ITERATIVE SCENE LEARNING IN VISUALLY GUIDED PERSONS' FALLS DETECTION

Anastasios Doulamis, Konstantinos Makantasis

Decision Support Laboratory
Technical University of Crete, 73100, Chania, Crete, Greece
phone: + (30) 28210 37430, fax: + (30) 28210 69410, email: adoulam@ergasya.tuc.gr
web: www.depm.tuc.gr

ABSTRACT

This article describes a fast real time computer vision algorithm able to detect humans' falls in complex dynamically changing visual conditions. The algorithm exploits single cameras of low cost while it requires minimal computational cost and memory requirements. Due to its affordability it can be straightforwardly implemented in large scale clinical institutes/home environments. In this paper, we evaluate the performance of this algorithm into two different real-world conditions. The evaluation was performed for long time and concerns robustness compared to other humans' activities, false positive/negative estimates, all in real time.

1. INTRODUCTION

Population in developed countries (Europe, USA, Japan, etc) is ageing. This has as consequence that the aging population will of course be subject to all relevant health problems that this population portion is associated with. However, the quality of life for elderly is associated with their ability to live independently and with dignity without having the need to be attached to their children, grand-children or any other person whose help would they need for their daily life and social behavior [1]. On the other hand, according to medical records, traumas resulting from falls in the third age have been reported as the second most common cause of death for the Elderly. The most socially worst thing for falls is the subsequent effects that this causes; movement impairments, bones fractures, partial paralysis, along with other concomitant consequences to their lives and the surroundings [1]. This is also important with people who suffer from dementia.

The dementia is one of the problems that hinder these people's ability to have such an independent life, making necessary the presence and monitoring of their daily activities by care-givers. Dementia (from Latin de- "apart, away" + mens (genitive mentis) "mind") is the progressive decline in cognitive function due to damage or disease in the brain beyond what might be expected from normal aging. Although dementia is far more common in the geriatric population, it may occur in any stage of adulthood. This age cutoff is defining, as similar sets of symptoms due to organic brain dysfunction are given different names in populations younger than adulthood. An estimated 26.6 million people worldwide were afflicted by Alzheimer's in 2006; this number may quadruple by 2050.

For this reason, a major research effort has been conducted in the recent years for automatically detecting persons' falls especially for the Elderly. Such identification is a prime research issue in computer vision society due to the complexity of the problem as far as the visual content is concerned. For instance, the algorithm should ideally (a), detect falls in real time (or at least just in time), i.e., without losing the resolution accuracy for the fall detection, (b) be robust to background changes and illumination changes, (c) be robust when more than one persons' are present in the scene, (d) identify falls occurring in any position with respect to the camera and (e) be tolerant to camera changes (active cameras).

1.1 Previous Research Approaches

Currently, the most common way for detecting persons' falls are through the use of specialized devices, such as accelerometers [3][2],[4],[5] floor vibrations [6], combination of accelometry with barometric pressure [7], wearable equipment [8], gyroscope sensors [9], or combination/fusion of them [10]. However, such approaches are device dependent and present a series of drawbacks; they prevent the elderly- or cognitive disable persons from being normally function as all we do in our lives since they impose them to wear specialized devices.

A more research challenging alternative is the use of visual cameras. A characteristic work for fall detection is the *Asynchronous Temporal Contrast* (ATC) *vision sensor* which features sub-millisecond temporal resolution [11]. Then, region centroids are calculated and falls are alerted when significant *vertical velocity* is detected. The work of [12] proposes a multiview camera system for detecting the falls. In this approach, a Layer Hidden Markov Model (LHMM) is adopted to model motion activity, while the posture classification is performed by a fusion unit that independently fused information provided by the other cameras under a fuzzy logic context. A pattern analysis algorithm that discriminates falls events form slip events using energy maps is reported in [13]. Finally, a 3D active vision has been proposed in [14].

1.2 The Proposed Contribution

In this paper, we evaluate the performance of a real-time computer vision algorithm, demanding low processing and memory requirements, able to identify humans' falls in complex, dynamic in terms of background visual content, and unexpected environments, like the ones encountered in real-world clinical and/or home conditions. Very recently, we have proposed some other variations of the presented approach with the main goal to increase fall detection with respect to dynamic visual changes, camera motion, illumination variations and be independent from the position of the fall. In particular, in [15], the author has proposed a new iterative motion estimation algorithm for accelerating computational processing and he has exploits rules schemes to conclude to alerts. However, as the same author mention in [16] "motion features are sensitive to noise" and therefore the work of [15] is still not sufficient for a large scale implementation of long-time monitoring. Towards this direct learning vector quantization techniques have been proposed in [17] as a powerful tool for detection falls in real-life and dynamically modified conditions. In [18], the approach of [17] has been extended using more sophisticated background modelling techniques, exploiting concepts of Gaussian Mixtures Modelling. However, although these algorithms can be implemented in real-time, they demand a relatively high computational load, preventing from a large scale implementation for an affordable cost. This issue is addressed in this paper by proposing a computationally efficient but visually robust fall detection service. In particular, we perform a low cost and affordable algorithm able to detect humans' falls and we evaluate this algorithm in real world conditions.

2. A VISUAL FALL DETECTION OF LOW COST AND MEMORY

The first step of the proposed algorithm is to identify the foreground objects(s) and separate it (them) from the background. Then, we analyze the trajectories of the moving objects to find out the falls.

2.1 Foreground Objects' (Humans) Detection

In general background subtraction techniques are not so simple in computational cost and memory and they fail for a large scale implementation. For this reason, in this article, the intensity of motion vectors along with their directions is exploited to identify humans' movements. However, motion vectors are still very sensitive to luminosity changes and color/camera parameter variations. For instance, a different focus of the camera, which is a continuous usual process of the current camera sensors may result in estimating of large intensity values of motion vectors, though no motion information is encountered in the captured image frames [16]. For this reason, initially we apply the methodology of [19] so as to detect corners, edges and other salient points on video frames and these points to be considered as "good" locations for estimating motion vectors. This way, a significant reduction of the cost is accomplished. To further accelerating the time, avoiding the eigenvalues calculation of [19], we estimate the absolute difference of two subsequent frames and then thresholding this difference to get a binary mask indicating areas of high motion information. Then, the feature points are extracted by constraining the aforementioned areas of high motion activities on the foreground object mask in previous frames. Then, we spatially sampling the initially detected feature points by retaining the local maximum feature points, within a neighbouring region. Finally, we apply morphological operators to clarify the results from noise.

Then, we overcome the problem of decoding for the video frames by exploiting the motion vectors of the MPEG encoding data streams vectors as motion vectors. Of course, these vectors have been calculated for encoding purposes and not as motion detectors but they can provide a rough estimate of the activity in the scene. Motion vector information is embedded only in P and B frames while it is absent from I frames. Although there exists techniques that permit estimation of motion vectors for the I frames by taking into account the embedded knowledge on neighbouring P and B frames with a low computational complexity, we adopt in this paper to eliminate motion information and thus background updating for I frames so as to minimize the additional cost. Since usually I frames are encoded every 10 or 12 frames this assumption is not critical for the performance of the visual fall detection service.

2.2 Background Modelling and Updating

To reduce false alarms of falls detection under dynamic visual conditions, we need an updating mechanism for the background content. In addition, we need this mechanism to be computationally efficient so that it can be included in devices of low capabilities allowing large scales applications scenarios of IT technology in aiding the elderly or persons' with dementia.

In our case motion information is used as background updating stimuli, instead for foreground object extraction. Background updating takes into consideration motion information. In particular, regions which are spatially far away from the motion activity segments are selected to be background areas. Then, using background subtraction techniques, we provide estimates of the foreground objects.

More specifically, we assume that the background is updated at every frame instance. In particular, we initially estimate the intensity in motion activity estimation. Let us denote this intensity as E within a foreground area. If E is greater than a threshold value then, in the area of detecting a foreground object, we should update the background values since a foreground object has appeared and covered parts of the background. Otherwise there is no important variation in the scene which imposes that there is no need for background updating.

When background updating should take place, we filter the detected regions so that the ones that are far away from the estimated moving regions to be selected as new background while the ones that are close to the foreground objects to be considered as ambiguous regions. These regions are defined by a rectangular that includes the left-right upper-down extreme boundary motion vectors locations.

2.3 Updating Foreground Regions

The foreground objects are directly estimated from the background model, using a background subtraction approach. This mask is then used for estimating good feature points for motion activity which are exploited for background updating.

2.4 Falls' Detection

One possible measure for detecting falls is through the use of vertical velocity. This is considered in this paper as one possible metric. The vertical velocity is measured as the movement of the horizontal line of a connected object instead of the centre of gravity to avoid blenching of humans. However, vertical velocity is not always an appropriate measure in the process of humans' falls identification. This is due to the fact that a fall can be occurred in opposite directions of a cameras' position. To address this difficulty, we

a) estimate the accumulation of the vertical over several frames times (to handle cases erroneous estimation of the foreground mask is extracted resulting in abrupt changes of vertical velocity) and

b) constrain the cumulative vertical velocity on ratio of the foreground object height among a small series of successive image frames.

More formally, let us denote as $v(k)$ the height of the foreground object at the kth image frame and as $v(k-m)$ the respective height at the *k-m* frame where *m* is a fixed small integer number, e.g., four or five. Then, a metric for falls alert is estimated as

$$FD = \sum_{i=k-v}^{k} \frac{y(i)-y(i-1)}{y(i)} \frac{v(k)}{v(k-m)} \qquad (1)$$

We call this metric as second option in the experiments. If this metric is higher than a threshold then falls are detected. Otherwise, other human activities are supposed to be executed in the scene. The adoption of the ratio $v(k)/v(k-m)$ is to help to system for distinguish other human activities, such as sitting, walking, bending, etc. These way false alarms are reduced. In addition, we adopt an adaptive threshold selection in order to face the cases where the foreground object is far away or close to the camera. In particular, we adopt a linear relation with respect to the object area, which is calculated as the square sum of the number of the foreground object pixels.

3. SYSTEM EVALUATION

The evaluation of the results was conducted in two different environments [1]. The first concerns a martial arts school in Chania Greece while the second the location of real-life home environments of persons with dementia with the assistance of Trikala City municipality. The code was implemented in OpenCV. All results have been taken using the system for continuous test for more than 4 months.

(a) (b) (c)

Figure 1 – Characteristics examples of the environment recorded along with the background changes

The system was tested in different cases, including camera position (meaning that active cameras scenarios can be also supported), changes in the illumination conditions rapid and fluctuations in the background. We should mention that the results in the martial arts school in Chania were focused on background which the sun light is reflected on mirrors making the illumination changes really demanding (see Figure 1).

During the experimentation process, some persons may fall in every direction according to the camera while they can perform normal every day activities, like sitting on a chair. In Figure 2, we present an example of two humans detected by the proposed algorithm.

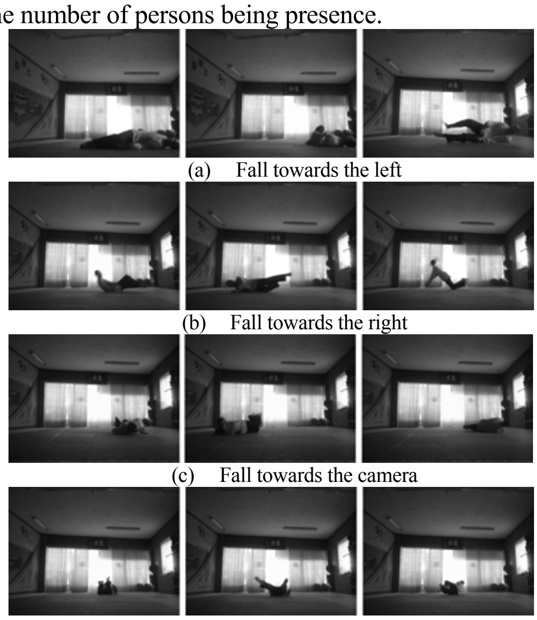
(a) (b)

Figure 2 – Examples of two persons detected.

3.1 Evaluation in the Martial Arts School in Chania

3.1.1 Background Changes

To make things more complicated, video background was changing dynamically. New objects were appeared in the scene and existing objects changed position. Below are shown three pictures of background. Figure 1 shows some characteristics examples of background changes. In Figure 3 (a) the curtains are closed while in Figure 3 (b) the curtains are opened and in Figure 3 (c) curtains are opened and one bench was appeared in the scene. We impose no restrictions on the number of persons being presence.



(a) Fall towards the left

(b) Fall towards the right

(c) Fall towards the camera

Figure 3 – Examples of different falls.

3.1.2 Humans' Actions

Experimental actions include several actions such as (a) falls, (b) appearance/disappearance of objects, and (c) normal activities. Falls were made in every direction according to the camera. This includes falls to the right, to the left, with forward motion and backward motion in regard to the camera position. Some examples of the falls are presented in Figure 3. During experimentation process, several objects were used, such as benches and balls to simulate normal activities, like sitting or playing with the ball, and falls, like falling from the bench. Normal activities simulated during the experiment, included leaning forward to tie the shoelace, laying down on the floor, sitting on the bench, sitting down on the floor. These normal activities may look like a fall, but they are not a real fall, so they used to check false negative rate and consequently the performance of the system. Examples of normal activities are shown in Figure 4.

Figure 4 – Examples of different normal humans' activities tested.

The motion vectors calculated in our system is depicted in Figure 5.

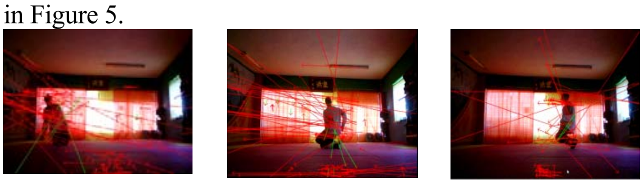
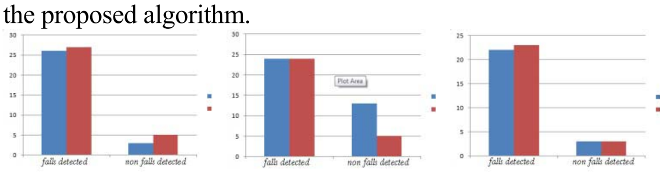


Figure 5 – Motion vectors for some human's activities/falls.

As described before, two different versions of the algorithm were tested. With the first version only the velocity of the vertical motion is checked to establish a fall [modification of equation (1) without the ratio $v(k)/v(k-m)$]. While in the second case the system is tracking the motion of a point which is a result of Eq. (1). False positives and negative are presented in Figure 6. This figure verifies the success of the proposed algorithm.



(a) (b) (c)

Figure 6 - Overall Performance: false positive/false negative. (a) vertical speed of 85 and the ratio of Equation (1) to 0.95. (b) vertical speed of 100 using no ratio in (1). (c) vertical speed of 100 and ratio of equation (1) to 1.

Figure 7 shows the action how false positives and negatives behave with respect of different values of vertical velocity and ratio $v(k)/v(k-m)$. It is clear from Figure 7 that performance depends on the number of fall detections and the number of non fall detections, however increasing the number of fall detections and decreasing the number of non fall detections are two competitive goals.

3.2 Demo Room in Trikala Municipality

Figure 8 shows tow frames of the video being captured in Trikala demon room. The background changes with the position of a chair.

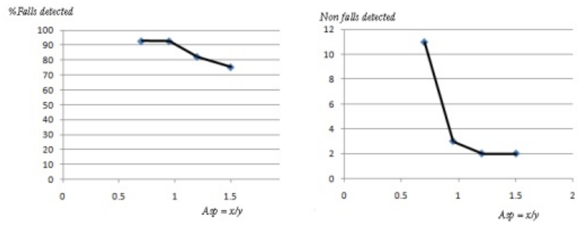


Figure 7- Average performance of the fall and non-fall detection using the equation (1)

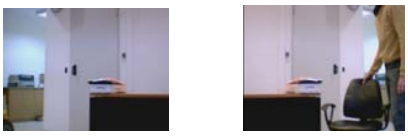
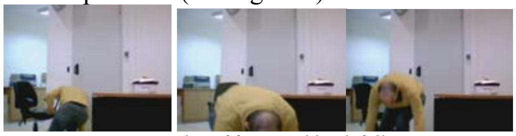
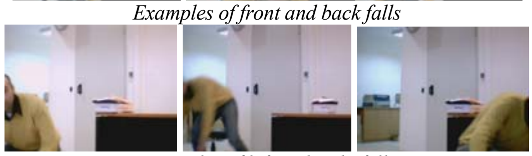


Figure 8 – Background changes in Trikala demo room.

Except the video background structure, luminosity, as well, was remaining constant due to the fact that the room is an indoor place. In this way, experiment environment was much simpler than the first case. During the experimentation process one person simulated falls, in every direction according to the camera position, and every day normal activities. Again, simulated falls were made in every direction according to the camera position. This includes falls to the right, to the left, with forward motion and backward motion in regard to the camera position. (see Figure 9).



Examples of front and back falls



Examples of left and right falls

Figure 9 **-** Examples of different falls.

Normal activities simulated during the experiment. These may look like a fall, but they're not a real fall. By using them we can check the performance of the system. Examples of normal activities are shown in Figure 10. The motion vectors for fall and non fall motions are also depicted in Figure 11.

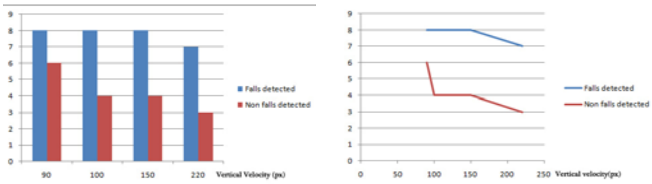
Figure 10 - Examples of normal activities.

Figure 12 describes the overall performance, for the first version of the algorithm, i.e., without the use of ratio $v(k)/v(k-m)$. Figure 12(a) shows the number of false positive and false negative while Figure 12(b) how the results are affected by the vertical velocity parameter. For Figure 12(a), the vertical velocity was set up to 150pixels. Finally, Figure 13 shows the results with respect to the ratio $v(k)/v(k-m)$.

(a) Fall motion vectors

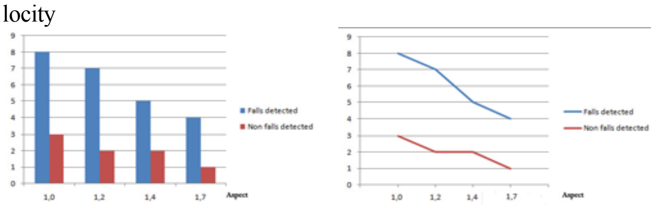
(b) non fall motion vectors

Figure 11 - Motion vectors for fall and non fall movements.



(a) (b)

Figure 12 - Fall detection performance with respect to vertical velocity



(a) (b)

Figure 13 - Fall detection performance with respect to the ratio $v(k)/v(k-m)$

4. CONCLUSION

This paper presents a new real-time computer vision algorithm able to identify falls in dynamic real-world environments via visual observations. The algorithm requires low cost and memory requirements and thus it can be applied in large scale. This paper presents an extensive evaluation of the algorithm into two different real-world environments, the martial arts of Chania and a demo room in Trikala municipality.

REFERENCES

[1]. ISISEMD European Union Funded project, http://ec.europa.eu/information_society/apps/projects/factsheet/index.cfm?project_ref=238914.

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

[3]. Shuangquan Wang, Jie Yang, Ningjiang Chen, Xin Chen, Qinfeng Zhang, "Human Activity Recognition with User-free Accelerometers in the Sensor Networks.," *Inter. Conf. on Neural Networks and Brain,* pp. 1212-1217, 2005.

[4]. T. Le, R. Pan, "Accelerometer-based Sensor Network for Fall Detection," *IEEE Conf. on Biomedical Circuits and Systems*, pp. 265 –268, 2009

[5]. Chin-Feng Lai, Sung-Yen Chang, Han-Chieh Chao, Yueh-Min Huang, "Detection of Cognitive Injured Body Region Using Multiple Triaxial Accelerometers for Elderly Falling," IEEE Sensors Journal, Vol. 11, No. 3, pp. 763 – 770, 2011.

[6]. Y. Zigel, D. Litvak, I. Gannot, "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*, Vol. 56, No. 12, pp. 2858 – 2867, 2009

[7]. F. Bianchi, S.J. Redmond, M.R. Narayanan, S. Cerutti, and N.H. Lovell, "Barometric Pressure and Triaxial Accelerometry-Based Falls Event Detection," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 18, No. 6, pp. 619 – 627, 2010.

[8]. N.M Nyan, F. Tay, E. Murugasu, "A Wearable System for Pre-impact Fall Detection," *J. Biomechanics*, Vol, 41, No. 16, 3475—3481, 2008.

[9]. N.M. Nyan, F. Tay, A.W. Tan, K.H. Seah, "Distinguishing Fall Activities from Normal Activities by Angular Rate Characteistics and High-speed Camera Characterization," *Medical Engineering & Physics*, Vol. 28, No. 8, pp. 842—849, 2006.

[10]. A.K. Bourke, K.J. O' Donovan,, G. Ó Laighin, "The Identification of Vertical Velocity Profiles using an Inertial Sensor to Investigate Pre-impact Detection of Falls," *Medical Engineering & Physics*, Vol. 30, No. 7, pp. 937—946, 2008.

[11]. Z. Fu, T. Delbruck, P. Lichsteiner, E. Culurciello, "An Address-Event Fall Detector for Assisted Living Applications," *IEEE Transactions on Biomedical Circuits and Systems TBCAS*, , Vol. 2, No. 2, pp. 88-96, June 2008.

[12]. N. Thome, S. Miguet, S. Ambellouis, "A Real-Time Multiview Fall Detection System: A LHMM-Based Approach," *IEEE Trans. on CSVT*, Vol. 18, No. 11, pp. 1522—1532, 2008.

[13]. T. Liao, Chung-Lin Huang, "Slip and Fall Events Detection by Analyzing the Integrated Spatiotemporal Energy Map," 20th International Conference on Pattern Recognition (ICPR), 2010, pp. 1718 – 1721. 2010

[14]. G. Diraco, A. Leone, P. Siciliano, "An active vision system for fall detection and posture recognition in elderly healthcare," *Design, Automation & Test in Europe Conference & Exhibition* (DATE), pp. 1536 – 1541, 2001

[15]. N. Doulamis, "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

[16]. N. Doulamis, "Visual Fall Alert Service in low Computational Power Device to Assist Persons' with Dementia," IEEE 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, Rome, November 2010.

[17]. A. Doulamis, N. Doulamis, I. Kalisperakis and C. Stentoumis, "A Real-time Single-Camera Approach for Automatic Fall Detection," *ISPRS Commission V, Close Range Image measurements Techniques*, Newcastle upon Tyne, 22-24 June 2010.

[18]. A. Doulamis I. Kalisperakis,Ch. Stentoumis and N. Matsatsinis, "Self Adaptive Background Modeling for Identifying Persons' Falls," *IEEE Conference on Semantic and Digital Media Technologies*, Limassol, Cyprus, 2010.

[19]. J. Shi and C. Tomasi, "Good Features to Track," *IEEE Inter. Conference on Computer Vision and Pattern Recognition,* pp. 593-600, 1994.