Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008)
25-27 December, 2008, Khulna, Bangladesh

Intelligent Video Surveillance for Monitoring Fall Detection of Elderly in Home Environments

Homa Foroughi[1], Baharak Shakeri Aski[2], and Hamidreza Pourreza[1]

[1]Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Khorasan, Iran

[2]Islamic Azad University of Ramsar, Ramsar, Mazandaran, Iran

***Abstract* — Video Surveillance is an omnipresent topic when it comes to enhancing security and safety in the intelligent home environments. In this paper we propose a novel method to detect various posture-based events in a typical elderly monitoring application in a home surveillance scenario. These events include normal daily life activities, abnormal behaviors and unusual events. Due to the fact that falling and its physical-psychological consequences in the elderly are a major health hazard, we monitor human activities with a particular interest to the problem of fall detection. Combination of best-fit approximated ellipse around the human body, projection histograms of the segmented silhouette and temporal changes of head position, would provide a useful cue for detection of different behaviors. Extracted feature vectors are fed to a MLP Neural Network for precise classification of motions and determination of fall event. Reliable recognition rate of experimental results underlines satisfactory performance of our system.**

***Index Terms* — Elderly Monitoring, Fall Detection, Home Surveillance, Human Shape, MLP Neural Network, Posture Recognition.**

I. INTRODUCTION

Intelligent video surveillance systems are receiving a great deal of interest especially in the fields of personal security and assistance. These systems are built in order to accomplish several tasks from detection of human presence to recognition of their activities. In the past few decades, vision-based surveillance has been extensively applied on industrial inspection, traffic control, security systems, and medical and scientific research.

One of the application areas of video surveillance is monitoring the safety of elderly in home environments. In the case of elderly people living on their own, there is a particular need for monitoring their behavior, such as a fall, or a long period of inactivity. Fall in the elderly is a major public health problem and may lead to injury, restricted activities, fear or death. A fall incident not also causes many disabling fractures but also has dramatic psychological consequences that reduce the independence of the person. It is shown in [12] that 28-34% elderly people in the community experience at least one fall every year, and 40-60% of the falls lead to injury.

Human society is experiencing tremendous demographic changes in aging since the turn of the 20th century. Thus with the population growing older and increasing number of people living alone, supportive home environments able to automatically monitor human activities are likely to widespread due to their promising ability of helping elderly people.

Nowadays, the usual solution to detect falls is wearable sensors [1], [3]. These autonomous sensors are usually attached under the armpit, around the wrist, behind the ear's lobe or at the waist. These devices integrate accelerometer and/or inclinometer sensors and can monitor velocity and acceleration, vertical posture toward lying posture. However the problem of such detectors is that older people often forget to wear them, indeed their efficiency relies on the person's ability and willingness to wear them. Moreover in the case of noncontact sensors, they often provide fairly crude data that's difficult to interpret. Although, computer vision systems try to extract some considerable features from video sequences of movement patterns to detect falls. The data provided by cameras are semantically richer and more accurate than standard sensors.

In this paper we present a novel video analysis based approach for monitoring human activities with a particular interest to the problem of fall detection. The reminder of the paper is organized as follows: in section II we briefly review some existing vision-based fall detection systems, in section III our proposed system is introduced and then more technical details are described in section IV. Experimental results are represented in section V and finally we conclude in section VI.

II. RELATED WORK

Recently some research has been done to detect falls using image processing techniques. A simple method was used in [5], [6] based on analyzing aspect ratio of the moving object's bounding box. This method could be inaccurate, depending on the relative position of the person, camera, and perhaps occluding objects. The works in [2], [7] used the normalized vertical and horizontal projection of segmented object as feature vectors. To overcome occluding objects problem, some researchers have mounted the camera on the ceiling: Lee [8] detected a fall by analyzing the shape and 2D velocity of the

1-4244-2136-7/08/$20.00 ©2008 IEEE

person. Nait-Charif [9] tracked the person using an ellipse and inferring falling incident when target person is detected as inactive outside normal zones of inactivity like chairs or sofas. Rougier [4] used wall-mounted cameras to cover large areas and falls were detected using motion history image and human shape variation. Other systems used the audio information or using 3D trajectory and speed of head to infer events [11]. These mechanisms tend to be more complex and need more additional cost. Despite the considerable achievements that has accomplished on this field in the recent years, there are still some clear challenges to overcome:

- Visual fall detection is inherently prone to high levels of false positive as what appears to be a fall might not be a fall, but a deliberate movement towards the ground. In other words, most of current systems [6], [7], [9], [10] are unable to discriminate between real fall incident and an event when person is lying or sitting down abruptly.
- Existent fall detection systems tend to deal with restricted movement patterns and fall incidents are usually detected in contrast with limited normal scenarios like walking; however in real home environments various normal /abnormal motions occur.

III. PROPOSED SYSTEM OVERVIEW

This paper proposes a novel method for monitoring human posture-based events in a home environment with focus on detecting three types of fall. We tried to simulate real life situations; to this aim, we considered comprehensive movement scenarios consisting normal daily life activities such as walking, running, bending down, sitting down and lying down, some abnormal behaviors like limping or stumbling, and also unusual events like falling. Furthermore several scenarios of falling have been regarded. These ten different postures are described in section V.A in more details.

Since we are interested in analyzing the motion occurring in a given window of time, firstly we need to obtain the segmentation of moving objects. Therefore, a background estimation procedure is performed to separate motion from background. After the silhouettes are acquired, the next step involves extracting features. Since changes in the human shape can discriminate if the detected motion is normal (e.g. the person walks or sits) or abnormal (e.g. the person falls), we analyze the shape changes of the extracted silhouettes in the video sequence. It seems that combination of approximated ellipse around the human body, horizontal and vertical projection histograms and temporal changes of human head position; would provide a useful cue for detecting different behaviors. By using the motion feature, the key motion information can be encoded without any accurate visual tracking. The extracted features will then be fed to the MLP neural network with back propagation learning scheme for precise classification of motions.

IV. TECHNICAL DETAILS

A. Foreground Segmentation

Background subtraction is a particularly popular method for motion segmentation and attempts to detect moving regions in an image by differencing between current image and a reference background image in a pixel-by-pixel fashion. In order to human segmentation and extracting the moving objects; here we use a background subtraction method described in [15], which is fairly robust and gives appropriate results on image sequences with shadows, highlights and high image compression.

B. Feature Extraction

One of the issues of major importance in a recognition system is feature extraction, i.e., the transition from the initial data space to a feature space that will make the recognition problem more tractable. So we analyze the shape changes of the extracted silhouettes in the video sequence. To this aim, three main features that retain the motion information of the actions and can be easily obtained are selected to form a feature vector.

- *Approximated Ellipse*

When a motion is detected, an analysis on the moving object is performed to detect a change in the human shape, more precisely in orientation and proportion. The person is then approximated by an ellipse using moments [4]. An ellipse is defined in a 5D parameter space by its center $(\bar{x},\bar{y})$, its orientation and the length *a* and *b* of its major and minor semi-axes [13]. The authors believe that this representation of a person is rich enough to support recognition of different events. For a continuous image *f(x, y)*, the moments are given by:

$$m_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^p y^q f(x,y)\,dxdy \quad \text{for } p,q = 0,1,2,\ldots \tag{1}$$

The center of the ellipse $(\bar{x},\bar{y})$ is used to compute the central moment as follows:

$$\mu_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y)\,d(x-\bar{x})\,d(y-\bar{y}) \tag{2}$$

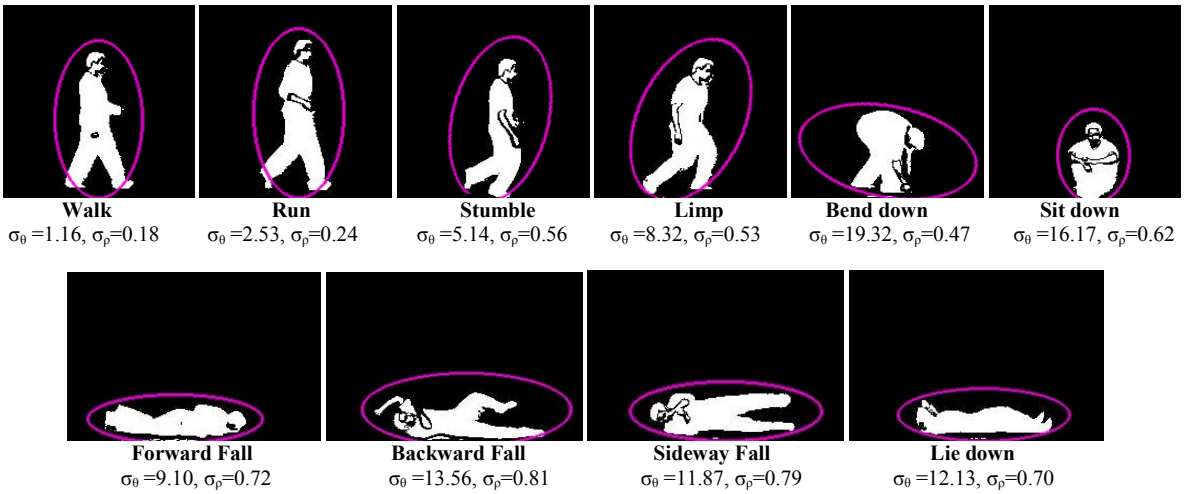


Fig. 1. Approximated Ellipses of Different Human Postures

The angle between the major axis of the person and the horizontal axis *x* gives the orientation of the ellipse and can be computed as follows:

$$\theta = \frac{1}{2}\arctan\left(\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}\right) \tag{3}$$

I_{min}, I_{max}, - the least and greatest moments of inertia-are computed by evaluating the eigenvalues of the matrix:

$$J = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix} \tag{4}$$

Then, the major semi-axis a and the minor semi-axis b of the best fitting ellipse are given by:

$$a = (4/\pi)^{1/4}\left[\frac{I_{max}^{3}}{I_{min}}\right]^{1/8} \quad b = (4/\pi)^{1/4}\left[\frac{I_{min}^{3}}{I_{max}}\right]^{1/8} \tag{5}$$

With *a* and *b*, we can also defined the ratio of the ellipse = *a/b* (related to its eccentricity). In order to detecting changes in the human shape discriminate an abnormal behavior from other normal activities, two values are computed:

- **The orientation standard deviation of the ellipse**: If a person falls perpendicularly to the camera optical axis, then the orientation will change significantly and will be high. If the person just walks, will be low.
- **The ratio standard deviation of the ellipse**: If a person falls parallelly to the camera optical axis, then the ratio will change significantly and will be high. If the person just walks, will be low.

Note that and are computed for a 2 second duration. As Fig. 1 demonstrates, and are completely distinctive for different postures.

• *Projection Histograms*

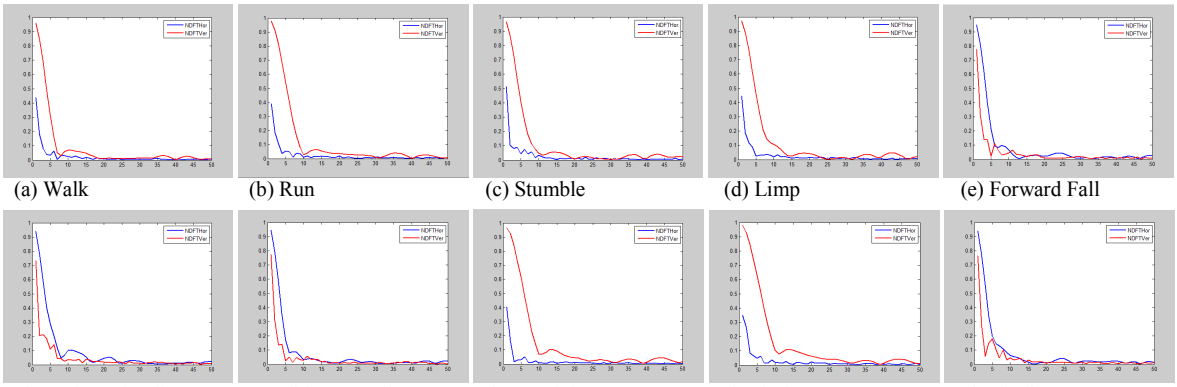
The shape of a 2D binary silhouette can be represented by its projection histogram [13]. The horizontal and vertical projection histogram of foreground is obtained by calculating the number of foreground pixels row wise and column wise. Since projection histograms vary according to location of object in the scene, it's necessary to perform normalization step first. The widely used approach for normalization is rescaling the detected silhouette to a fixed length *M* [2]. However, the normalization parameter is usually obtained by trial and error and also it is posture dependant and strongly sensitive to different postures. To overcome these drawbacks, we propose exploiting *Discrete Fourier Transform* method for normalization.. Let us denote the image size as $N \times M$, so we can consider the foreground *F* as cloud of (x_p, y_p) points. Then the horizontal and vertical projection histograms of foreground F can be defined as follows:

$$Hz(y) = \left|(x_p, y_p) \in F,\ y_p = y\right| \tag{6}$$
$$Vt(x) = \left|(x_p, y_p) \in F,\ x_p = x\right|$$

Applying DFT on $Hz(y)$ and $Vt(x)$ results in:

$$DFT_{Hz(h)} = \frac{1}{N}\sum_{i=1}^{N} Hz(i)\, e^{-2j\pi\frac{hi}{N}} \qquad h = 0, \ldots, N-1 \tag{7}$$

$$DFT_{Vt(v)} = \frac{1}{M}\sum_{i=1}^{M} Vt(i)\, e^{-2j\pi\frac{vi}{M}} \qquad v = 0, \ldots, M-1$$



(a) Walk (b) Run (c) Stumble (d) Limp (e) Forward Fall

(f) Backward Fall (g) Sideway Fall (h) Bend down (i) Sit down (j) Lie down

Fig. 2. Projection Histograms

Magnitudes of these DFT coefficients tend to decay for large values of h and v. Besides, different postures magnitudes differ mostly in the first few term.

Thus we extract first fifty significant DFT coefficients i.e. $DFT_{Hz(1)} \ldots DFT_{Hz(50)}$ and $DFT_{Vt(1)} \ldots DFT_{Vt(50)}$ are selected and normalized by:

$$NDFT_{Hz(h)} = \frac{\left|DFT_{Hz(h)}\right|}{\left|DFT_{Hz(1)}\right|} \quad h = 2, \ldots, 50 \tag{8}$$

$$NDFT_{Vt(v)} = \frac{\left|DFT_{Vt(v)}\right|}{\left|DFT_{Vt(1)}\right|} \quad v = 2, \ldots, 50$$

Normalized magnitudes of fifty significant DFT coefficients of different postures are shown in Fig. 2.

• *Temporal Changes of Head Position*

Head position is a potentially powerful and intuitive pointing cue if it can be obtained accurately and noninvasively. We choose to track the head because it is usually visible in the scene (no occlusion problems), and has a large movement during a fall. Our proposed method tries to estimate a person's head position in every frame of a video sequence and further infer his attentive behavior. To localize the head of the person, firstly silhouette is enclosed by minimum bounding rectangle and then the topmost detected point of the bounding box is marked in each frame [14]. Then the absolute difference values in successive frames are obtained. The difference in topmost point of the head over successive frames forms this part of our feature vector, and applying appropriate threshold on the vertical displacement of topmost point, results in proper differentiation of various behaviors. The curves obtained from sequence of successive head positions are shown in Fig. 3. Notice that the movement of head is computed for a 3 second duration.

C. Posture Classification

Neural Network is an interesting approach for analyzing time-varying data. Here, for each activity which has to be recognized a solitary neural network has been used. So for recognition of n activities, n neural networks exist. In this study we use a four-layered MLP network with backpropagation learning schema. MLP is a supervised neural network that can have multiple inputs and outputs and multiple hidden layers with arbitrary number of neurons. The activation function is $a_j = f(\sum_i w_{ji} a_i)$, where the w_{ji} is the weight of jth node in current layer to ith node in previous layer and a_i is the output of ith node in previous layer. $f(x)$ is sigmoid function. All available data are random divided into two subsets, for training and testing respectively.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Setup

In order to evaluate the overall system performance, we apply the proposed approach to a set of videos recorded in our lab. Here we consider indoor environment settings with single fixed camera monitoring static scene. Distance of person to the camera is approximately 4-5 meters. We could expect three different kinds of behavior: Normal, Unusual and Abnormal.

a) Normal Daily Activity: Five different normal daily activities are considered: Walking, Running, Bending down for catching something and rising up, Sitting down on the floor and standing up and Lying down.

b) Unusual -Fall-: As most falls occur during intentional movements initiated by the person, they happen mainly in forward or backward: stumbling on an obstacle during walking, backward slip on wet ground.

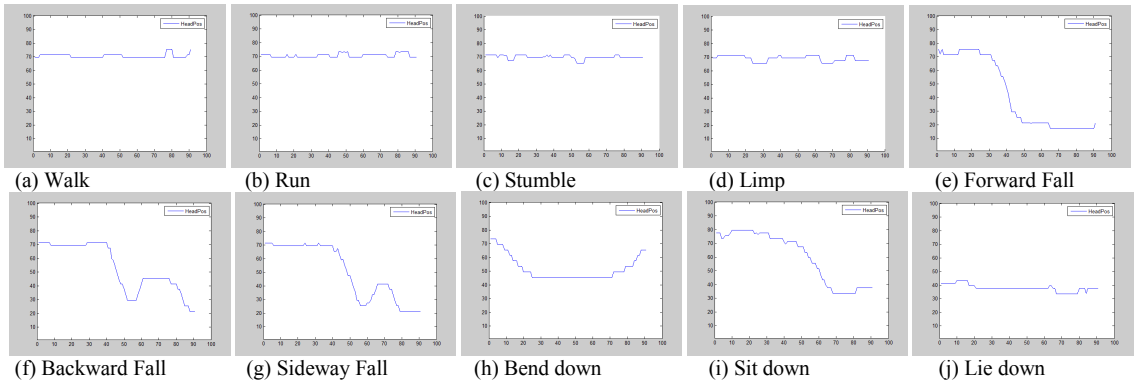


Fig. 3. Vertical Movement of Person's Head

But in some cases the fall occurs sideways, either during a badly controlled sit to stand transfer. In this case, the person frequently tries to grip the wall [1].

c) Abnormal

Stumble: Subjects were asked to walk in an unusual way, e.g. as if they were suffering a balance deficiency such as dizziness. In these situations person is unable to keep his balance or symmetry and synchrony of his movement. Body movements suggest the person is in a dubious condition.

Limp: Limping may be caused by unequal leg lengths, experiencing pain when walking, muscle weakness, disorders of proprioception, or stiffness of joints.

50 subjects with different height, weight and genders whose ages ranged from 20 to 30 were asked to participate in the project. We repeated 10 kinds of activities by 5 times in the experimental space.

B. Evaluation of the Proposed Method

TABLE I
RECOGNITION RATE FOR VARIOUS EVENTS

Events	N_a	N_c	N_f	R
Walk	125	114	11	91.20
Run	125	112	13	89.60
Stumble	125	112	13	89.60
Limp	125	110	15	88.00
Forward Fall	125	116	9	92.80
Backward Fall	125	118	7	94.40
Sideway Fall	125	114	11	91.20
Bend down	125	113	12	90.40
Sit down	125	112	13	89.60
Lie down	125	118	7	94.40

The experimental results show that the proposed system is able to achieve promising results under most test conditions. Table I represents robust recognition rate of experimental results. N_a refers to number of actions, N_c is number of correctly detected events, N_f is number of falsely detected events and R is the recognition rate.

To evaluate performance of the proposed system, we use two famous criteria that are widely used in fall detection systems. Sensitivity is the capacity to detect a fall and Specificity is the capacity to detect only a fall:

$$Sensitivity = \frac{TP}{TP + FN} \quad , \quad Specificity = \frac{TN}{TN + FP}$$

The definitions of TP, FP, FN and TN are as follows:

Fall Incident / System Recognition	**Occurs**	**Does Not Occur**
Positive	*TP*	*FP*
Negative	*FN*	*TN*

According to section V.A dataset of our experiments is composed of video sequences representing 875 normal/abnormal activities and 375 simulated falls. Table II itemizes all the results obtained with our dataset. According to above information, Sensitivity and Specificity of the proposed system would be 92.80 and 97.60 respectively.

TABLE II
EVALUATION OF RECOGNITION RESULTS

Fall Incident / System Recognition	**Occurs**	**Does Not Occur**
Positive	*348*	*21*
Negative	*27*	*854*

VI. CONCLUSIONS

In this paper we developed a novel real-time video surveillance system principally dedicated to fall detection. We claim that the proposed system is not just an ordinary

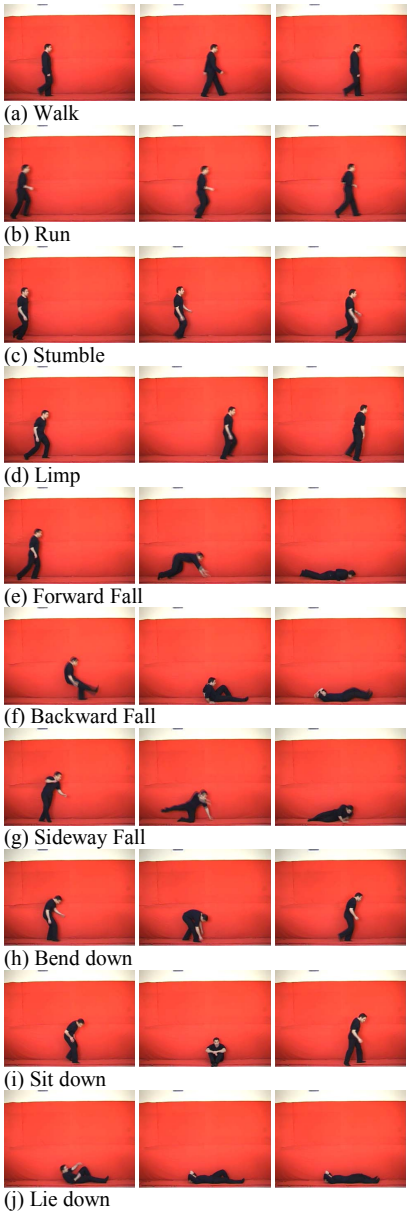


Fig.4. Example of Each Motion

human fall detection system; it has many applicable properties and can be employed in different surveillance systems. Moreover while existing fall detection systems are only able to detect occurrence of fall behavior, the proposed system is able to detect type of fall incident (forward, backward or sideway). Since temporal changes of the human shape give crucial information on human activities, we used combination of approximated ellipse around the human body, horizontal and vertical projection histograms and temporal changes of human head position as feature vectors. Also, our experiments indicate that MLP Neural Network is totally a suitable classifier for human motion recognition. Reliable average recognition rate of experimental results (91.12%) underlines satisfactory performance and efficiency of our system.

REFERENCES

[1] N Noury, A Fleury, P Rumeau, A Bourke, G Laighin, V Rialle, J Lundy "Fall detection - Principles and Methods" Conf Proc IEEE Eng Med Biol Soc. 2007 ;1 :1663-1666

[2] Arie Hans Nasution and S. Emmanuel; Intelligent Video Surveillance for Monitoring Elderly in Home Environments, International Workshop on Multimedia Signal Processing (MMSP), Greece, October 2007.

[3] Sixsmith A, Johnson N. "Smart sensor to detect the falls of the elderly", IEEE Pervasive Computing, vol. 3, no. 2, pp. 42–47, April-June 2004.

[4] Caroline Rougier, Jean Meunier, Alain St-Arnaud, Jacqueline Rousseau, "Fall Detection from Human Shape and Motion History Using Video Surveillance" Advanced Information Networking and Applications Workshops, 2007, 21st International Conference on, Vol. 2 (2007)

[5] B. T¨oreyin, Y. Dedeoglu, and A. C¸ etin. HMM based falling person detection using both audio and video. In IEEE International Workshop on Human-Computer Interaction, Beijing, China, 2005.

[6] S.-G.Miaou, P.-H. Sung, and C.-Y. Huang, "A Customized Human Fall Detection System Using Omni-Camera Images and Personal Information" Proc of Distributed Diagnosis and Home Healthcare(D2H2) Conference, 2006

[7] R. Cucchiara, A.Pratti, and R.Vezani, "An Intelligent Surveillance System for Dangerous Situation Detection in home Environments" , in Intelligenza artificable, vol.1, n.1, pp. 11-15, 2004

[8] T. Lee and A. Mihailidis. An intelligent emergency response system: preliminary development and testing of automated fall detection. Journal of telemedicine and telecare, 11(4):194–198, 2005.

[9] Nait CH, McKenna SJ. "Activity summarisation and fall detection in a supportive home environment", Int. Conf. on Pattern Recognition (ICPR), 2004.

[10] Tao, J., M.Turjo, M.-F. Wong, M.Wang, "Fall Incidents Detection for Intelligent Video Surveillance" , ICICS2005

[11] C.Rougier, J.Meunier, A. St-Arnaud, J.Rousseau, "Monocular 3D Head Tracking to Detect Falls of Elderly People", International Conference of IEEE Engineering in Medicine and Biology Society, Sept 2006

[12] Lin,C.-W.,et al.,"Compressed-Domain Fall Incident Detection for Intelligent Home Surveillance". Proceedings of IEEE International Symposium on Circuits and Systems,ISCAS 2005,2005:p.2781-3784.

[13] L. Wang, From blob metrics to posture classification to activity profiling, Proceedings of the 18th International Conference on Pattern Recognition , IV: 736-739, 2006

[14] A. Madabhushi and J.K. Aggarwal, "A Bayesian approach to human activity recognition," Proc. Second IEEE Workshop on Visual Surveillance, pp.25–32, June 1999.

[15] M.Shakeri, H.Deldari, H.Foroughi, "A Novel Fuzzy Background Subtraction Method Based on Cellular Automata for Urban Traffic Applications", 9th International IEEE Conference on Signal Processing, ICSP'08