertical motion of the body, which provides useful information for fall detection. The 2-D (image) velocity of the person has also been used to detect falls by (Lee and Mihailidis, 2005; Sixsmith and Johnson, 2004). However, a problem with the 2-D velocity is that it is higher when the person is near the camera, so that the thresholds to discriminate falls from a person sitting down abruptly, for instance, can be difficult to define. Rougier *et al*. (2006) proposed an algorithm that used simple shape analysis and in Rougier *et al*. (2007) used head tracking for detecting fall.

Caroline *et al*. (2011) used an elaborate shape analysis based on person's silhouette and Procrustes distance. Thome *et al*. (2008) proposed a method that

Corresponding Author: B. Yogameena, ECE Department, Thiagarajar College of Engineering and TechnologyMadurai, Tamil Nadu, India

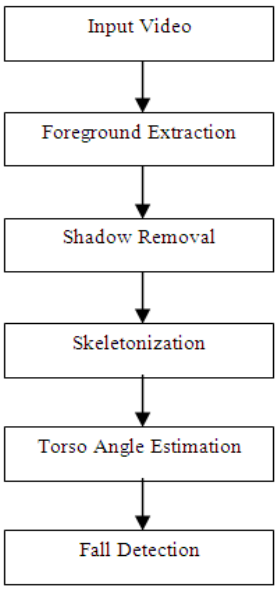


Fig. 1: Proposed method

used Markov model to distinguish falls from walking activities. The features used for motion analysis were extracted from a metric image rectification in each view. Anderson *et al*. (2009) analyzed the states of a voxel person obtained from two cameras and fall detection was achieved with a fuzzy hierarchy. Auvinet *et al*. (2008) proposed to exploit the reconstructed 3-D silhouette of an elderly person for fall detection. An alarm was triggered when most of this volume was concentrated near the floor. False alarm rate increases in case of normal activity. The objective of this study is to develop an algorithm that detects the fall based on leg angle and torso angle.

METHODOLOGY

Algorithm overview: The overview of the algorithm is shown in Fig. 1. The first stage of the surveillance system is foreground extraction. This blob detection subsystem detects the foreground pixels by subtracting the background that is modeled using Gaussian Mixture Model. Then, shadow removal and morphological operations are done to enhance the image. The second stage is the star skeletonization that is done for leg angle and torso angle determination. In this study, Gaussian Modelling of shadow is developed by extracting maximal Gabor responses in the optimal color space where camera is assumed to be static.

Background subtraction and shadow removal: In the first stage, the background is modelled using Gaussian Mixture Model (GMM) and the foreground pixels are detected from the background model. Then, these pixels are separated as subject regions and rough shadow regions using projection histogram based approach (Ibrahim and Anupama, 2005). The pixels belonging to shadow region are classified as significant and insignificant and the suitable color space has been assigned using optimal color space rule. The rule for the classification of shadows (Shan *et al*., 2007) is given by Eq. (1), (2) and (3):

$$D_s(x,y) = \begin{Bmatrix} 1, I_0(x,y) - B_i(x,y) \geq Th_s \\ 0, otherwise \end{Bmatrix} \quad (1)$$

$$D_l(x,y) = \begin{Bmatrix} 1, I_0(x,y) - B_i(x,y) \geq Th_l \\ 0, otherwise \end{Bmatrix} \quad (2)$$

$$\varepsilon = \Sigma D_l / \Sigma D_s \quad (3)$$

where, D_s and D_l are intensity differences between original and background images, Th_s, Th_l are thresholds of insignificant and significant shadows respectively and, $Th_s<Th_l$, I_0 is the original image, B is the background image, ε is the shadow classifier.

Based on the pixel being classified as significant or insignificant, the appropriate color space is chosen before modeling the shadow pixels. Once the shadow region and corresponding color space have been determined the Gabor filter kernels are used to build a proper shadow model to model the pixels as a shadow pixel. The Gabor filters have received considerable attention in image processing applications because they posses optimal localization properties in both spatial and frequency domains. Therefore Gabor functions are ideal for accurate texture identification of the surface which is related with shadow (Caleanu *et al*., 2007). Also Gabor filter response can be represented as a sinusoidal plane of particular frequency and orientation, modulated by a Gaussian envelope (Caleanu *et al*., 2007). Considering the advantages of Gabor filters which include the robustness to illumination changes, multi-scale and multi-orientation nature, the following form of a normalized 2D Gabor filter function in the frequency domain is employed for the extraction of maximal response of spatial frequency and orientation for the shadow pixels as in Eq. (4) and (5):

$$\psi(x,y,f,\theta) = \exp\left(-\frac{x^2+y^2}{\sigma^2}\right)\exp(2\pi i x) \quad (4)$$

$$x' = x\cos(\theta) + y\sin(\theta) \quad (5)$$

where, ψ is the Gabor kernel, f is the spatial frequency, θ is the orientation,σ is the standard deviation of the Gaussian kernel and it depends upon the spatial frequency. The responses of convolutions in a shadow region at location (x, y) are given in spatial domain as in Eq. (6) and it represents important features for the shadow pixels:

$$r_{\xi}(x, y, f, \theta) = \psi(x, y: f, \theta) * \xi(x, y)$$
$$= \int_{-\alpha}^{\alpha} \int \psi(x - x_x, y - y_x: f, \theta)\xi(x_x, y_x)dx_x dy_x \quad (6)$$

where, $\xi(x, y)$ is the shadow pixel in region and $r_{\xi}(x, y, f, \theta)$ represents the Gabor response. For each shadow pixel, Gabor filters with multi-scale and multi-orientation are first used to extract Gabor features from input frames. Here, the magnitude of complex outputs of Gabor convolutions are used as features. The maximal response of the pixels in the shadow region is obtained by multiplying it with Gabor kernel in frequency domain.

These maximal Gabor responses are modeled as Gaussian since the original shadow data is random variable and subsequently these responses are also random and large in size. Once these responses are assumed as Gaussian, the likelihood is derived as in Eq. (7):

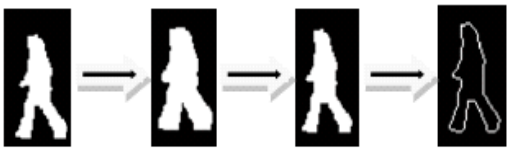
$$L_1(i, j) = \exp(-0.5(X-m_f)^T C^{-1}(Y-m_\theta) \quad (7)$$

where, $L_1(i, j)$ is the proposed Gaussian shadow model for the Gabor response of the shadow pixels, X and Y are the Gabor response obtained by varying the spatial frequency, f and orientation θ, respectively, m_f is the mean value of the Gabor filter response when θ is kept constant , m_θ is the mean value of the Gabor filter response when f is kept constant and C is the covariance of both responses. To eliminate the shadows completely, the Gaussian model fitted earlier is used as a tool on the currently available pixels representing the moving object along with their shadow. These pixels are subjected to a Gabor filtering action as developed earlier in modeling with the filter function and the Gabor response is represented as F_g.

The mean spatial frequency m_f and the mean orientation m_θ which were responsible for producing maxima during modeling are used to produce Gabor responses. The likelihood function of the response is given by Eq. (8):

$$L_2(I, j) = \exp(-0.5(F_g(i, j)-m_f)^T C^{-1}(F_g(i, j)-m_\theta) \quad (8)$$

where, $L_2(i, j)$ is the likelihood response due to Gabor action. Therefore a foreground pixel in the current frame convolved with Gabor giving response is subjected to a likelihood estimation resulting in a likelihood value. This shadow likelihood is thresholded to decide whether the current pixel is a shadow or foreground. Though many thresholding schemes are available, the threshold (th_s), which is the mean value of the Gabor responses obtained from modeling, is considered as threshold, since the likelihood estimation is assumed as Gaussian which follows non-skew distribution.



Foreground of moving target | Dilation | Erosion | Border extraction

Fig. 2: Pre-processing for skeletonization

Star skeletonization: There will be spurious pixels detected, holes in moving features, "inter-lacing" effects from video digitization processes and other anomalies. Foreground regions are initially filtered for size to remove spurious features and then the remaining targets are pre-processed before motion analysis is performed.

The first pre-processing step is to clean up anomalies in the targets. This is done by a morphological dilation followed by erosion. This removes any small holes in the target and smoothes out any interlacing anomalies. This effectively robustifies small features such as thin arm or leg segments. After the target has been cleaned, its outline is extracted. The process is shown in Fig. 2.

Fujiyoshi *et al*. (2004) proposed the use of star skeletonization procedure for analyzing the motion of human targets. The standard star skeleton techniques for skeletonization such as distance transformation and thinning are computationally expensive and moreover are highly susceptible to noise in the target boundary. The method adapted in this study provides a simple way of detecting only the gross extremities of the target to produce star skeleton. The main idea is, the simple form of skeletonization extracts the broad internal motion features of a target and is employed to analyze the target's motion (Fujiyoshi *et al*., 2004). Then the contour for a human blob is extracted as shown in Fig. 3. The centroid (x_c, y_c) of the human blob is determined by using the following Eq. (9):

$$x_c = \frac{1}{N_b}\sum_{i=1}^{N_b} x_i$$

$$y_c = \frac{1}{N_b}\sum_{i=1}^{N_b} y_i \quad (9)$$

where, (x_c, y_c) represent the average contour pixel position, (x_i, y_i) represent the points on the human blob contour and there are a total of N number of points on the contour. The distance d_i from the centroid to contour points is given by (10):

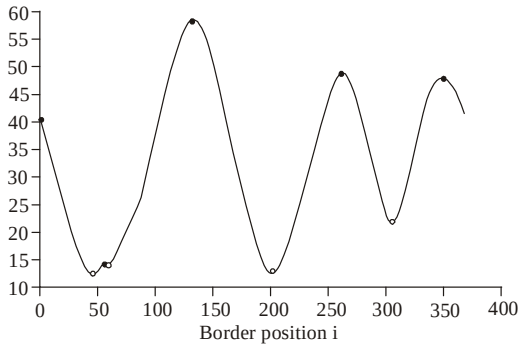


Fig. 3: Plot of skeleton extreme points

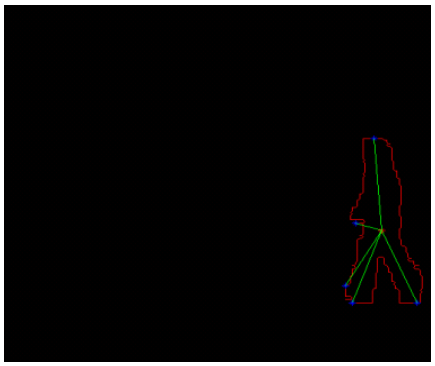


Fig. 4: Skeleton features for normal walk

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \quad (10)$$

From the d_i plot, the local maximum points are collected and their corresponding plot is shown in Fig. 3. The star skeletonization is formed as shown in Fig. 4.

Another prompt to analyze the motion of the target is its posture. Using motion cues based on the star skeleton, it is possible to determine the posture of a moving human. For the cases in which a human is moving in an up-right position, it can be assumed that the lower extreme points are legs and so choosing these points to analyze cyclic motion seems to be a reasonable approach (Fujiyoshi *et al*., 2004). In particular, the left-most lower extreme points (l_x, l_y) are used as the cyclic points. However, it is not necessary that the same leg is detected at all times, because the cyclic structure of the motion will still be evident from this point's motion. If $\{x^s_i, y^s_i\}$ is the set of extreme points, (l_x, l_y) is chosen according to the following conditions (11) and (12):

$$(l_x, l_y) = \left(x_i^s, y_i^s\right) : x_i^s = \min_{y_i^s < y_c} x_i^s \quad (11)$$

Then the angle (l_x, l_y) makes with the vertical angle θ is calculated as:

$$\theta = \tan^{-1} l_x\text{-}x_c \,/\, l_y\text{-}y_c \quad (12)$$

A further cue to determine the posture of moving human is the inclination of the torso. This can be approximated by the angle of upper-most extreme point of the target. This torso angle Φ can be determined in exactly the same manner as θ and the leg angle for walking and running is shown in Fig. 3. The cutoff frequency was set as 0.1 to get appropriate extreme points in this proposed study. At last the leg angle θ, torso angle Φ and the skeleton motion in a sequence are given as input vectors for the Relevance Vector Machine.

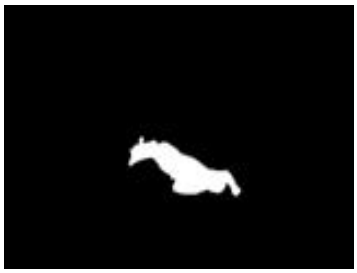
There are three main advantages of this type of skeletonization process. It is not iterative and is, therefore, computationally cheap. It also explicitly provides a mechanism for controlling scale sensitivity. Finally, it does not rely on priori human model.

Relevance vector machine: Fall is an action classification problem that requires RVM for classification. The Relevance Vector Machine is a powerful algorithm, useful in classifying data in to species. The proposed Relevance Vector Machine (RVM) classification technique has been applied in many different areas of pattern recognition. The skeleton points and the motion cues for each blob are selected as features. The RVM is a Bayesian regression framework, in which the weights of each input vector are governed by a set of hyper parameters. These hyperparameters describe the posterior distribution of the weights and are estimated iteratively during training. Most hyper parameters approach infinity, causing the posterior distributions of the corresponding weights to zero. The remaining vectors with nonzero weights are called relevance vectors. In this study, Relevance Vector Machine (RVM) technique is used for the classification of human action such as normal or fall.

RESULTS AND DISCUSSION

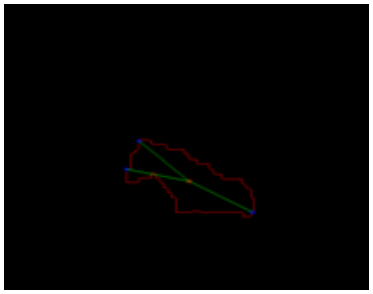
The efficiency of the proposed algorithm has been evaluated by carrying out extensive works on the simulation of the algorithm on benchmark datasets. The method processes about 24 frames per sec for colour images. To demonstrate the performance of the proposed method, different fall action sequences are taken from CAVIAR (http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/) and (Auvinet *et al*., 2010). These datasets contain fall of an individual person from different camera view points. The video sequences were converted into frames and the background subtraction was obtained using GMM shadow removal has been done next to it. Then the star skeletonization was

Input Image



Foreground Extracted Image

Edge Detected Image



Skeleton Image

Fall Detected Image

Fig. 5: Results for benchmark dataset

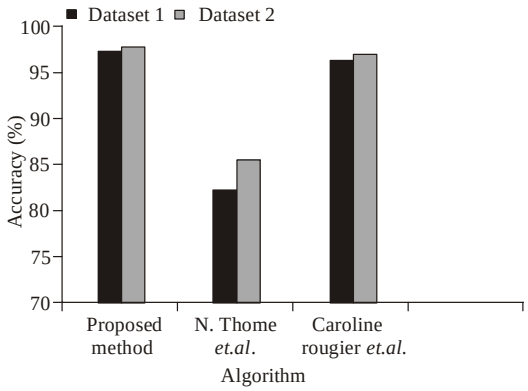


Fig. 6: Comparative results of fall detection

used to obtain the motion cues and skeleton features have been obtained clearly. Finally the skeleton features and the motion cues were given as input for the Relevance vector machine algorithm to classify the fall from the normal one which was indicated in red color as shown in Fig. 5.

To evaluate the proposed approach the classification accuracy has been computed and compared with existing state-of-the-art methods and it is shown in Fig. 6.

To analyze our recognition results, we compute the sensitivity, the specificity and the accuracy obtained with our RVM classifier as follows:

True Positives (TP): Number of falls correctly detected
False Negatives (FN): Number of falls not detected
False Positives (FP): Number of normal activities detected as a fall
True Negatives (TN): Number of normal activities not detected as a fall

- **Sensitivity:** Se = TP/ (TP+FN)
- **Specificity:** Sp = TN/ (TN+FP)
- **Accuracy:** Ac = (TP+TN)/(TP+TN+FP+FN)

Proposed algorithm is validated using bench mark datasets, from Auvinet *et al*. (2010) the Sensitivity: Se = 95.83%, Specificity: Sp = 97.5%, Accuracy: Ac = 96.67% is obtained.

A good fall detection method must have a high sensitivity, which means that a majority of falls are detected and a high specificity, which means that normal activities are not detected as falls. The accuracy must also be high.

CONCLUSION

In this study, a video surveillance algorithm for classifying human fall and normal actions is described. First the foreground blobs are detected using adaptive mixtures of Gaussians. Then shadow removal algorithm is applied to eliminate other objects and shadows. Subsequently, skeleton features are extracted for each individual. These features reduced the training time and also improved the classification accuracy. The features are learnt through a Relevance vector machine to classify the individual's actions into two classes. The Accuracy is also increased by selecting the appropriate kernel Gaussian which also reduces the computational complexity. This facilitates the proposed algorithm that is able to detect abnormal actions of an individual such as normal or fall with high accuracy.

REFERENCES

Anderson, D., J. Keller, M. Skubic, X. Chen and Z. He, 2006. Recognizing falls from silhouettes. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS '06, Columbia, MO, pp: 6388-6391.

Anderson, D., R.H. Luke, J.M. Keller, M. Skubic, M. Rantz and M. Aud, 2009. Linguistic summarization of video for fall detection using voxel person and fuzzy logic. Comput. Vis. Image Understand., 113(1): 80-89.

Auvinet, E., L. Reveret, A. St-Arnaud, J. Rousseau and J. Meunier, 2008. Fall detection using multiple cameras. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2008, 20-25 Aug., Canada H3C 3J7, pp: 2554-2557.

Auvinet, E., C. Rougier, J. Meunier, A. St-Arnaud and J. Rousseau, 2010. Multiple Cameras Fall Dataset. Technical Report 1350, DIRO-Université de Montréal, Retrieved from: http://www.iro.umontreal.ca/~labimage/Dataset/.

Caleanu, C., D. Huang, V. Gui, V. Tiponut and V. Maranescu, 2007. Interest operator versus Gabor filtering for facial imagery classification. Pattern Recogn. Lett., 28: 950-956.

Caroline, R., M. Jean, A. St-Arnaud and R. Jacqueline, 2011. Robust video surveillance for fall detection based on human shape deformation. IEEE T. Circuits Syst. Video Techn., 21(5): 611-622.

Fujiyoshi, H., A.J. Lipton and T. Kanade, 2004. Real-time human motion analysis by image skeletonization. IEICE T. Inform. Syst., E87-D (1): 113-120.

Ibrahim, M.M. and R. Anupama, 2005. Scene adaptive shadow detection algorithm. World Acad. Sci. Eng. Techn., 42: 88-91.

Lee, T. and A. Mihailidis, 2005. An intelligent emergency response system: Preliminary development and testing of automated fall detection. J. Telemed. Telecare, 11(4): 194-198.

Nait-Charif, H. and S. McKenna, 2004. Activity summarisation and fall detection in a supportive home environment. Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, 23-26 Aug., UK, 4: 323-326.

Rougier, C., J. Meunier, A. St-Arnaud and J. Rousseau, 2006. Monocular 3-D head tracking to detect falls of elderly people. Conf. Proc. IEEE Eng. Med. Biol. Soc., 1: 6384-6387.

Rougier, C., J. Meunier, A. St-Arnaud and J. Rousseau, 2007. Fall detection from human shape and motion history using video surveillance. 21st International Conference on Advanced Information Networking and Applications Workshops, AINAW '07, 21-23 May, Montreal, QC, 2: 875-880.

Shan, Y., F. Yang and R. Wang, 2007. Color space selection for moving shadow elimination. Proceeding of 4th IEEE International Conference on Image and Graphics, 22-24 Aug., Air Force Eng. Univ., Xi'an, pp: 496-501.

Sixsmith, A. and N. Johnson, 2004. 2004. A smart sensor to detect the falls of the elderly. IEEE Pervasive Comput., 3(2): 42-47.

Thome, N., S. Miguet and S. Ambellouis, 2008. A real-time, multiview fall detection system: A LHMM-based approach. IEEE T. Circuits Syst. Video Technol., 18(11): 1522-1532.

Toreyin, B.U., Y. Dedeoglu and A.E. Cetin, 2005. HMM based falling person detection using both audio and video. Lect. Notes Comput. Sc., 3766: 211-220, DOI: 10.1007/11573425-21.