CAMERA BASED FALL DETECTION USING MULTIPLE FEATURES VALIDATED WITH REAL LIFE VIDEO

Glen DEBARD[a,e] [1], Peter KARSMAKERS[a,f], Mieke DESCHODT[b,d], Ellen VLAEYEN[b,d], Jonas VAN DEN BERGH[c], Eddy DEJAEGER[b], Koen MILISEN[b,d], Toon GOEDEMÉ[c,e], Tinne TUYTELAARS[e], Bart VANRUMSTE[a,f]

[a]*MOBILAB: Biosciences and Technology Department, K.H.Kempen, Belgium*

[b]*Center for Health Services and Nursing Research, K.U.Leuven, Belgium*

[c]*Lessius, Campus De Nayer, Belgium*

[d]*Department of Internal Medicine, Division of Geriatric Medicine, University Hospitals Leuven, Belgium*

[e]*ESAT-PSI, K.U.Leuven, Belgium*

[f]*ESAT-SCD, K.U.Leuven, Belgium*

Abstract. More than thirty percent of persons over 65 years fall at least once a year and are often not able to get up again unaided. The lack of timely aid can lead to severe complications such as dehydration, pressure ulcers and death. A camera based fall detection system can provide a solution. In this paper we compare four different fall features extracted from the dominant foreground object, as well as various combinations thereof. All tests are executed using real life data, which has been recorded at the home of 4 elderly, containing 24 falls. Experiments indicate that a fall detector based on a combination of aspect ratio, head speed and fall angle is preferred. The preliminary detector, which still has a substantial false alarm rate with a precision of 0.257(±0.073) and a promising recall of 0.896(±0.194), gives insights for further improvement as is discussed.

Keywords. Fall Detection, Video Surveillance, Assisted Living

Introduction

Many older persons fall and are not able to get up again unaided. Thirty to forty-five percent of the persons aged 65 or older living at home and more than half of the elders living in a nursing home fall at least once a year. One out of three up to one out of two older persons fall more than once every year [1]. Ten to fifteen percent of those who

[1] Corresponding Author: Glen Debard, Mobilab K.H.Kempen, Kleinhoefstraat 4, 2440 Geel, Belgium; E-mail: glen.debard@khk.be.

fall, suffer severe injuries. The lack of timely aid can even lead to more severe complications (e.g. dehydration, pressure ulcers and even death). Although not all falls lead to physical injuries, psychological consequences are equally important, leading to fear of falling, losing self-confidence and fear of losing independence [1]. Taking the ongoing aging of the population into account, it is obvious that a manner to detect fall incidents is getting more and more important.

The existing technological detectors are mostly based on wearable sensors. However a market study of SeniorWatch [2] discovered that the sensors are not worn at all times (e.g. at night). Also, in case the device is button operated, like a Personal Alarm System, the person often is not able to activate the alarm system due to complexity of issues around the use of call alarms [3]. A camera based system can overcome these disadvantages.

In the last decade, several research groups have focused on a camera based fall detection algorithm. A simple approach is to inspect the aspect ratio of the bounding box of the moving object as used in [4]. Notice however that the aspect ratio can also be altered due to occlusions. This could induce false positives. Willems et al. [5] use a combination of this feature and the fall angle. Lee and Mihailidis [6] detect a fall by analyzing the shape and 2D velocity of the person. Rougier et al. [7] use wall-mounted cameras to cover large areas and falls are detected using motion history images and human shape variation. Other systems use the 3D trajectory and the speed of the head to infer events [8]. Here, we compare and combine several of these features.

A major drawback of all these studies, is the fact that they use simulated data. The falls have been recorded in artificial environments and the simulators are mostly younger persons. The goal of our work is the development and evaluation of a prototype of a camera based fall detection system using real life data. For this, we have installed cameras to monitor the falls of four older persons at their home for 6 months.

This paper gives an overview of our fall detection algorithm and the preliminary results of the validation of the algorithm using exclusively our real life video. Section 1 describes the proposed fall detector, while Section 2 shows the results and discusses the main challenges. In Section 3 we discuss these results. Section 4 concludes the paper.

1. Methods

Our fall detection algorithm consists of four main parts: video acquisition, person tracking, fall detection and alarm generation (see Figure 1). The video is first converted to grey level images. This way there is no need to alter the processing if we switch to near-infra red at night. The alarm generation is not implemented at this stage. The next sections explain the subjects and the data set, person tracking, features for fall detection and fall detector in further detail.

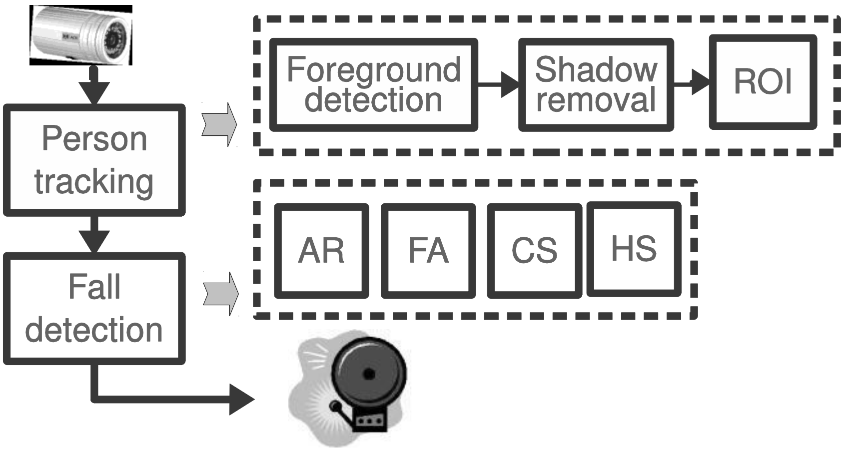

Figure 1: Overview of the system (ROI: region of interest detection; aspect ratio (AR), Fall angle (FA), center speed (CS) and head speed (HS): different features, see text)

1.1. Subjects and dataset

During the acquisition phase, we have installed four camera systems consisting of 4 wall-mounted IP camera's for 6 months. Three in one nursing home and one at the home of an elderly woman. The age of the participants is in the range of 83 to 95 years old. During these 6 months, we have captured 24 falls and recorded 14000 hours of video. To our knowledge, this is a unique dataset. To capture these events, we received the approval of the Medical Ethics Committee of the Leuven University Hospitals and all participants gave their written informed consent. The resolution of the images is 640 by 480 pixels. To validate the algorithm, we use for each of the 24 falls, the camera on which the person is best visible. From this video, we select a fragment of 20 minutes long with the fall occurring in the last 2 minutes of the video. Our current system does not use the post-fall information (i.e., the person lying on the floor). Each video is divided in non-overlapping timeslots of 2 minutes long. For each timeslot, the fall features are extracted and the maximum values during that timeslot are used for further analysis. In total this results in 240 epochs, of which 24 are labeled as a fall.

1.2. Person Tracking

1.2.1. Foreground Detection

We first need to segment out the foreground. For this we use a background subtraction technique based on an approximate median filter [9]. The technique uses one background estimate. This estimate is compared pixel by pixel with the current frame and is updated as follows. In case the pixel in the current frame is brighter (resp. darker) than the one in the background, the background intensity level is incremented (resp. decremented) with one. The foreground can then be determined by calculating the difference between the current frame and the background. In case it is larger than a given threshold (in our case 10 out of 256 intensity levels), the pixel is considered as foreground. Otherwise, it is a background pixel. The advantages of the approximate median filter are its low memory consumption, fast computation and robustness. The drawbacks are its rather slow update to large changes in illumination and the fact that the foreground is influencing the background from the moment that it appears. This influence leads to the appearance of a ghost figure. When a person is sitting on the couch for a longer period, the background is updated to incorporate the person into the background. If he stands up, the region of the couch that was occluded previously will also differ from the background and it is detected as foreground. This can influence the extraction of the features to detect a fall.

1.2.2. Shadow Removal

A shadow cast by a moving object is also detected as foreground since it makes the covered pixels appear darker. This makes the foreground erroneous. To remove this shadow, we use the property that a shadow only changes the intensity of the pixel while the texture of the covered region does not change [10]. As a result, the texture of the shadow is correlated with the corresponding texture of the background image. Jacques and Jung describe in [11] the usage of the cross correlation (CC) to see how good the detected foreground pixels match the background pixels. In case the CC is higher than a certain threshold and the pixel is darker in the current image, then the pixel is classified as shadow. Also other changes in illumination can be eliminated using this technique when removing the constraint that the pixel has to be darker in the current image. Jacques and Jung state that a threshold for the CC of 0.98 together with a 5×5 neighborhood gives a good result. These values are also used in our experiments.

1.2.3. ROI Detection

The next step in our algorithm is the determination of a region of interest (ROI). We first use an erosion/dilation step on all foreground pixels. Followed by a connected components analysis to determine the foreground objects. The largest object in the foreground is selected and considered to correspond to the person.

1.3. Fall Detection Features

Using the person, we extract four features to detect a fall, including: aspect ratio (AR), Fall angle (FA), center speed (CS) and head speed (HS) (see Figure 2).

Figure 2: Extraction of fall features: full line: bounding box, fine dash line: bounding ellipse, diamond: center of gravity, octagon: head position (The black box is for privacy reasons)

The AR is calculated as the ratio of the width of the bounding box (BB) around the foreground object and its height. A low AR represents an upright person, while a high AR might point to a person lying down. The angle of the person in the image can be defined as the angle between the long axis of the bounding ellipse (BE) and the horizontal direction. A person that is standing, has an angle of close to 90 degrees. A small angle represents a person lying down (if seen from a side-view). The FA is the change in angle over a fixed timespan (2 seconds in our experiments). A FA close to 90 degrees can indicate a fall.

A person, and certainly an older person, typically moves with a low speed. In contrast, most of the falls have a portion with high speed movement. Based on this observation, we use two fall features related to speed, CS and HS. CS is the speed of movement of the center of gravity (CG). This CG has the advantage that it is rather stable. Small changes in appearance of the person give only small changes in the CG. But an occlusion of the lower body, which happens frequently, causes the CG to move upwards. The head, on the other hand, is visible in most cases. In [12] Foroughi et al. define the head as the highest point of the object. Here we use the highest end of the main axis of the BE as head position. The speed itself is then defined as the amount of pixels that the point has shifted between two adjacent frames in the video divided by the time between these two frames.

1.3.1. Fall Detection with SVM

In this section we propose a Support Vector Machine (SVM) [13] based fall detector which classifies a timeslot (by its features) either as a “normal” or as an “abnormal” event, i.e. a fall.

When observing the data it can be noticed that classes are imbalanced (in most cases “normal” behavior is seen, falls are rare) and class distributions are overlapping (the limited set of features being used might not clearly discriminate all “normal” events from falls). Without any precautions SVM prediction might result in a simple majority vote ignoring the existence of falls. To address this problem the SVM learning objective was modified such that different weights are applied to misclassifications depending on the class [14]. In the SVM learning objective errors for the minority class are multiplied by w while majority errors are multiplied by $1 - w$. In which manner w is determined is explained later.

In order to validate the fall detector the available data set is randomly partitioned into a training set, containing 66% of the data, and an independent test set with the remaining data. The training set is then used to estimate the SVM model parameters and a set of hyper-parameters. The test set is only used for validation.

The hyper-parameters used in this paper are (a) the weight *w*, (b) a regularization parameter and (c) the Radial Basis Function (RBF) kernel bandwidth. These are selected using cross-validation and a grid search maximizing the Area Under the Curve (AUC) of a Receiver Operating Characteristic (ROC) curve. The ROC curve is computed by varying the threshold on the distances of considered data examples to the separating hyperplane which is defined by the SVM model. In order to reduce random effects induced by partitioning the data averaged AUC scores are computed on different data partitionings.

Additionally, feature selection is performed by executing a greedy forward search. Firstly, 4 univariate SVM models (each based on 1 different feature and trained using the procedure explained above) are compared in terms of AUC. Next, the best feature (corresponding to the best SVM model) is retained and combined with each of the remaining features in a bivariate SVM model. The best feature set is retained and the procedure is repeated to find the best feature set with incremented cardinality. Note that features were standardized to have zero mean and unit standard deviation.

2. Results

Given our 4 features, SVM models were estimated using the procedure described in the previous section. Results were averaged over 10 different partitionings of training and test set.[2] Table 1 lists the averaged AUC scores and the corresponding standard deviations for SVM models based on different feature sets. Individual AUC scores are computed on the independent test set. Figure 3 and Figure 4 respectively present the ROC and Precision Recall curves of the 4 best performing, in terms of AUC, SVM models. It can be observed that the combination of AR, HS and FA is preferred. Using this feature set SVM outputs an averaged operating point with a recall of 0.896(±0.194) and a precision of 0.257(±0.073).

Table 1. Results from SVM

Feature set	AUC
{AR}	0.882(±0.058)
{FA}	0.528(±0.092)
{CS}	0.843(±0.046)
{HS}	0.872(±0.052)
{AR,HS}	0.872(±0.052)
{AR,HS,FA}	0.909(±0.047)
{AR,HS,FA,CS}	0.864(±0.062)

[2] Note that for each feature set the same set of data partitionings was used.

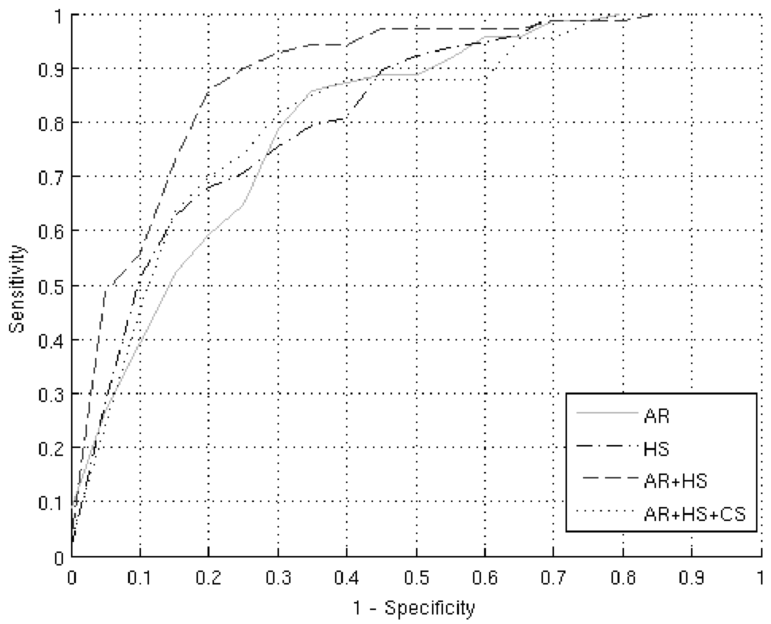

Figure 3: ROC Curve

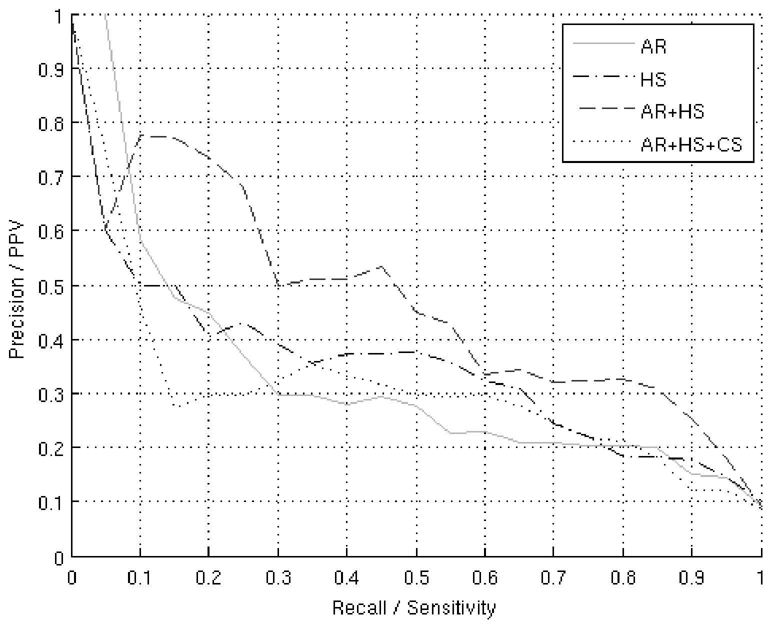

Figure 4: Precision Recall graph

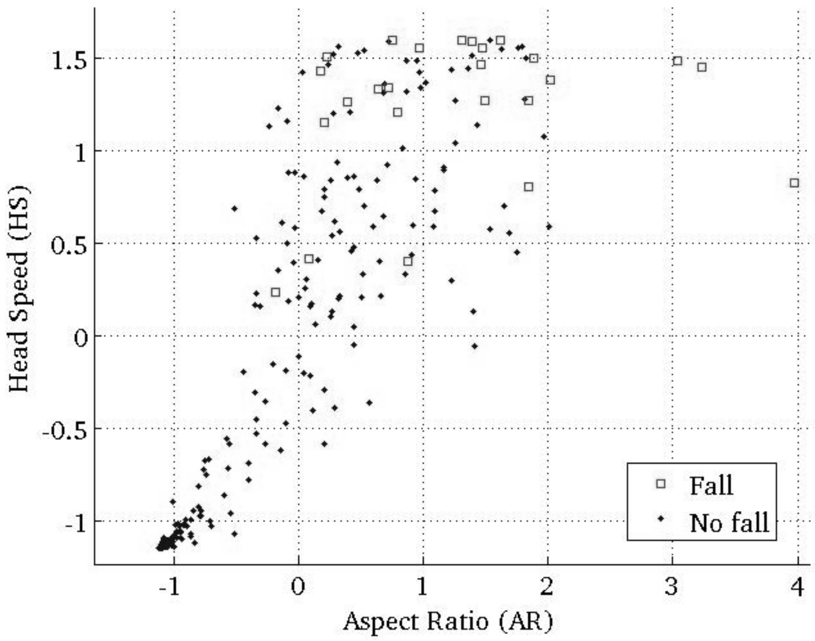

Figure 5: Class distribution using AR and HS

Considering Figure 3, we notice that the precision quickly drops when increasing the recall. This behavior can be explained by Figure 5 presenting the distribution of the data when considering features AR and HS. Here we can see that there are quite a number of non-falls that are close to the fall-cluster. Closer visual inspection reveals that 90% of these have 4 main causes. In 25% of the cases 2 persons are present in the room. In 20% of the cases another foreground object has almost the same size as the person. In both cases, the system often switches to the other person or object, resulting in large motions and changes in AR. In 25% of the cases, the person's image is split in 2 blobs which are almost the same size. Situations were such an event occurs include: over-illumination, person is wearing a shirt that is similar to the background or starts to be integrated in the background by the background update. This often results in a deviating AR as well as large motions as the system jumps back and forth between the different parts. Finally, in 20% of the cases there is interference of a ghost figure or moved furniture.

In some situations segmentation of falls is difficult. Elderly often use a walking aid, which can roll further in case of a fall. Additionally rooms are often full of furniture, which causes the person to fall against or even displace the furniture. This often has a strong influence on the fall detection features.

At the same time, many of the falls are atypical, e.g. falling while getting out of the sofa or falling when kneeling to pick something up. These atypical falls increase the within-class variability, yet are underrepresented in simulated datasets.

3. Discussion

Comparing our results with those reported in the literature [4][8], we have a similar to even higher detection rate, but a higher false alarm rate. The higher false alarm rate can be explained by the challenging nature or our data set, including various sources of errors that were previously largely ignored. In real life, falls only occur in rare cases. It is thus important to significantly decrease the number of false alarms to get a usable fall detection system.

Most of the false alarms can be solved by using more advanced techniques. The largest improvement can be introduced by using a tracker. This avoids large motions and changes in appearance caused by jumping back and forth between different foreground blobs of different (parts of) persons or other objects. Also a more advanced foreground detection, that is robust to continuous changes in illumination, slow movement of older persons, different types of light sources and possible over-illumination, can give a large improvement. Using Mixture of Gaussians however showed no improvement on first sight. A means to detect a person in the foreground, like for example the person detector of Felzenswalb et al. [15] can also reduce erroneous foreground objects. This detector is only trained for standing persons, but can still help to select the interesting foreground object.

Additional improvements are possible using other fall features (e.g. posture or other appearance-based approaches), integrating information of several cameras and certainly by integrating the post fall information.

4. Conclusion

Fall detection is becoming more and more important to ease the fears of an older person or someone with an increased fall risk. In this way these persons are able to live longer independently in a more comfortable way. In this paper we gave an overview of our ongoing research, which is unique in the way we use real life data. We have shown that under real life conditions, various sources of errors emerge such as other persons, moving furniture, walking aids, etc. that significantly increase the number of false alarms, yet have previously been largely ignored. Our preliminary fall detector shows a promising recall of 0.896(±0.194) and a precision of 0.257(±0.073). This calls for further research into more discriminative fall features, as well as better foreground detection algorithms, including tracking and person detection.

References

[1] K. Milisen, E. Detroch, K. Bellens, T. Braes, K. Dierickx, W. Smeulders, S. Teughels, E. Dejaeger, S. Boonen, and W. Pelemans, *Falls among community-dwelling elderly: a pilot study of prevalence, circumstances and consequences in Flanders*, Tijdschr Gerontol Geriatr, vol. 35, no. 1, pp. 15-20,2004.

[2] SeniorWatch, *Fall detector: Case study of European 1st seniorwatch project,* Tech. Rep., SeniorWatch, 2001.

[3] Fleming J. and Brayne C., *Inability to get up after falling,subsequent time on floor, and summoning help: prospective cohort study in people over 90*, BMJ, vol. 337, no. v17 1, pp. a2227, 2008.

[4] S. G. Miaou, P.H. Sung, and C.Y. Huang, *A customized human fall detection system using omni-camera images and personal information*, Distributed Diagnosis and Home Health-care, pp. 39-42, 2006.

[5] J. Willems, G. Debard, B. Vanrumste, and T. Goedemé, *A video-based algorithm for elderly fall detection*, Medical Physics and Biomedical Engineering World Congress, WC2009, Sept. 2009.

[6] T. Lee and A. Mihailidis, *An intelligent emergency response system: preliminary development and testing of automated fall detection*, Journal of Telemedicine and Telecare, vol. 11, no. 4, pp. 194-198, 2005.

[7] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, *Fall detection from human shape and motion history using video surveillance*, in Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st International Conference on, 2007, vol. 2, pp. 875-880.

[8] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, *Monocular 3d head tracking to detect falls of elderly people*,2006.

[9] N. J. B. McFarlane and C. P. Schofield, *Segmentation and tracking of piglets in images*, Machine Vision and Applications, vol. 8, no. 3, pp. 187-193, May 1995.

[10] D. Grest, J. Frahm, and R. Koch, *A color similarity measure for robust shadow removal in real-time*, Vision, Modeling and Vizualization, 2003.

[11] J. C. S. Jacques and C. R. Jung, *Background subtraction and shadow detection in grayscale video sequences*, The XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI05), 2005.

[12] H. Foroughi, B.S. Aski, and H. Pourreza, *Intelligent video surveillance for monitoring fall detection of elderly in home environments*, in Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on, 2008, pp. 219-224.

[13] V.N. Vapnik, Statistical learning theory, Wiley, New York, 1998.

[14] E. Osuna, R. Freund, and F. Girosi, *Support vector machines: Training and applications*, AI Memo 1602, Massachusetts Institute of Technology, 1997.

[15] P. Felzenszwalb, D. Mcallester, and D. Ramanan, *A discriminatively trained, multiscale, deformable part model*, in IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Anchorage, Alaska, June 2008.