Middle-East Journal of Scientific Research 15 (8): 1178-1185, 2013
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2013.15.8.2407

Intelligent Monitoring System of the Elderly's Fall Based on Video Sequences

[1]Kh. Rezaee, [2]E. Azizi, [3]J. Haddania and [4]A. Delbari

[1,2,3]Biomedical Engineering Department Hakim Sabzevari University, Sabzevar, Iran
[1,3]Center for Research of Medical Technologies Sabzevar University of Medical Sciences, Sabzevar, Iran
[4]Department of Neurobiology, Care Sciences and Society (NVS) Karolinska Institute, Stockholm, Sweden

Abstract: Every year thousands of the elderly suffer serious damages such as articular fractures, broken bones and even death due to their fall. In this paper, based on the analysis of images taken from the elderly's movement, an efficient system has been proposed that, in the first phase, simulates the movement of the elderly by detecting their abnormal walking. Thus by combining several important features, including an estimate of body angle, representation of the Motion History Image (MHI) and estimate of the magnitude and direction of movement, the speed of the falling is calculated. This system has been implemented on a set of 57425 video frames received from the elderly residing in Farzanegan Health Care Center in Mashhad city in Iran and the video sequences containing the actual occurrence the of falling. There were only 48 falls and 163 like falls on video sequences with mean age of 66.5±6.1 years. Out of a total of about 58 people including 43 men and 15 women, 39.65% were faced with falls at ambient of health center. All the sequences were randomly converted into four Movie categories with these details: AVI format, 120×160 pixels resolution and 15 fps. Compared to such techniques as Shieh and Rougier, 94.1% average accuracy (AAC), 92% sensitivity and 94.47% specificity indicate the ability of the system in identifying the incidents similar to the fall. The high speed data processing of the algorithm, 92.91% detection rate (DR) and insignificant false alarm rate (FAR), 5%, distinguishes the proposed system from similar ones, particularly due to its intelligent monitoring and its real time detection of the elderly's fall.

Key words: Image Processing · Video Sequence · Elderly's Fall · HSV Space · Motion History Image

INTRODUCTION

Today, many countries are faced with the growing population of the elderly each year. In 2000, the elderly aged above 65 constituted one-eighth of the world's population, i.e. a population equivalent to 750 million people [1]. Based on demographics released in 2010, it is estimated that in 2035, one-third of Europe's population will be above 65 years old [2]. Iran is one of the countries with young population and in near future, it has to deal with an aging population. Falls and loss of balance is a common threat to the health of the elderly. It can affect the quality of life, increase maintenance costs and lead to adverse physical, psychological and social conditions, or even death [3]. Studies show that 25% to 47% of elderly suffer from falls once or more and this figure adds up to 50% among the elderly who live in health care centers. Since this accident can jeopardize the performance and independence of the elderly, the identification of people who are at risk of falling is of paramount importance [5]. Therefore, the first step in the prevention of this incident is to alleviate the side effects of falling [6]. If the elderly are not able to inform people in case of their fall, it might aggravate the damages of this incident and in some cases lead to the loss of their life. Therefore, a smart and efficient system to detect falls of elderly people seems

Corresponding Author: Khosro Rezaee, Department of Computer and Biomedical Engineering, Hakim Sabzevari University, Sabzevar, Iran.

essential. The techniques that have been proposed to date to signify people's falls can be classified into three general categories:

- Sensor networks and wearing sensors.
- The use of gyroscopes, accelerometer and devices for detecting the vibrations caused by falls.
- Monitoring the dynamic state based on video analysis.

In the system designed by Alexander *et al*. [7] sensor network techniques have been used for monitoring and online surveillance of the elderly. The main shortcoming of wireless sensor network system is that the elderly need to wear bulky clothes and in case they forget to do so, they will not be able to declare their situation at the time of the fall.

Furthermore, this technique will be inoperable in case the person loses his/her consciousness after the fall. Vibration analysis instruments, gyroscope, status belts and pressure gauge board are other methods designed according to individuals' manner of movement. In 2008, Bourke *et al*. [9] created a secure threshold for fall detection algorithms which used a biaxial gyroscope. The simultaneous combination of the accelerometer system of the elderly's movement and the estimate of the movement's direction was also proposed by Nyan [10]. The vibration analysis system is also used in detecting the fall of the elderly or the disabled. Zigel also enhanced the instrument used for detecting the floor vibration in 2009 by adding a sound sensor [11]. Among all current systems, the real-time systems which detect the fall of people based on analysis of video images have the highest efficiency and accuracy. As to the designing of video surveillance algorithm, Naseimento [12] proposed methods based on the analysis of machine vision in detecting changes in individuals' position. In 2012, Liao and Huang [13] detected slips and falls of people based on Bayesian networks. Each of the current systems has strengths and weaknesses with regard to their detection. Low accuracy, low processing speed, lack of real-time response to the events in some of the above cases and high levels of positive error are among the weaknesses of such systems.

MATERIALS AND METHODS

The system was implemented on 57425 video frames taken from Farzanegan Health Care Center in Mashhad and video sequences containing the occurrence of the falls. There were only 48 falls and 163 like falls on video sequences. To determine the accurate falls, recurrent falls in a cohort of elderly people with mean age of 66.5±6.1 years over 160 hours of recorded video. Out of a total of about 58 people including 43 men and 15 women, 39.65% were faced with falls at ambient of health center. In other hand, 17 men and 6 women had frequent falls with mean age of 69.5±5.4 years. Associated with falls among the elderly, 31 cases (91.66%) were related to the accidents or sudden imbalance, 3 cases (5.55%) associated with the devices and accessories and 2 cases (2.77%) were related to the other events. The method proposed for detection of the elderly's falls in this paper draw mainly on video processing techniques to detect the target area. The algorithm has been shown in in Figure 1 and the main body of the system will be introduced in following sections accordingly.

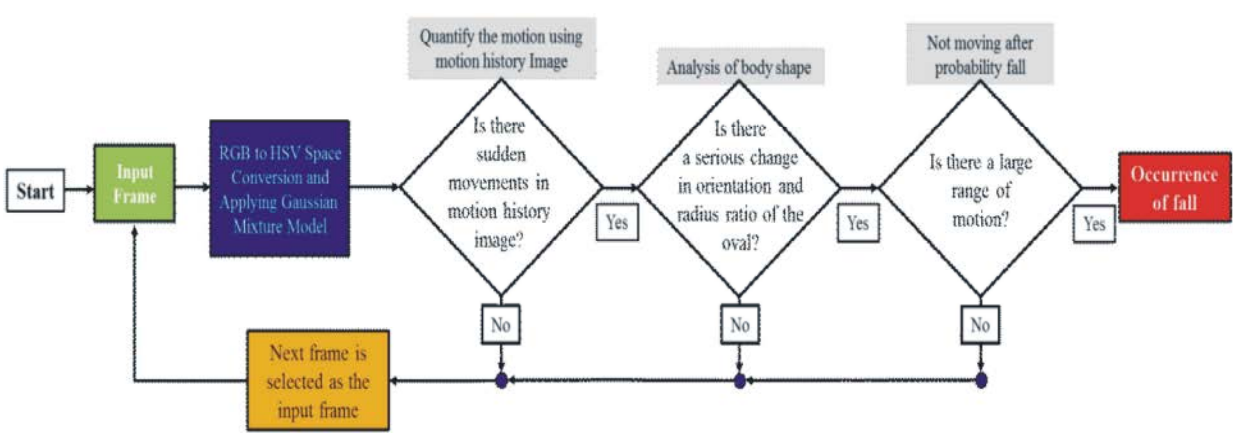

Fig. 1: Conversion from RGB into HSV space for input frames

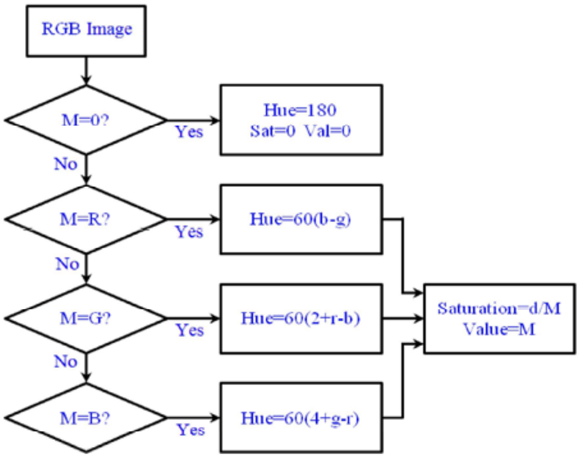

Fig. 2: Conversion from RGB into HSV space for input frames

Conversion from RGB to HSV Space: In many machine vision and image processing algorithms, the intensity of disproportionately high or low light, such as shades in separating a special part of the image, produces error. The color space, which is the result of the removal of unwanted effects of light, is the conversion to HSV space. Using the features of color space in HSV environment can reduce the complexities between the image surface and the intensity of unwanted light which causes errors. In calculation of H section, it is assumed that *M=max (R, G, B), m=min (R, G, B)* and *d=M-m*. The values of r, g and b are also calculated according to *r= (M-R)/d*, *g= (M-G)/d* and *b= (M-B)/d*. The main function of frame conversion from RGB into HSV space is minimizing the effects of the individual's shadows in images, which is the major cause of errors in mode separation. RGB image is converted into HSV space diagram as shown in Figure 2.

Gaussian Mixture Model: The subtraction of the backgrounds of a video sequence from each other is a method by which the target foreground section in each frame is separated. Among the conventional methods of thresholding, Gaussian mixture model is one of the techniques used for separation and display of different background images [14]. According to equation (2.1), Gaussian model for each pixel is the Gaussian density, i.e.:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \tag{2.1}$$

In this equation, μ is the mean intensity of pixels brightness and σ is the variance. By constructing probability density function (PDF) with μ and σ variables, each pixel in each frame is distributed by K mixture in Gaussian model and the probability of the pixel having the value of X_n at the time of N is calculated according to equation (2.2):

$$P(x_N) = \sum_{j=1}^{K} w_j \eta\left(x_N, \theta_j\right) \tag{2.2}$$

In which w_k is the weight parameter of K Gaussian factor and □(x,è$_k$) is the normal distribution of K factor, which is calculated according to equation (2.3).

$$\eta(x,\theta_k) = \eta(x,\mu_k,\Sigma_k)\frac{1}{(2\pi)^{D/2}|\Sigma_k|^{1/2}}\exp(Z) \tag{2.3}$$

μ_k and Σ_k are respectively mean and covariance of K factor and $Z = \left[-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right]$. The number of distribution K is estimated according to the w_k *divided* σ_k merit function and the first distribution of (B) is used as a foreground model. B function is calculated according to equation (2.4) sufficiency

$$B = \arg\min_b \left(\sum_{j=1}^{b} w_j > T\right) \tag{2.4}$$

where T threshold is the lowest decimal value in the foreground model. During the conversion, the foreground and background are separated. Figure 3 shows a set of frames taken from a video sequence as well as the application of conversion to HSV space and Gaussian mixture model.

Fig. 3: Images from top left to the right show 89 to 99 frames of a sample video sequence from the fall of an individual and in the bottom column, the images produced after using Gaussian mixture model and thresholding has been presented in binary format.

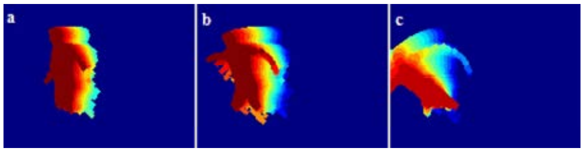

Fig. 4: Motion History Images for (a) 67 to 78 frames, (b) 79 to 90 frames and (c) 91 to 102 frames of a sample video sequence or 132 video frames.

The Anatomical Changes of Human Body During a Fall: After separation of the foreground and background of the image, the location and position of the person is identified. The main advantage of this process is the identification of the person's posture relative to the horizontal and vertical axes.

After the removal of the additional image pixels and by defining an estimated oval that specifies the position of the person, information about the shape, form and direction of the person's movement can be obtained. In two-dimensional coordinate system, the estimated oval is made of (x, y) center, φ direction and d_1 and d_2 diameters. When a change in the manner of movement is made, the analysis of two indexes is of paramount importance.

- Standard deviation of the movement direction ($\square_{\varphi}$) in the estimated oval.
- Calculating the ratio of d_1 and d_2 as well as standard deviation $\square_{d1/d2}$.

Motion History Image (MHI): The process of moving in successive frames of a video sequence can be considered as a type of a memory. The person's motion history in the combined image or images, which defines the precedent of the person's movement, is one of the techniques used for displaying the process of movement in the elderly, which according to the binary sequence of the motion area of the person, is modeled based on $D(Pix_x, Pix_y, t)$ taken from the sequence of the main image $I(Pix_x, Pix_y, t)$. Each pixel in the images representing the precedent of individual motion by P_{MMI} is shown according to equation (2.5) which is, in fact, the temporary memory of movement in each point that occurs in time interval T ($1 \leq T \leq N$) (N is the number of sample frames in the input video sequence).

$$P_{MMI} = \begin{cases} T & D(Pix_x, Pix_y, t) \\ \max(0, P_{MMI}(ST) - 1) & otherwise \end{cases} \tag{2.5}$$

where *ST* is $(Pix_x, Pix_y, t\text{-}1)$ and in the frames showing the previous movements, the pixels with higher brightness change show greater motion. To quantify the person's movement, the C_{motion} coefficient is estimated according to motion memory in frames which is recorded every 500ms. This coefficient is calculated according to equation (2.6)

$$C_{motion} = \frac{Pix_{\mathrm{Re}d\,Level}}{Pix_{\mathrm{Re}dLevel} + Pix_{OtherLevel}} \tag{2.6}$$

This equation is expressed based on the total number of red pixels and non-red pixels in the area show the previous frame movements which changes in time interval [1-0]. In Figure 4, the algorithm of motion memory has been applied to three sets of 12 frames from three sample video sequences, which represents the constant, immediate or variable motions and calculation of C_{motion} coefficient.

The calculation of motion coefficient is necessary for detecting the falls in the elderly. It is not enough, however. By calculating the gravity center of the person in the threshold frames, an estimated oval is illustrated. The analysis of the direction and angle of this oval according to horizontal and vertical direction plays an important role in identifying a person's fall.

Determining the Direction of a Person's Motion (φ angle) and Eccentricity (RHO): The visual arrays from input film, in which each matrix cell has m×n dimensions, form one frame of the input video. In addition, in the matrix representing the gravity center of the oval, N row

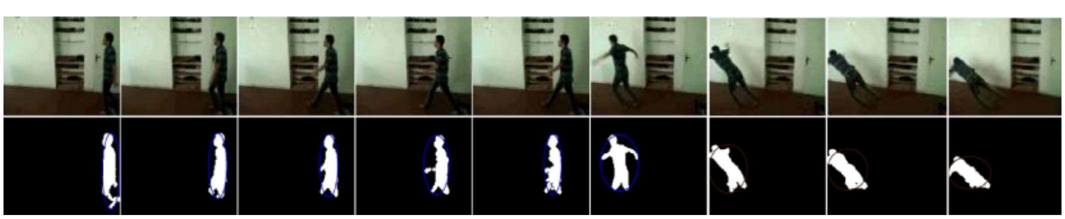

Fig. 5: Images from top left to the right, 40, 45, 53, 61, 66, 89, 95, 97 and 99 frame and in bottom columns from left to the right, the change of oval based on variation in one's posture has been depicted and C_{motion} has crossed the defined threshold in frame 95.

is in fact x and y coordinates of the center of the oval in which N is the number of frames in input video sequence. The direction of both large and small diameters of the oval shows the direction of the oval in two-dimensional coordinates and the ratio of these two diameters shows the eccentricity of the elderly person. In Figure 5, a set of 9 frames has been shown in which an estimated oval based on the formal mold of a person's body has been depicted.

RESULTS AND DISCUSSION

In 3 steps, the detection process of a fall in video images was carried out. These three steps were:

- Quantification of movements and modeling accordingly.

When C_{motion} changes from the defined threshold, which has been obtained according to the statistical calculations of video images, the movements similar to the occurrence of a fall can be detected.

- Analysis of the body shape and form in the binary frame

One of the main parameters which will undergo tremendous changes is $\square_{\varphi}$ or standard deviation in the direction of motion. In average, of 96 tested video sequences, $\square_{\varphi}$ was about 15 degree and the ratio of $\square_{d1/d2}$ was 0/9.

- The inactivity of the elderly after the fall.

The parameters that might change after a fall or during inactivity of the person are expressed as follow:

- C_{motion} >5%
- Standard deviations $\square_x$ and $\square_y$ which are both smaller than 2.
- Standard deviations in the estimated oval equations, which under conditions $\square_{d1} < 2$, $\square_{d2} < 2$ and $\square_{\varphi} < 15^o$ shows the inactivity or stillness of the elderly after suffering from a fall.

The system was implemented on 57425 video frames taken from Farzanegan Health Care Center in Mashhad and video sequences containing the occurrence of the falls. There were only 50 falls in video sequences. All sequences were randomly converted into 4 categories of Movie with these details: AVI format, 120×160 pixels resolution and 15 fps.

In Table 1, Average accuracy (AAC) and detection rate (DR) and false alarm rate (FAR) have been calculated according to equations (3.1)-(3.3) which are used for measuring the accuracy of detection in video sequences.

$$AAC = \left(\frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}\right) \times 100 \tag{3.1}$$

$$DR = \frac{Number\ of\ True\ Positive}{Number\ of\ Fall\ Events} \tag{3.2}$$

$$FAR = \frac{N_{FP}}{N_{FP} + N_{TN}} \tag{3.3}$$

In these equations,

N_{TP}: The number of frames containing the occurrence of a fall that has been detected by algorithm.

N_{FN}: The number of frames containing the occurrence of a fall which algorithm has not been successful in their detection.

N_{TN}: The number of frames that does not contain the occurrence of any fall and algorithm has not identified them.

N_{FP}: The number of frames that does not contain the occurrence of any fall but the algorithm has misidentified them.

Table 1: The received video sequence, calculation of mean accuracy, DR and FAR of the proposed algorithm

		Number of frames in which the elderly person falls		Number of frames with normal movement or similar falls.				
Video Clips	No. Frames	N_{TP}	N_{FN}	N_{TN}	N_{FP}	Average Accuracy (AAC)	Detection Rate (DR)	False Alarm Rate (FAR)
Movie 1	12758	9	1	31	2	93.02%	90.00%	6.06%
Movie 2	16892	11	2	39	2	92.59%	84.61%	4.87%
Movie 3	19656	15	2	68	4	93.25%	88.23%	5.55%
Movie 4	8119	8	0	16	1	96.00%	100%	5.88%
Total	57425	43	5	154	9	93.36%	89.58%	5.52%

Table 2: Comparison of the proposed system with methods commonly used for detection of the falls

Methods	Author	Sensitivity (%)	Specificity (%)
Sensor networks and Gyroscopes	Bourke [9]	100	97.5
	Nayan [10]	100	92.5-97.5
Accelerometers and Vibration Systems	Alwan [16]	93.28-100	100
	Zhang [17]	89.1	Unknown
Algorithms based on Machine Vision	Rougier [18]	88	87.5
	Vish [19]	100	100
	Tao [20]	82-96	Unknown
	Shieh [21]	82.2-92.2	90.9-97.3
	Our Work	~90	~95

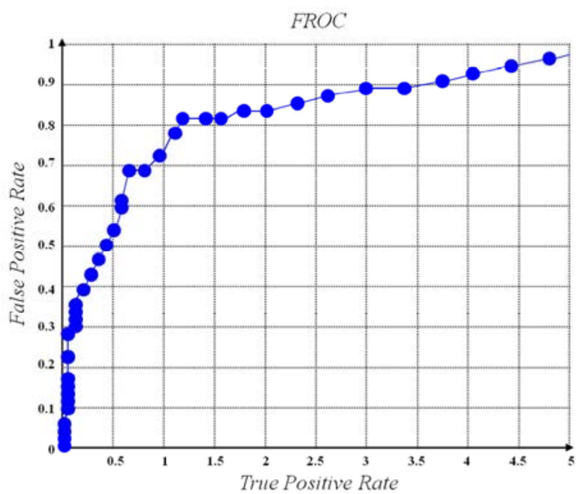

Fig. 6: FROC diagram for a set of near-falls and falls of the elderly.

ROI is also the target area in the image which includes various types of falls in the elderly. In (3.4), (3.5) specificity and sensitivity were respectively equal to 94.47% and 89.58%.

$$SP = \left(\frac{N_{TN}}{N_{TN} + N_{FP}}\right) \times 100 \tag{3.4}$$

$$SE = \left(\frac{N_{TP}}{N_{TP} + N_{FN}}\right) \times 100 \tag{3.5}$$

It should be noted that the above method is a new technique and a comparison has been drawn between the performance of this system and other methods used for detection of the elderly's fall. Different methods use various databases. The proposed method in this paper, however, has been applied to a greater database. At first, it should be noted that the use of a single camera would decrease the sensitivity and specificity of the third technique in contrast to other two techniques. Thus, the occurrence of error is only natural.

Using multiple cameras, which simultaneously identify the fall of the elderly, increases the accuracy, sensitivity and specificity of this system more than the other two techniques. Inconsistency is another problem that may occur when decreasing the frames and the resulted time delay might result in lower sensitivity. Figure 6 shows FROC diagram for a set of 40 near-fall cases, 25 of which were relatively severe. According to the diagram, the high specificity of the system is due to proper detection of the cases without the occurrence of any fall, which algorithm has been able to properly discriminate from the actual occurrence of the falls. The high sensitivity of the system has also been due to correct detection of the falls and low errors. Users are more likely to sensor networks because sensor networks provide greater sensitivity and average, are higher, the sensitivity and specificity of 98% and 95% respectively.

Table 2 shows the greater ability of the system in detecting the occurrence of a fall. Accelerometers and vibration analysis systems are facing with the sensitivity between 90%-95% and specificity higher than 95%. While the sensitivity and specificity of the algorithm based on machine vision is respectively 85-95% and 90%-100%. Therefore, systems based on machine vision techniques in terms of accuracy and sensitivity is competitive in performance compared to other techniques such as sensor network and accelerometers and vibration analysis systems. Unlike other techniques, the method of video processing has the advantage which is doing intelligent monitoring of the elderly person and is comfortable wearing heavy clothing. Also accelerometers and vibration analysis systems haven't the ability to implement in anywhere. Figure 7 shows changes of RHO and C_{motion} depending on the number of video frames. Figure 8 displays the standard deviation and the standard deviation of theta eccentricity (RHO) respectively. Fall events are detected by analysis of these figures in amplitude of the standard deviation and the standard deviation of theta eccentricity.

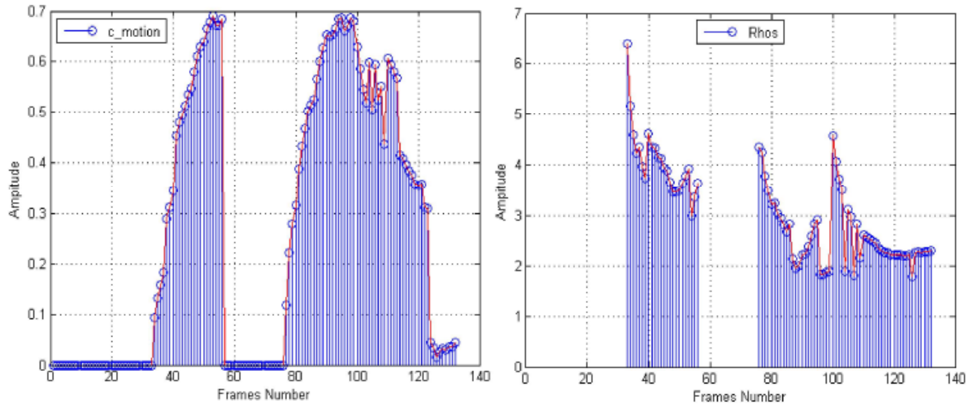

Fig. 7: Drawing RHO changes and C_{motion} in order to the number of frames of video sequences.

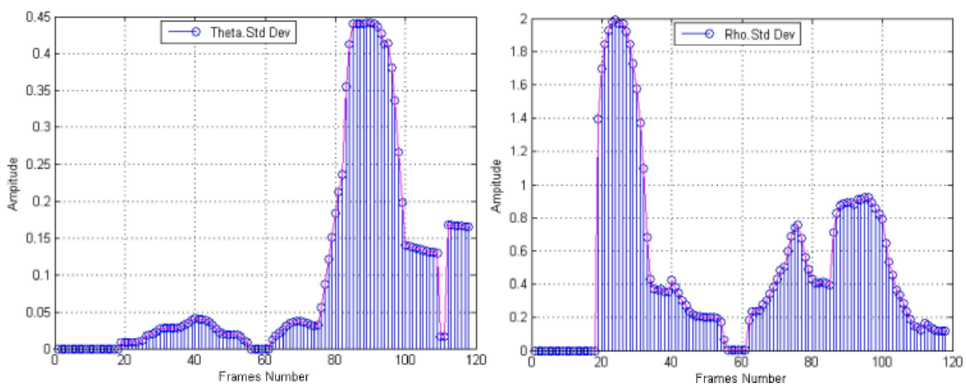

Fig. 8: Theta standard deviation and standard deviation of eccentricity (RHO) from left to right respectively is shown.

CONCLUSION

Based on the analysis of video sequences and using machine vision and image processing techniques, the proposed system detects the occurrence of the falls with higher accuracy and speed. The practical results and final simulation of the algorithm showed that the system, though using a single camera to capture the movements of the elderly, produced higher sensitivity and specificity compared to similar methods. 94% accuracy, low false alarm rate and high detection rate have dramatically increased the reliability of the system performance. The prediction of the fall of the elderly based on the analysis of video images will be among the future researches of the author. It would be based upon motion equations and biomechanics joints of the elderly. Implementation and installment of this system in elderly nursing homes and the hospital wards immunize the elderly vulnerable to the fall against the terrible consequences of this unfortunate accident.

REFERENCES

1. Newton, RA., 1995. Standing balance abilities of elderly subjects under altered visual and support surface conditions. Physical therapy Canada. Winter 47(1): 25-29.
2. Ambient Assisted Living Joint Programme (2008), available from http://www.aal-europe.eu.
3. Bal oh Rw, T.D. Fife, Z. Ling, T. Socotch, K. Jacobson, T.B. Bell and K. Kirch, 2000. Comparsion of Static and Dynamic Posture Graphy in Young and Older Normal People. JAGS, 42(4): 405-412.
4. Wegner, L., C. Kisner and D. Nichols, 1997. Static and Dynamic Balance Response in Persons with Bilateral Knee Osteoartherritis. JOSPT, 25(1): 13-18.
5. Atwater, S.L., T.K. Crawe, J.C. Deitz and P.K. Richardson, 1990. Inter Rater and Test-Restart Reliability of Two Predictive Balance Tests. Physical Therapy, 70(2): 79-87.

6. Satterfied, KS., 2001. Balance Testing Helps Identify Elderly at Risk of Multiple Falls. American Otological Society.
7. Alexander, G.L., M. Rantz, M. Skubic, M.A. Aud, B. Wakefield, E. Florea and A. Paul, 2008. Sensor systems for monitoring functional status in assisted living facility residents. Research in Gerontological Nursing, 1(4): 238-244.
8. DirectAlert, 2010. Wireless emergency response system. URL: http:// www.directalert.ca/ emergency/ help-button.php.
9. Bourke, A.K. and G.M. Lyons, 2008. A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. Medical Engineering and Physics, 30(1): 84-90.
10. Nyan, M., F.E. Tay and E. Murugasu, 2008. A wearable system for pre-impact fall detection. Journal of Biomechanics, 41(16): 3475-3481.
11. Zigel, Y., D. Litvak and I. Gannot, 2009. A method for automatic fall detection of elderly people using floor vibrations and sound - proof of concept on human mimicking doll falls. IEEE Transactions on Biomedical Engineering, 56(12): 2858-2867.
12. Nascimento, J.C. and J.S. Marques, 2008. Performance evaluation of object detection algorithms for video surveillance. IEEE Transactions on Multimedia, 8(4): 761-74.
13. Yi, T., Chung Lin Huang and C.H. Shih, 2012. Slip and fall event detection using Bayesian Belief Network. Pattern Recognition, 45: 24-32.
14. Stauffer, C. and W.E.L. Grimson, 1999. Adaptive background mixture models for real-time tracking. Proc. CVPR, Fort Collins, Colorado, USA, pp: 246-252.
15. Pong, P.K. and R. Bowden, 2001. An Improved Adaptive Background Mixture Model for Real time tracking with Shadow Detection. In Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, AVBS01.
16. Alwan, M., G. Demiris and Z. He, 2006. Technology for successful ageing. Proceedings of the 28th IEEE Embs Annual International Conference.
17. Zhang, T., J. Wang, P. Liu and J. Hou, 2006. Fall Detection by Embedding an Accelerometer in Cellphone and Using KFD Algorithm. IJCSNS International Journal of Computer Science and Network Security, 6(10): 277-284.
18. Rougier, J., J. Meunier, A. St-Arnaud and J. Rousseau, 2007. Fall detection from human shape and Motion history using video surveillance. In: International on Advanced Information Networking and Applications Workshops, 21-23 May 2007; 2: 875-880.
19. Vishwakarma, V., C.A. Mandal and S. Sural, 2007. Automatic detection of human fall in video. In: International Conference on Pattern Recognition and Machine Intelligence, pp: 616-23.
20. Tao, J., M. Turjo, M.F. Wong, M. Wang and Y.P. Tan, XXXX. Fall incidents detection for intelligent video surveillance. In: Proceedings of the Fifth International Conference on Information, Communications and Signal Conference Processing, pp: 1590-4.
21. Shieh, W.Y., and J.C. Huang, 2011. Falling-incident detection and throughput Enhancement in a multi-camera video-surveillance system. Medical Engineering & Physics, 10(16): 28-38.