CAMERA BASED FALL DETECTION USING REAL-LIFE VIDEO

G. Debard[1], *M. Deschodt*[2], *E. Vlaeyen*[2], *J. Van den Bergh*[3], *E. Dejaeger*[4], *K. Milisen*[2], *T. Goedemé*[3], *T. Tuytela*

[1]MOBILAB: Biosciences and Technology Department, K.H.Kempen, Belgium
[2]Center for Health Services and Nursing Research, K.U.Leuven, Belgium
[3]Lessius, Campus De Nayer, Belgium
[4]U.Z.Leuven, Belgium
[5]ESAT-PSI, K.U.Leuven, Belgium
[6]ESAT-SCD, K.U.Leuven, Belgium

ABSTRACT

More than thirty percent of older persons 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 present our fall detection algorithm by first explaining the object detection and afterwards the used fall features. We show the performance of the fall features separately and then also the performance of the different combinations using a support vector machine. Both are measured using our real-life dataset. We conclude that the usage of the aspect ratio of the bounding box combined with the speed of the head gives the best result with a sensitivity of 0.778 and a PPV of 0.366.

***Index Terms*—** Fall Detection, Video Surveillance, Support Vector Machine, Assisted Living

1. 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] [2]. Ten to fifteen percent of those who 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] [3]. Taking the ongoing ageing 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 [4] discovered that the sensors are not worn at all times. Also in case the device is button operated, like a Personal Alarm System (PAS), the person often forgets to press the button to generate the alarm. A fall occurring at those moments will not be detected. A camera based system can overcome these disadvantages.

In the last decade, several research groups have focussed on a camera based fall detection algorithm. A simple method is to base on the analysis of the moving objects bounding box's aspect ratio as used in [5] and [6]. Notice however that the aspect ratio can also be altered due to occlusions, and the relative position of the person with respect to the camera. This could induce false positives. Willems *et al* [7] use a combination of the previous techniques combined with the fall angle. Lee [8] detects a fall by analysing the shape and 2D velocity of the person. Rougier [9] uses wall-mounted cameras to cover large areas and falls are detected using motion history image and human shape variation. Other systems use 3D trajectory and the speed of the head to infer events [10].

A mayor 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 TETRA-funded [**?**] FallCam project uses another approach. The goal of the project is the development of a prototype of a camera based fall detection system. This system should be verified using real live data. For this, we have installed cameras to monitor the falls of four older persons at their home for 6 months. The age of the participants is in the range of 83 to 95 years old. During this period of time, we have captured 25 falls and recorded 14000 hours of video. To our knowledge, this is a unique dataset.

This paper gives an overview of our fall detection algorithm and the preliminary results of the validation of the algorithm using exclusively our real-live video.

The next section 2 describes our used methods, while section 3 shows the results. In section 4 we discuss these results to conclude the paper in section 5.

2. METHODS

Our fall detection algorithm consists of four main parts: video acquisition, person tracking, fall detection and alarm generation. Fig. 1 shows this in a schematic overview. The video is 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 for the moment. The next sub-chapters will explain the person tracking and the features for fall detection in further detail.

2.1. Person Tracking

2.1.1. Foreground Detection

We first need to know the background, for this we use a background subtraction technique based on an approximate median filter. The algorithm was developed in 1995 by McFarlane and Schofield to track piglets [11]. The technique uses one background estimate. This estimate will be compared pixel by pixel with the current frame and is updated as follows. In case the pixel in the current frame is brighter

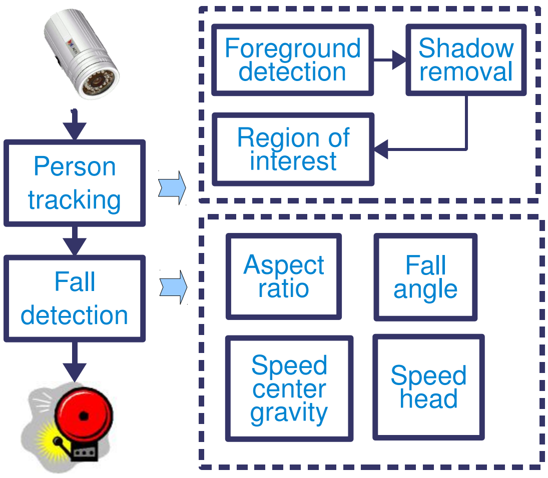

Fig. 1. Overview of Fall Detection Algorithm

than the one in the background, the background will be incremented with one, and vice versa.

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 a foreground pixel. 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 slower 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 will be updated to incorporate the person into the background. If he stands up, the region of the couch that was obstructed 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.

2.1.2. *Shadow Removal*

A shadow cast by a moving object will also be detected as foreground since it makes the covered pixels appear darker. This makes our foreground erroneous. To remove this shadow, we can use the property that a shadow only changes the intensity of the pixel, the texture of the covered region does not change [12]. From this follows that the texture of the shadow is correlated with the corresponding texture of the background image.

Jacques *et al.* describe in [13] the usage of the cross correlation (CC) to see how good the detected foreground pixels match the background pixels. For a pixel $I(i,j)$, they propose to calculate the CC as follows:

$$CC(i,j) = \frac{ER(i,j)}{E_B(i,j).E_{T_{ij}}} \qquad (1)$$

$$ER(i,j) = \sum_{n=-N}^{N} \sum_{m=-N}^{N} B(i+n,j+m).T_{ij}(n,m) \qquad (2)$$

$$E_B(i,j) = \sqrt{\sum_{n=-N}^{N} \sum_{m=-N}^{N} B(i+n,j+m)^2} \qquad (3)$$

$$E_{T_{ij}} = \sqrt{\sum_{n=-N}^{N} \sum_{m=-N}^{N} T_{ij}(n,m)^2} \qquad (4)$$

Where T is the current frame, B is the calculated background, $T_{ij}(n,m)$ is a $(2N+1)*(2N+1)$ neighbourhood around pixel $I(i,j)$, E_T is the energy in this neighbourhood and $E_B(i,j)$ is the energy in the corresponding neighbourhood of the background image. A value of the CC close to one defines a high correlation. In case the calculated CC is higher than a certain threshold L_{cc} and the pixel is darker in the current image, then the pixel is part of a shadow. Also other changes in illumination can be eliminated using this technique if removing the constraint that the pixel has to be darker in the current image.

$$CC(i,j) \geq L_{cc} \cap E_{T_{ij}} < E_B(i,j) \qquad (5)$$

Jacques *et al.* state that a threshold for the CC of 0.98 together with a 5*5 neighbourhood (N=2) gives a good result, which we agree on.

2.1.3. *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. After this a connected components analysis is executed to determine the foreground objects. The largest object in the foreground is then defined as being the person. To minimize noise and interference, the object has to be larger than a certain threshold. In our case, minimum 17500 pixels gave the best performance. From this object we can start to extract the features to detect a fall.

2.2. Fall Detection Features

Using the object that was extracted in the first part, we can extract four features to be used to detect a fall. These are: aspect ratio (AR) of the bounding box, fall angle, speed of movement of the centre of gravity and the speed of movement of the head.

The aspect ratio of the bounding box is calculated as the division of the width of the bounding box and its height. A low aspect ratio represents an upright person, while a high aspect ratio points 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 and the horizontal plane of the image. A person that is standing, has an angle of 90 degrees. While an angle of close to zero represents a person lying down if seen from a side-view. The fall angle is the difference between this angle in the current frame and the angle from a defined time in the past. We use 2 seconds between the frames to measure the angles. A fall angle that is close to 90 degrees can be a fall.

A person, and certainly an older person, moves typically with a low speed, while in contrast to this most of the falls will have a portion with a high speed of movement. We calculate two different speeds. The speed of the centre of gravity and the so called head speed. The centre of gravity has the advantage that it is rather stable. Small changes in appearance of the person gives only small changes in the centre of gravity. But an occlusion of the lower body, which happens frequently, causes the centre of gravity to move upwards. The head is however visible in most cases. Foroughi *et al.* describe in [14] that the head is the highest point of the object. We use the highest end of the main axis of the ellipse as head position. The speed itself is then defined as the amount of pixels that the point has shifted between two adjacent frames divided by the time between these two frames.

2.3. Data Set

During the acquisitioning phase of the project, we have recorded 25 real live falls. The resolution of the images is 640 by 480 pixels. For each of the 22 other falls, we have created a video of 20 minutes long. The fall occurs in the last part of the video. We did not use the information of the person lying on the ground. Each video was divided in timeslots of 2 minutes long. For each timeslot, the four features are extracted and the maximum values during that timeslot are used as feature.

On this data set, many different tests have been executed. First the different features are evaluated separately. In this case, a threshold value is changed to find the best performance for each feature. After this, we evaluated the usage of a Support Vector Machine (SVM) to automatically find the optimal combination of these different features.

2.4. Support Vector Machine

General text about SVM will be delivered by Peter. Maybe the evaluation of the SVM has to be done differently? Search for which parameters of the SVM, we get the best mean performance? We created 25 randomly chosen test and training sets. For each training set we search for the SVM that delivers the best results. For each SVM we calculated the specificity, sensitivity and positive predictive value. By calculating the mean and standard deviation (Stdev), we can show the expected performance of an SVM that is trained on the complete data set.

3. RESULTS

First the different fall detection features are handled separately. For each feature we search for a threshold that gives the best performance. The best results are shown in table 1.

Additional to looking at the features separately, we used a Support Vector Machine to find a good combination of the different features. The best results are shown in table 2.

4. DISCUSSION

Table 1 shows that the aspect ratio and both the speed of the centre of gravity and the head give similar results. The sensitivity is between 0.72 and 0.80, so 72 to 80 percent of the falls are detected. The positive predictive value is however rather low at 0.17, meaning that less than one on five is a real fall. The fall angle however generates a huge number of false detections.

From table 2, we can conclude that a combination of features gives a better result. The combination of the aspect ratio of the bounding box and the speed of the head gives the best result with a sensitivity of 0.778 and a PPV of 0.366. Using the fall angle dramatically decreases the sensitivity.

The reason for the very high number of false detections of the fall angle, can be found in the fact that the bounding ellipse can change considerably with small changes in the contours it surrounds. A person that extends one arm will already make the ellipse turn. The false detections for the three other features have two main reasons: the quality of the foreground segmentation is not yet high enough and the presence of more than one large foreground object.

As explained in section 2.1.1, we make usage of an approximate median filter. One of the disadvantages is that the person disappears in the background when immobile, which leads to a ghost figure. When the person and the ghost are connected, the bounding box and ellipse are deformed and lead to erroneous feature values. If another person enters the room or some other object is moved, we see more than one blob in the foreground. If the size is very close to each other, it is possible that the object that we use to extract our features is the older person in the first frame and a caretaker in the next frame. This causes a high speed movement. This can be dealt with by restricting the maximum speed, but a better solution is the usage of a tracker.

During a fall, often other types of interference occur. A rollator can roll further while the person is falling or the person can fall against and even move furniture. Both cases lead to a foreground that contains a lot more than only the person, and to features that are significantly different then expected. Our data set contains several of these examples. Three other falls happen while the older person is supported by a caretaker, in this case the falls are not detected, but in this case it is also not needed.

5. CONCLUSION AND FUTURE WORK

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 project, which is unique in the way that we use real-life data. We discussed our foreground detection algorithm and fall features. We presented the performance of the single features and of the combination using an SVM. We concluded that an SVM trained using the aspect ratio of the bounding box and the speed of the head of the biggest object gives the best results.

Future work will be in improving the performance of the foreground detection and using a tracker to follow more than one object.

6. REFERENCES

[1] Koen Milisen, Els Detroch, Kim Bellens, Tom Braes, Katrien Dierickx, Willy Smeulders, Stefan Teughels, Eddy Dejaeger, Steven Boonen, and Walter 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] M. E. Tinetti, "Preventing falls in elderly persons," *New England Journal of Medicine*, vol. 348, no. 1, pp. 42–49, 2003, 57 MASSACHUSETTS MEDICAL SOC/NEJM WALTHAM 630WY.

[3] P. A. Stalenhoef, J. P. M. Diederiks, J. A. Knottnerus, A. D. M. Kester, and Hfjm Crebolder, "A risk model for the prediction of recurrent falls in community-dwelling elderly: A prospective cohort study," *Journal of Clinical Epidemiology*, vol. 55, no. 11, pp. 1088–1094, 2002, 39 PERGAMON-ELSEVIER SCIENCE LTD OXFORD 628TM.

[4] SeniorWatch, "Fall detector: Case study of european ist seniorwatch project," Tech. Rep., SeniorWatch, 2001.

[5] B. U. Toreyin, Y. Dedeoglu, and A. E. Cetin, "Hmm based falling person detection using both audio and video," in *Computer Vision in Human-Computer Interaction, Proceedings*,

Table 1. Results of separate fall features

Feature	Threshold	True positive	True negative	False positive	False negative	Sensitivity	Specificity	PPV
Aspect Ratio	2.5	18	250	85	7	0.72	0.746	0.174
Angle	80	22	86	248	3	0.88	0.257	0.081
Speed of centre	180	19	247	87	6	0.76	0.739	0.179
Speed of head	640	20	241	93	5	0.80	0.722	0.177

Table 2. Results of combined fall features using a support vector machine

Fall Features	Sensitivity		Specificity		PPV	
	Mean	Stdev	Mean	Stdev	Mean	Stdev
All	0.484	0.227	0.727	0.193	0.274	0.194
All without Fall angle	0.798	0.105	0.766	0.087	0.319	0.094
Aspect Ratio + Speed of centre	0.801	0.134	0.731	0.111	0.294	0.092
Aspect Ratio + Speed of head	0.778	0.115	0.807	0.098	0.366	0.106

N. Sebe, M. S. Lew, and T. S. Huang, Eds., vol. 3766 of *Lecture Notes in Computer Science*, pp. 211–220. Springer-Verlag Berlin, Berlin, 2005, Proceedings Paper 10 HEIDELBERGER PLATZ 3, D-14197 BERLIN, GERMANY BDG83.

[6] S. G. Miaou, Pei Hsu Sung, and Chia Yuan Huang, "A customized human fall detection system using omni-camera images and personal information," *Distributed Diagnosis and Home Healthcare*, vol. 0, pp. 39–42, 2006.

[7] Jared Willems, Glen Debard, Bart Vanrumste, and Toon Goedemé, "A video-based algorithm for elderly fall detection," Medical Physics and Biomedical Engineering World Congress, WC2009, Sept. 2009.

[8] 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, 18 ROYAL SOC MEDICINE PRESS LTD LONDON 937FL.

[9] Caroline Rougier, Jean Meunier, Alain St-Arnaud, and Jacqueline 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.

[10] Caroline Rougier, Jean Meunier, Alain St-Arnaud, and Jacqueline Rousseau, "Monocular 3d head tracking to detect falls of elderly people," 2006.

[11] 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.

[12] Daniel Grest, Jan-Michael Frahm, and Reinhard Koch, "A color similarity measure for robust shadow removal in real-time," Vision, Modeling and Vizualization, 2003.

[13] 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.

[14] 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.