International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII, Part 5
Commission V Symposium, Newcastle upon Tyne, UK. 2010

A REAL-TIME SINGLE-CAMERA APPROACH FOR AUTOMATIC FALL DETECTION

Anastasios Doulamis [a,*] Nikoalos Doulamis[b], Ilias Kalisperakis[b], Christos Stentoumis[b]

a: Technical University of Crete, Decision Support Systems Lab, 73100, University Campus Chania, Creece {adoulam@cs.tua.gr}

b: National Technical University of Athens, Dept. of Rural and Surveying Engineering, 15773, Athens, Greece (ilias_k@central.ntua.gr,)

Commission VI, WG VI/4

KEY WORDS: Humans' fall detection, dynamic background modelling, visual analysis, motion estimation

ABSTRACT: Elderly monitoring through video surveillance systems may be a solution for detecting humans' falls. Traumas resulting from falls in the third age have been reported as the second most common cause of death, especially for the elderly. However, such monitoring architectures are usually too costly due to the high human resources required. Furthermore, they violate privacy. For this reason, a strong research interest is recently focusing on the use of advanced computer vision tools for detecting falls of elderly people. In this paper, we propose a real-time computer vision–based system capable of automatically detecting falls of elderly persons in rooms, using a single camera of low cost and thus low resolution. The system is automatically reconfigured to any background changes which cover not only light modifications but also texture alterations. In addition, the proposed method is not sensitive to camera movement, permitting the use of active cameras in person monitoring. Another important innovation of the proposed scheme is its independence from the direction of the fall, in contrast to reported approaches in which fall is usually identified only as rapid vertical motion. To achieve these goals, the system initially extracts the foreground object from the background using a background subtraction approach. Dynamic background modelling is achieved using an iterative approach which actively updates parts of the background texture, according to the level of motion in the scene and the location of the moving object. Regarding estimation of motion information, our approach exploits motion vectors information which is directly available from the MPEG coded video streams. This information is projects onto "good" descriptors to improve system reliability.

1. MANUSCRIPT

Recently, there is a boost in ageing population which is expected to further increase in the forthcoming years, making the research of information society towards improving elderly living a pressing need. According to the American Administration on Aging the older population, (persons 65 years or older) numbered 38.9 million in 2008 (the latest year for which data is available) in USA (www.aoa.gov). This represents 12.8% of the U.S. population, about one in every eight Americans. By 2030, there will be about 72.1 million older persons, more than twice their number in 2000. Even worsen statistics are noticed for most of European countries (Richard, 2003). In this direction, fall detection for the elderly is of primary importance. Traumas resulting from falls in the third age have been reported as the second most common cause of death. Additionally, it causes most of the times major accidents and movement impairments with concomitant consequences to their lives and the surroundings (Wang, 2005a). Thus, as falls and fall-related injuries remain a major challenge in the public health domain, reliable and immediate detection of falls is important so that adequate medical support can be delivered.

Currently, the most common way for recognizing humans' falls is to use accelometers (Brownsell, 2004) or wearable devices (Nyan, 2008) or gyroscope sensors (Nyan, 2006) or even combination of different sensors (Bourke, 2008). However, such approaches prevent elderly from an independent living since they force to wear medical devices all the time, some of them cannot be hidden. In addition, they are mostly dependent on device specifications and functionalities. For instance, they either assume that the patients are functionally capable of pushing buttons (this is not the case e.g., of dementia patients) or expect them to wear specialized devices or even demand expensive equipment not always affordable by small clinics and nursing institutes.

Another useful application towards falls detection is to use wireless sensor networks (Wang, 2005b). Wireless patient monitoring systems improve quality of life for the subject by granting them more freedom to continue their daily routine, which would not be feasible if wired monitoring equipment were used. However, again these networks are suitable only for specific/ constrained environments.

Elderly monitoring through video surveillance systems may be a solution used in clinics or nursing institutes. However, such architectures are usually too costly due to the high human resources required. Furthermore, they violate privacy (someone survey private moments of the elderly). For this reason, we need to embed intelligent mechanisms into the monitoring systems so that they focus only on detecting salient events of interest (falls) without transmitting any other visual information and thus maintaining privacy.

Recently in the literature some works have been presented that exploit visual information for detecting falls. Some of these approaches are based on humans' motion estimation in complex background conditions, such as the ones encountered in our homes. Furthermore, machine learning and cognition algorithms are included in these approaches to distinguish falls from other daily humans' activities, such as sitting, walking, standing, carrying objects, etc.

In particular in (Fu, 2008) an *Asynchronous Temporal Contrast* (ATC) *vision sensor* which features sub-millisecond temporal resolution has been proposed for detection of humans' falls. Having estimated the moving regions, their centroids are then extracted and a fall is alerted when a significant vertical velocity is calculated. Furthermore, the work of (Thome, 2008) exploits multiple views for detecting the falls using a Layer Hidden Markov Model (LHMM). In (Qian, 2008) work, a background subtraction method is adopted to identify humans' and then cascading supoort vector machines are exploited to categorize different types of humans' activities. Finally, (Foroughi, 2008) adopts an elliptical model to estimate the shape of humans' body and then exploits support vector machines to categorize these shapes as falls or other humans' activities.

However, these algorithms either assume static background models which perform poorly when dynamic changes of the background are encountered (as is the case of real–life situations), or exploit multiple camera information to define human's height and consequently his/her fall. Furthermore, the computational complexity of some of the aforementioned systems is high, preventing the real-time performance which is necessary for large scale implementation. Fall should be detected in real time (or almost in real time) but also not be confused with other actions (sitting, bending or even leaving/entering the camera view). Furthermore, specific directions of the falls are assumed, usually perpendicular to the camera view which is too restrictive for real life applications. In a real world surveillance system, a person can move in all directions of the room, and thus his/her fall does not necessarily imply vertical motions. Besides, the system should be robust to any background change and thus demand minimum re-configuration to dynamic changes of the environment and lighting conditions.

In (Doulamis N., 2010) a visual fall detection scheme has been proposed using an iterative motion estimation algorithm which is constrained by time and shape rules. The concept behind this paper is that humans' motion should be coherent and thus any significant modification in motion and shape of the tracked objects should re-force a new estimation of the visual background conditions unless a fall is detected. Thus, (Doulamis N., 2010) work is sensitive to environmental changes and it can be only applicable for slightly modified backgrounds, while the algorithm fails in case of occlusions, or presence of multiple persons. A more robust approach for extracting foreground objects from background in more complicated visual conditions have been proposed in (Doulamis A., 2010). This work uses adaptive boosting strategy (adaboost) to model the visual statistics of the background/foreground objects and thus improving the visual analysis. However, the work of (Doulamis A., 2010) does not directly deal with fall detection while it assumes a limited number of foreground objects and thus it is not suitable for the persons' fall application scenario.

In this paper, we propose a real-time computer vision–based system capable of automatically detecting falls of elderly persons in rooms, using a single camera of low cost and thus low resolution. The system is automatically reconfigured to any background changes which cover not only light modifications but also texture alterations. *In addition, the proposed method is not sensitive to camera movement, permitting the use of active cameras in person monitoring.* Another important innovation of the proposed scheme is its independence from the direction of the fall, in contrast to reported approaches in which fall is usually identified only as rapid vertical motion.

The first component of the proposed system is the automatic extraction of the foreground object from the background. This is performed using a background subtraction approach, enriched with *on-line learning strategies* able to work in highly dynamic scenes. Dynamic background modelling is achieved using an iterative approach which actively updates parts of the background texture, according to the level of motion in the scene and the location of the moving object. This is achieved using non-linear approaches for modelling the background such as the Learning Vector Quantization. This technique is very fast and thus background modelling can be calculated in real-time.

Regarding estimation of motion information, our approach exploits MPEG coded motion vectors so as to minimize the cost in computing the iterative algorithm since n decoding is required. However, MPEG motion vectors are estimated for coding purposes and not for analysis and thus they are very sensitive to noise (Doulamis, 2000). For this reason, we estimate only the most confident motion vectors as the ones derived using edge-based information. Towards this direction, the Shi-Tomasi algorithm is taken into account (Shi, 1994). Selection of good features increases the robustness in detecting motion in a scene. On the contrary, application of conventional motion detector methodologies to the all image pixels would result in motion vectors of significant magnitude in regions where no real motion exists due to slightly camera adjustments and lighting conditions.

More specifically, in case that a large connected component of motion is detected in the scene the background is dynamically updated. In particular, all non-moving regions are considered as the new background candidate and therefore this information is accumulated to the existing background models. On the contrary, the moving regions are considered intact.

The second component of the proposed system is responsible for detecting a human's fall by exploiting the dynamic foreground localization and tracking methodology described above. The main difficulty in this case is the need to distinguish human's falls from other regular human activities (sitting, bending). In particular, the goal is to minimize false positive and false negative results and thus significantly increase the reliability of the fall detection system, making it applicable to real world conditions. This is even more challenging under a single camera system due to the geometric limitations in depth reconstruction as compared to stereovision.

2. FOREGROUND EXTRACTION

The first step of the proposed Person Fall Detection sensor is the extraction and robust tracking of the foreground object. However, robust and continuous foreground detection is a very challenging research task, especially in real-life application scenarios where complex and dynamic background conditions are encountered. Usually, foreground extraction is accomplished using background modelling methods followed by background subtraction techniques. In this paper, we extract foreground object using classification based background modelling approaches, which are described in the following.

2.1 Classification Background Modelling

In this paper, we model the background visual statistics a modification of the algorithms of (Heikkila, 1999) using non-linear learning strategies as if of Learning Vector Quantization-LVQs (Kohonen, 2001).

According to (Heikkila, 1999), let us denote as $I(t)$ an image (or a video frame) at a time instance t. Let us also denote as $B(t)$ the background region at the same tine t. Then, an image pixel is said to belong to a foreground region if

$$|I(t) - B(t)| > \tau \tag{1}$$

where τ is a predefined threshold. Morphological 3x3 closing operators can be applied on (1) to remove noisy pixels and thus improve the accuracy detection. Then, the background is updated as

$$B(t+1) = aI(t) + (1-a)B(t) \tag{2}$$

where a is kept small to prevent artificial prevent artificial "tails" forming behind moving objects.

Then, two background corrections are applied:

1. If a pixel is marked as foreground for more than *m* of the last *M* frames, then the background is updated as $B(t)=I(t)$. This correction is designed to compensate for sudden illumination changes and the appearance of static new objects.

2. If a pixel changes state from foreground to background frequently, it is masked out from inclusion in the foreground. This is designed to compensate for fluctuating illumination, such as swinging branches.

This algorithm, however, is very sensitive to illumination changes, although it is very simple. For this reason, we modify it in this paper introducing non-linearlities via a learning vector quantization algorithm. This way, we group together similar statistics of the background reducing noisy oscillations from frame to frame and while keeping the computational complexity as low as possible. The main characteristic of the LVQ algorithm is its ability to be implemented in real time making the proposed humans' fall detection algorithm applicable in real-life conditions.

2.1.1 Background Modeling Using on-Line Learning

Let us assume that we have available the color statistics for the background. The exact way on how we perform this automatically is described in Section 3. Then, the proposed scheme aims at learning these background statistics dynamically as new content is being captured so as to classify these regions as background or foreground areas. Let as denote as $S_b(t)$ the training set of the background at a time instance *t*. In this paper, the training set expresses color information of background region.

Given the set $S_b(t)$, one can estimate a number of classes (color representatives) that they model as much as possible the visual characteristics of the background, as is described by $S_b(t)$. Let us denote as $\mathbf{m}_i$, with $i=1,2,\ldots,K$ the *K* representatives for the background. The optimal number for $\mathbf{m}_i$ can be estimated through the generalized Lloyd-Max algorithm. This algorithm can be implemented through the k-means (Tan, 2005).

The k-means implementation for the Lloyd Max algorithm actually handles the incoming data in a batch mode. This means that it groups together background regions that share common color properties without taking more in account the current statistics than the previous ones. That is, it defines the *K* color representatives around which the background pixels will be related.

Another dynamic implementation for the $\mathbf{m}_i$ representatives is through Learning Vector Quantization (LVQ). LVQ is one example of competitive learning. The goal here is to have the network "discover" structure in the data by finding how the data is clustered. In vector quantization, we assume there is a codebook which is defined by a set of *M* prototype vectors (*M* is chosen by the user and the initial prototype vectors are chosen arbitrarily). An input belongs to the ith cluster if *i* is the index of the closest prototype (closest in the sense of the normal Euclidean distance). This has the effect of dividing up the input space into a Voronoi tessellation. In particular, in LVQ approach, we randomly pick one sample out of the available ones. Then, we find the "winning" node for this sample, i.e., the cluster whose the distance from this sample is smaller than any other distance with other clusters, meaning that the selected sample and the respective node winner node if the optimal choice we can made at this stage of the algorithm. Then, we update the weights for the winning cluster to take into account this new information.

$$|\mathbf{w}_k - \mathbf{x}| < |\mathbf{w}_i - \mathbf{x}|, \quad \text{for all}\, i \tag{3}$$

In particular, the algorithm starts at the 0$^{\text{th}}$ iteration in which it assumes that *K* representative codebooks have been randomly derived, denoting as $\mathbf{m}_i(0)$ where in this case we have added the dependence of the $\mathbf{m}_i$ on the iterations of the algorithm. Then, picking up a random sample $\mathbf{x}$ and denoting as $\mathbf{m}_{k'}(n)$ the best (closest) representative for this sample at the nth iteration of the algorithm, then, this representative is updated so that

$$\mathbf{m}_k(n+1) = \mathbf{m}_k(n) + \mu(\mathbf{x} - \mathbf{m}_k(n)) \tag{4}$$

while the other representatives remain intact. Parameter μ defines the convergence rate for achieving optimality.

<u>Main Steps of the Proposed Hybrid Algorithm</u>

1. Take a set of *M* previous backgrounds.
2. Assume that the representatives for these backgrounds are available
3. Group each incoming sample with respect to these representatives
4. If some samples within a group deviates a lot create a new group
5. Update the representative values using the LVQ methodology
6. Merge groups whose representatives are too close

Table 1. Summary of the main steps of the proposed hybrid scheme for estimating background representatives.

In our case, the background content can be dynamically modified. Thus, for each captured frame a new background set $S_b(t)$ is estimated. If we take into account only the statistics for the current background, then we totally ignore (forget) the previous information which may be important for future background samples. Suppose for instance a background object which is fully occluded by humans as they are randomly walking in the surveyed area. If we would forget the previous samples, then we cannot recall the color statistics for this object so as to correctly categorize the background even for future information.

In this paper, we adopt a hybrid methodology which appropriately modifies the previous knowledge exploiting on-line learning methodologies. Let us assume that *M* previous backgrounds are available. Then, their statistical properties can be included in sets $S_b(t)$, $S_b(t\text{-}1)$, $S_b(t\text{-}2)$,…, $S_b(t\text{-}M)$. For each of the previous backgrounds, we have estimated some representatives using the proposed algorithm. We should mention that the number of representative vectors per background is not constant since this actually depends on dynamicity of the content. Then, each new incoming sample is grouped with respect to these representatives. These groups are merged or split with respect to algorithm proposed in Section 2.1.2. Then, we update the values of the representative samples using the LVQ algorithm

2.1.2 Automatic Estimate of the Number of Representatives

If the distance of a picked sample is far away from all *K* representatives, a new representative is created since this means that the sample cannot be described by any of the existing codebooks. In this case, we set the new representative equals the new sample and we follow as in Section 2.1.1.

Similarly, if the values of two representatives are too close we remove one of the two since they refer to almost similar content. Between the two, we remove the representative that less describes the variations and information of all samples assigned to it. This way, we automatically update the number of representative per background area. Table 1 summarizes the main steps of the proposed algorithm.

3. CONFIDENT SELECTION OF THE BACKGROUND SAMPLES

In this section, we describe an algorithm that estimates the most confident regions in an image that should be characterized as background. This is accomplished by collecting a set of images (video frames) which localise the foreground region and thus they indirectly refer to the background. It should be mentioned that we pay no particular attention on whether foreground objects are present when the background modelling algorithm is activated. These objects will be actually removed from the scene due to the dynamic nature of our method. This constitutes the main innovation of the proposed algorithm; the background can be dramatically changed from time to time and despite that the proposed scheme is able to be adapted to the new conditions (after the elapse of some frames). As a result, it is probable for some frames (especially right after a major change of the background content) to derive a rough approximate description

3.1 Creating the background Samples

The main difficulty of the aforementioned approach is the automatic estimation of the set of background region statistics, or in other words, the automatic estimation of the background region. In this paper this is achieved indirectly by firstly detecting the foreground region and then estimating the background as the complementary set of the foreground.

For the foreground object, we assume that it possess the largest portion of motion activity. Of course, parts of the background can be moved in such a dynamic environment (either due to periodic oscillation, or due to sensitivity in capturing noise, or even due to dramatic environmental changes,-illumination changes, background objects that are moving, etc). However, the largest and most coherent part of moving activity is for the foreground object. In the following, we exploit such property to identify the foreground region in a frame and then indirectly the background area.

The motion vector information as being directly available from the Moving Picture Expert Group (MPEG) compression is exploited in this paper as an indicator of motion activity. Other information can be used instead of motion vectors, as if the optical flow. However, motion vectors are available without requiring to decode the MPEG video streams. This saves computational complexity and storage capabilities.

Motion vectors are available for P and B frames of MPEG sequences. For the I frames one can estimate the moving information using knowledge from the previous and the following P/B frames.

It is clear that MPEG motion vectors have been estimated coding purposes and not for a moving region segmentation. This means that they present high noise as far as the actual objects' motion is concerned. In MPEG coding, motion vectors are embedded in video streams along with compensation errors so as to reduce the information needed to be stored by exploiting motion activities. As a result, even random motion vectors can be useful for decoding purposes (of course no efficient) since the compensation error (in this case, it takes large values) can correct any mistake in actual motion estimation. On the other hand, these values are not proper for detecting the moving information for a foreground region.

To improve motion analysis in this paper, we project motion information onto "good" ("feasible") descriptors. This way, instead of calculating the whole motion information in an holistic approach we restrict this analysis on motions that present high probability of being foreground region, rather than on areas that seems to be background.

Towards a feature-based selection, we use the algorithm of (Shi, 1994). Good descriptors are, for example, the corners in an image, the locations of zero crossing of the image Laplacian, the standard deviation in the spatial intensity, etc. However, even a region is rich in texture information can be poor in tracking. For instance, it can straddle a depth discontinuity or a boundary of a light reflection. Even worst, the good features can be also occluded and the trackers can be drifted far away from their original target.

For this reason, in this paper, we implement the work of (Shi, 1994) on corners. In particular, we initially calculate the minimal eigenvalue for every pixel of the source image and then performs non-maxima suppression around a neighboring area (in our case an area of 3x3 neighboring block is selected). The next step is to filter out the corners from possible noise. Thus, we exclude from the features the corners that present minimal eignevalues less than a preset threshold activity.

4. HUMANS' FALLS DETECTION

Detection of vertical motion trajectories for estimating humans' falls as adopted in (Fu, 2008) significantly restricts humans' position in the visual environment and the way that the fall takes place. Usually, when survey a room we place single cameras on the top corners of the columns and it quite probable the fall to take place in other than vertical positions with respect to the camera. Humans' are not walking in front of cameras. Instead, they are passing throughout the room in arbitrarily directions and falls can occur either across, or in an oblique way.

In addition, the recognition of a fall from the objects shape (Foroughi, 2008) is sensitive to the optical parallax that the object is be seen by the camera. Even in the seldom case that the fall will occur vertically to cameras' view, humans' shape is totally different if he/she falls towards the camera (dramatic reduction of the shape) or in the opposite direction to the view. This means in, other words, that it is probable to confuse a fall with another regular humans' activity, i.e., sitting. To be sit in a chair towards cameras' view can be too similar to fall in the opposite way. More complicated effects on humans' shape can be encountered in the usual case that falls are oblique or even horizontal with respect to the visual view.

To address this difficulty, in this paper, we estimate the fall by following the trajectory of the upper boundary of the foreground object. This criterion is much more sensitive to human fall events. In particular, let us denote as $y(t)$ the vertical position of the upper boundary of the detected foreground object at time *t*. Then, we define the vertical velocity (acceleration during the fall) as the time derivative

$$J = \frac{y(t+1) - y(t)}{dt} \tag{5}$$

It is clear that the aforementioned equation presents a measure of how the foreground object is changing through time. Thus, it can be considered as a measure for fall alert.

The main problem of (5) is that it does not take into account the size of the human object. It is clear, however, that human objects close to the camera are bigger than object far from the camera. Thus, we need to face differently each of these objects. For this reason, we normalize $y(t)$ with respect to the initial detected foreground height.

$$\bar{J} = \frac{[y\ (t+1) - y(t)]/\,y(t)}{dt} \tag{6}$$

Equation (6) can be seen as a proper metric for falls detection. The experimental results that we propose in Section 5 show the advantage of this method in recognizing falls in complicated visual environments.

5. EXPERIMENTAL RESULTS

Initially, we test the performance of the proposed algorithm in our lab using two different types of experiments. In the first experiment, the background is modified by moving objects in it. In the second experiment, the camera is moving to severely change the background visual conditions. Figure 1 shows some characteristics original samples for the first experiment, while Figure 2 for the second.

Figure 1. Characteristic frames for the 1st experiment.

Figure 2. Characteristic frames for the 2nd experiment.

Figure 1 shows a man who is moving in a room. Sometimes he is falling but the falls are detected by the system. He also makes other humans' activities, such as sitting or carrying objects. Moreover, sometimes he is approaching the camera. But in all cases, the fall detector algorithm works efficiently. In Figure 2, the man moves the camera to change the view. But even in this case, the background subtraction algorithm can efficiently identify the person and distinguishes falls from other humans' actions.

In Figure 3, we depict the background region as being extracted for the 1st experiment. We can see that at the first frames there is a gradual transition from the content of the original image to the background content. For this reason, initially in the background parts of the human are also included. This is, however, a temporal situation and lasts only for very few frames (bout ten). Afterwards, the system converges to the actual background region, and despite the movements of the objects and the complicated motions of the human the background can be extracted with high accuracy in all cases. We also observe from Figure 3 that gradually the background is updated with the current objects (as if the door) improving the fall detection accuracy.

This is more evident in Figure 4 in which at a time the camera is moving and the background significantly changes, We can see that the proposed algorithm is very robust and it can detects the background even in such dramatic changes within very few frames. In addition, despite the complexity of the background content the algorithm remains very stable.

Figure 3. Characteristic frames for the detected as background region for the Experiment 1.

Figure 4. Characteristic frames for the detected as background region for the Experiment 1.

Figure 5 illustrates the estimate for the foreground object using a background subtraction methodology. We should stress that the goal of this paper is not to apply a very sophisticated moving object detection scheme. On the contrary, our goal is to identify falls in a real-time methodology. The proposed algorithm is very fast and can be implemented in real-time despite the small time constrains that a video stream demands. As we can see from Figure 5, at the first frames the algorithm fails due to the transition period in identifying the background. We should mention that we impose no restrictions on the initial content of the video. Then, the algorithm can efficiently detect the objects and but fails in case of sever background changes. This failure, however, is temporary since within very few frames the algorithms can self improved again.

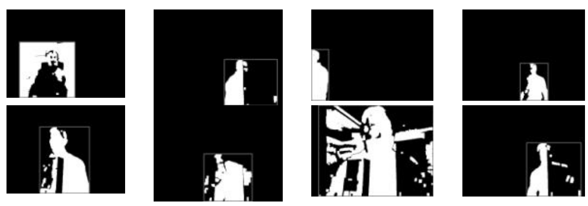

Figure 5. Foreground detection using the proposed scheme for the Experiment 1.

In addition, for the purposes of fall detection the algorithm is also very robust. Table 2 shows the accuracy results in terms of false positive and false negative errors. These criteria calculate the percentages of actual falls that are not detected by the system as well as the percentage of erroneous falls that have been detected as falls by the system (confusing among falls and other humans' activities). The first case is surely more severe, since we ignore incident of interest. Instead, high values for the second criterion means that we need to activate false alarms many times. The results have been derived as the average performance for two days videos in a real-life clinic.

Fall Detection Performance	
False Positives	False Negatives
11.60%	23.54%

Table 2. Fall detection results on a two-days real-life video stream recorded in a clinic.

6. CONCLUSIONS

In this paper, we propose a novel algorithm which can be used for detecting humans' falls. The algorithm is relied on a visual monitoring system. The proposed scheme is capable of automatically detecting the falls using a single camera of low cost and thus low resolution. The system is automatically reconfigured to any background changes which cover not only light modifications but also texture alterations. In addition, the proposed method is not sensitive to camera movement, permitting the use of active cameras in person monitoring. Another important innovation of the proposed scheme is its independence from the direction of the fall, in contrast to reported approaches in which fall is usually identified only as rapid vertical motion.

The system consists of two components. The first is the automatic extraction of the foreground object from the background. This is performed using a background subtraction approach, enriched with on-line learning strategies able to work in highly dynamic scenes. Dynamic background modelling is achieved using an iterative approach which actively updates parts of the background texture, according to the level of motion in the scene and the location of the moving object. The second is responsible for detecting a human's fall by exploiting the dynamic foreground localization and tracking methodology described above. The main difficulty in this case is the need to distinguish human's falls from other regular human activities (sitting, bending).

Experimental results on real-life video streams with complicated and dynamic visual backgrounds verify the fact that the proposed system can robustly and efficiently identified humans' falls even under very complex conditions.

7. REFERENCES

Bourke, A. K., O' Donovan, K. J., Ó Laighin, G., 2008. The identification of vertical velocity profiles using an inertial sensor to investigate pre-impact detection of falls. *Medical Engineering & Physics*, 30(7), pp. 937–946.

Brownsell S., and Hawley, M. S., 2004. Automatic fall detectors and the fear of falling. J. Telemed. Telecare., 10, pp. 262 266.

Doulamis, A., Doulamis, N., Kollias, S., 2000. Non-sequential video content representation using temporal variation of feature vectors. *IEEE Trans. on Consumer Electronics*, 46(3), pp. 758-768.

Doulamis, A., Doulamis, N., Dragonas, J., Miaoulis, G., Plemenos, D., 2010. Robust Foreground Detection and Tracking using Dynamic Foreground-Background Modeling Based on Adaptive Boosting. *International Conference on Artificial Intelligence and Graphics (3ia)*, Athens, Greece.

Doulamis, N., 2010. Iterative motion estimation constrained by time and shape for detecting person's falls. *ACM 3rd Inter. conference on Pervasive Technologies Related to Assistive Environments*, Samos, Greece.

Foroughi, H., Rezvanian, A., Paziraee, A., 2008. Robust fall detection using human shape and multi-class support vector machine. *IEEE Computer Vision, Graphics & Image Processing, (ICVGIP)*, India, pp. 413 – 420.

Fu, Z., Culurciello, E., Lichtsteiner, P., Delbruck, T., 2008. Fall detection using an address-event temporal contrast vision sensor. *IEEE International Symposium on Circuits and Systems (ISCAS)*, Seattle, WA, USA, pp. 424-427.

Heikkila J., Silven, O., 1999. A real-time system for monitoring of cyclists and pedestrians. *Second IEEE Workshop on Visual Surveillance* Fort Collins, Colorado, pp. 74-81.

http://www.aoa.gov/AoARoot/Aging_Statistics/index.aspx (access 10 April 2010).

Kohonen, T., 2001.Self-Organizing Maps. *Third ed. Berlin Heidelberg: Springer-Verlag*

Nyan, M.N, Tay, E.H.F, Tan, A.W., Seah, K.H., 2006. Distinguishing fall activities from normal activities by angular rate characteristics and high-speed camera characterization. *Medical Engineering & Physics*, 28(8), pp. 842–849.

Nyan, M.N., Tay, E.H.F, Murugasu, E, 2008. A wearable system for pre-impact fall detection. *Journal of Biomechanics*, 41(16), pp. 3475-3481.

Qian, H., Mao, Y., Xiang, W., Wang, Z.,2008. Home environment fall detection system based on a cascaded multi-SVM classifier. *10th IEEE International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 1567-1572.

Richard, B., 2003. Aging Europe finds its pension is running out. The New York Times, New York, USA.

Shi, J., Tomasi, C., 1994. Good features to track. *Intern. Conf. on Comp. Vis. & Patt. Rec*, pp 593-600.

Tan, P.T., Steinbach, M., Kumar, V., 2005. Introduction to data mining. Addison-Wesley, ISBN : 0321321367

Thome, N., Miguet, S., Ambellouis, S., 2008. A real-time multiview fall detection dystem: A LHMM-based approach. *IEEE Transactions on Circuits and Systems for Video Technology*, 18 (11), pp. 1522-1532

Wang, S., Yang, J., Chen, N., Chen, X., Zhang, Q., 2005a. Human activity recognition with user-free accelerometers in the sensor networks. *International Conference on Neural Networks and Brain*, pp. 1212-1217.

Wang, S., Yang, J., Chen, N., Chen X., Zhang, Q., 2005b. Human activity recognition with user-free accelerometers in the sensor networks. *Proc. International Conference on Neural Networks and Brain*, pp. 1212-1217.