Camera-based fall detection on real world data

G. Debard[1,6], P. Karsmakers[1,5], M. Deschodt[2,4], E. Vlaeyen[2,4], E. Dejaeger[4]
K. Milisen[2,4], T. Goedemé[3,6], B. Vanrumste[1,5,6] and T. Tuytelaars[6,7]

[1]MOBILAB: Biosciences and Technology Department, KHKempen, Belgium
[2]Center for Health Services and Nursing Research, KU Leuven, Belgium
[3]Lessius Mechelen, Campus De Nayer, Belgium
[4]Department of Internal Medicine, Division of Geriatric Medicine, University Hospitals Leuven, Belgium
[5]ESAT-SCD, KU Leuven, Belgium
[6]ESAT-PSI, KU Leuven, Belgium
[7]IBBT Future Health department, Belgium

Abstract. Several new algorithms for camera-based fall detection have been proposed in the literature recently, with the aim to monitor older people at home so nurses or family members can be warned in case of a fall incident. However, these algorithms are evaluated almost exclusively on data captured in controlled environments, under optimal conditions (simple scenes, perfect illumination and setup of cameras), and with falls simulated by actors.

In contrast, we collected a dataset based on real life data, recorded at the place of residence of four older persons over several months. We showed that this poses a significantly harder challenge than the datasets used earlier. The image quality is typically low. Falls are rare and vary a lot both in speed and nature. We investigated the variation in environment parameters and context during the fall incidents. We found that various complicating factors, such as moving furniture or the use of walking aids, are very common yet almost unaddressed in the literature. Under such circumstances and given the large variability of the data in combination with the limited number of examples available to train the system, we posit that simple yet robust methods incorporating, where available, domain knowledge (e.g. the fact that the background is static or that a fall usually involves a downward motion) seem to be most promising. Based on these observations, we propose a new fall detection system. It is based on background subtraction and simple measures extracted from the dominant foreground object such as aspect ratio, fall angle and head speed. We discuss the results obtained, with special emphasis on particular difficulties encountered under real world circumstances.

Keywords: Fall Detection, Video Surveillance, 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 [14, 24].

Ten to fifteen percent of those who fall, suffer severe injuries. [14] The lack of timely aid can lead to further complications such as dehydration, pressure ulcers and even death. Although not all falls lead to physical injuries such as hip fracture, psychological consequences are equally important, leading to fear of falling, losing self-confidence and fear of losing independence [4, 14]. Taking the ongoing aging of the population into account, it is obvious that adequately detecting fall incidents is getting more and more important. Indeed, a large study in the Netherlands reported an increase of fall-related hospital admissions from 1981 to 2008 by 137% [8]. Furthermore, falls are associated with substantial costs. For instance, the excess costs associated with treating hip fractures range between USD 11,241-18,727 in the first year following the fracture [7]. A study in the U.K. estimated the total cost (year 1999) related to injurious falls in those aged 75 and older to be almost 647 million [20].

The existing technological detectors are mostly based on wearable sensors. However, a market study of SeniorWatch [21] discovered that the sensors are not worn at all times (e.g. at night). Also, in case the device is button operated, as with a Personal Alarm System, some persons with (mild) cognitive impairment are not always able to activate the alarm system due to complexity of issues around the use of call alarms [4]. As a result, many falls remain undetected. A camera-based system, on the other hand, has the potential to overcome the limitations mentioned above, because it is contactless and does not require initiative of the person. On the downside, one or more cameras need to be installed in every room, increasing the cost of this system; the system is fixed; and only works indoor. Another disadvantage is that it is not possible to take the system along on a trip.

In the last decade, several research groups have focused on a camera based fall detection algorithm. However a major drawback of 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 camera based fall detection system using real life data. For this, we have installed cameras monitoring four older persons with an increased risk of falling at their place of residence for six months. Three of these persons are residing in a nursing home, since people with a history of falling are often institutionalized.

In the remainder of this paper, we first discuss how we captured our dataset and the challenges posed by the usage of real world data (Section 2). Next, we give an overview of earlier work (Section 3). In Section 4, we describe the fall detection algorithm we developed, followed by some preliminary results of the validation of our algorithm using the real life video data in Section 5. In Section 6 we discuss these results. Section 7 concludes the paper.

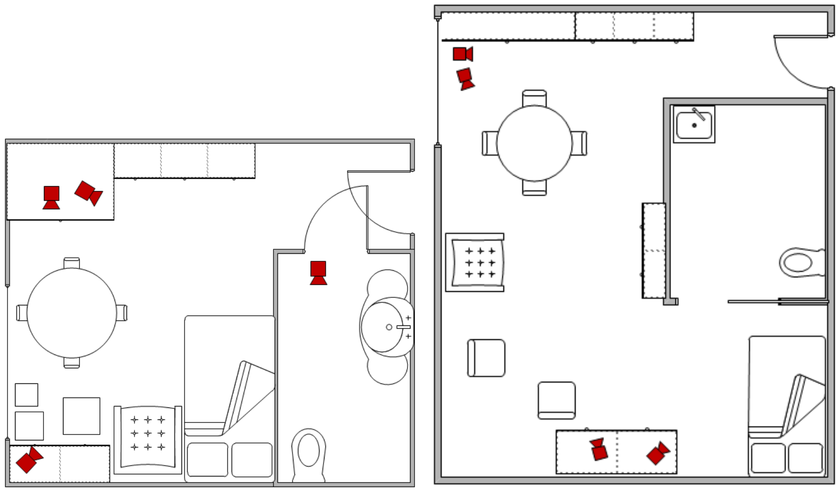

Fig. 1. Setup of cameras. Left panel: Room in nursing home. Right panel: Service flat

2 THE DATASET AND ITS CHALLENGES

2.1 Data collection

During the acquisition phase, we have installed four camera systems at the place of residence of four older persons. one at the home of an independently living older woman, one in a room of a nursing home and two in a service flat. Figure 1 shows how the cameras were installed in the nursing home. For privacy reasons, we did not plan to install a camera in the bathroom. However, the person in the nursing room asked us to install a camera there after falling twice at that location. We also provided a control panel that allowed the participant to switch off the system whenever wanted. However, only the cleaning personnel used this option.

The participants' age was in the range of 83 to 95 years old, and all of them had an increased risk of falling. Recordings were made during approximately six months, 7 days a week, and 24 hours a day. During these six months, we recorded over 14.000 hours of video and captured 24 falls. Most falls occured in two persons. The person living independently did not fall during our monitoring period, while one of the participants in a service flat only fell once. 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.

For each residence we used 4 wall-mounted IP cameras. We used a combination of ACTI ACM-1511 and AXIS 207 cameras. The ACTI cameras already had day/night vision. We changed the lenses of the AXIS cameras to one with a view angle of 80 degrees without a near-infrared filter. Additional near-infrared sources made it possible to record video in low-light conditions and during the night. We recorded images with a resolution of 640 by 480 pixels using a frame rate of 12 frames per second. Since we wanted to be able to analyze images in low light conditions or during the night, we used gray level images. To be able to store the data, we used M-JPEG compression. This reduced the disc space usage to 1.8 GB per hour.

Not only did the collection of this dataset allow us to evaluate prototype systems for camera based fall detection on real world data (see Section 5), it also provided us with valuable insights on the typical challenges that can arise when using real life data, both for fall and non-fall scenarios. While we cannot make the dataset publicly available because of privacy issues, we can comment on these general findings.

2.2 Data characteristics in a typical real life scenario

The analysis of the captured video shows some challenges that researchers developing fall detection systems should be aware of. Which ones are important depends on the algorithms used.

Image quality First, the quality of the camera in a real world scenario is typically lower than what is used in a lab setup. Indeed, from a practical point of view, to be cost-efficient, it is not possible to install high quality cameras. Moreover, it is necessary to monitor the person also in low-light conditions during the evening or night. Therefore, we also needed to record near-infrared, which is often more noisy. It is important to install as few cameras as possible. The usage of a camera and lens with view angle close to 90 degrees installed in the corner in of the room gives the best coverage. But the wide angle of the lens also decreases the spatial resolution of the camera.

Color information In near-infrared night images, no color information is available. But even during daytime when color information is available, it is not very reliable. Especially the different light sources in a house (sun light, fluorescent light, light bulbs, tv-screen, etc.) present some specific challenges. For example, during one of our preliminary tests, a person moved in front of a window, the sunlight was partially blocked, which changed the color of the incident light. Several methods for fall detection proposed in the literature [2] rely on color-based shadow detection algorithms to improve the output of a background subtraction algorithm. However, these are based on the assumption that when an area is covered by a shadow, this results in a significant change in brightness only without change in color information [6]. This assumption is not always met in real world circumstances. Hence color can be an unreliable source of information.

Overexposure The range of light intensities that occur during the day, is extensive. A good configuration of the camera is needed. Even then, the brightness of the sun can cause overexposure in some areas of the image. Careful placement of the cameras in the room can decrease the problem to some extent. Instead of pointing the camera to the window, it is better to attach it above the window,

Fig. 2. Examples of video frames with different illumination. Upper left: Sunlight causes overexposure at window. Upper right: Localized overexposure caused by halogen lamps. Lower left: Same room with minor overexposure. Lower right: Frame recorded at night using near infrared.

facing the room. However, since it is necessary to cover all areas of the room with a limited number of cameras, pointing them towards the windows cannot always be avoided. Also halogen lamps can cause overexposure, as well as special lighting conditions. Figure 2 shows an example of the same room at different moments with different kinds of illumination.

Image clutter Not only the change in illumination has to be taken into account, but also the changes occurring in the room itself. Rooms are often small, both in nursing homes as in private homes and older persons tend to collect a lot of furniture, which can have a sentimental value. When moving to a smaller residence, they want to take these along. As a consequence, rooms are often highly cluttered. When moving around in the room, the person is quite often partially occluded. Over longer time periods, furniture is also less static than one might expect (see also Figure 2). Furniture that is shifted, should therefore be dealt with appropriately by the system.

Walking aids Some older persons have difficulties walking unaided. Because of this, they sometimes use a walking aid like e.g. a rollator or a walking frame. The legs of the person and part of the lower body can be occluded by this. Moreover, the walking aid is another dominant foreground object, sometimes moving along with the person, sometimes put aside (see e.g. Figure 2 top left). Fall detection algorithms that rely on the person being the only or largest foreground object in the scene may not be able to cope with this situation.

Appearance changes The appearance of the person also changes over time, e.g. while getting (un)dressed or changing clothes. Under such conditions, relying on color or intensity distributions to track the person, may not be a good idea.

Other moving objects Other challenges are for example a television or a cupboard with doors that can be opened. Also a door is difficult to take into account. It is a large moving object, and what is behind the door can differ each time (e.g. an entrance door in a nursing home). A person that is lying in bed, is almost completely occluded by the sheets while sleeping. But getting out of bed, the sheets are folded back, which again represents a large moving object. Some methods based on motion history images (e.g. [19]) learn to ignore the motion in these image areas. However, this means that falls occurring at these locations are more likely to be ignored as well.

Motion patterns The behavior of an older person can differ significantly from that of a younger person. Analyzing our data, we observed that some persons stay seated in the same place for extended times during the complete day. The manner in which older persons move can differ significantly from younger persons, certainly with respect to the speed of movement, which can be extremely slow in some cases. Also the typical gait is different, with shorter strides.

2.3 Analysis of the observed fall incidents

As mentioned before, we monitored four persons and collected 24 falls. One person did not fall during the monitoring period, while a second person fell only once. The other two persons fell 10 and 13 times, respectively. Because the majority of the falls occurred in only two individuals, it is not possible to generalize our findings. Nevertheless, the recorded falls already give us some insight in the challenges their detection represents.

Use of walking aids Both persons with a high number of falls, often used a rollator walker. Half of the falls (n=12) occurred while using a walking aid. When the person was falling, the rollator was pushed forward, sometimes crossing a huge part of the room, or turning over. All these cases may interfere with the fall detection, either because the person is occluded behind it, or because it corrupts the extracted features. Figure 3 shows some examples of interference that a rollator walker can cause.

Initial pose Not all falls start from a standing pose. A fall can also start from a crouching or a bend over position, while picking something up. This occurs in five falls (21%). A fall can also happen in two steps. Sometimes the person was able to grab hold of a door or chair, but after a short time, had to let go and

Fig. 3. Two fall incidents with interference of a rollator walker. Upper panels: a fall where the rollator partially occludes the person. Lower panels: the rollator is pushed and rolls away from the person.

fell to the ground. This happened in two falls (9%). Five falls (21%) happened shortly after standing up or while preparing to sit down. This arises because an older person sometimes doesn't have enough strength in his/her legs to stand up or sit down slowly.

Occlusions and appearance Occlusions are another important challenge. In eleven falls (46%), the person was completely or partially occluded, either by the walking aid or by the furniture. In one case, the fall started in one room and ended in an adjacent one. Even with multiple cameras in the room, it is often impossible to get an unoccluded view of the person. In three falls (12%), the person was undressing, which drastically changed the appearance of the person.

Other moving objects One of the most occurring challenges are other moving objects in the scene. In 18 falls (75%), the furniture in the room was moved by the fall. Certainly chairs and tables are shifted easily, but also small and even larger cupboards can be moved during a fall. Moving doors are also common. In one case, a painting on the wall was shifted. The consequence is that sometimes the appearance of the room can change completely. We already mentioned that in some cases, the room is really filled with different pieces of furniture. In such a case, it is almost impossible to not hit something while falling down. Even when a

Fig. 4. Two fall incidents with moving furniture. Upper panels: The table and chairs are moved and the upper body of the person is occluded. Lower panels: The table, chairs and sofa are moved. The rollator is also fallen over and the person is almost completely occluded.

room is only modestly furnished, a fall against furniture will occur in most cases. Figure 4 shows some examples of this type of interference. Especially methods assuming a static scene and relying on background subtraction are affected by this. On the other hand, a sudden motion over a large part of the scene could by itself be a cue for fall detection.

Unbalanced data The final challenge is the ratio of fall to non-fall data. We have recorded a dataset that is really extensive. The persons that we monitored all had a high risk of falling. The numerous falls of two of our participants show this. But even in this case, the falls only represent a tiny portion of the available data. The performance of a fall detector is not only determined by its ability to detect a fall, but also by its ability to generate as few false alarms as possible. To test this, it is important to not only use the falls, but also part of the realistic non-fall data.

The usage of this real life data and the numerous challenges it represents, greatly increases the complexity in building a working fall detection system. In the following section we review the state-of-the-art, taking the challenges mentioned above into account. Next, in Section 4, we explain our preliminary fall detector in more detail.

3 RELATED WORK

Most systems described in the literature can be divided in two main approaches to the problem: those that try to detect the action of falling directly (e.g. [1, 2, 5, 12, 13, 18, 19, 22, 23, 25]), and those that instead detect unusual events in general (e.g. [15, 16, 19]). The latter rely on indirect evidence, such as prolonged inactivity at unusual locations, to infer fall incidents. Since normal behavior in terms of person appearance (actions or poses) is considered too broad and varied to model, these systems typically focus on spatio-temporal trajectories instead. By doing so, the problem of the large variability in appearance is circumvented. Moreover, since it is only needed to learn what normal behavior looks like, the unbalancedness of the data is not really an issue, nor is the variability in appearance of fall incidents. On the downside, what is normal behaviour in terms of spatio-temporal trajectories is typically location and camera (viewpoint) specific. Therefore, these systems usually need to be retrained for each new camera setup. Also, an unusual pattern does not imply the occurrence of a fall incident (or another event that would require intervention, for that matter). If, for instance, a person is ill, he/she may show various forms of unusual behaviour, such as staying in bed longer than usual, or going to the bathroom in the middle of the night. This may result in lots of false alarms.

Methods that more directly try to detect the dynamic event of falling, do not suffer from the above mentioned limitations. In this category, we again distinguish between methods building on simple cues like motion detection, often combined with domain knowledge (e.g. [1, 2, 13, 18, 19, 23]), and methods that build on recent advances in generic person detection and action recognition (e.g. [22]). While the latter may seem promising at first, the amount of training data seems insufficient to learn a reliable model for falls, especially when taking the large variability in appearance of the falls into account. Also the quality of the images is a limiting factor. Figure 5 shows the output of a state-of-the-art person detector / pose estimator [28] applied to some of our recordings. A tracker might improve these results to some extent, but we doubt whether it will be accurate enough to infer a fall from the change in pose. Finally, the needed computation time of these methods often does not allow for real-time processing.

It is possible to use more complex methods, like action recognition and person detection, but we believe the most promising approach at this moment to be a combination of relatively simple, low-level cues with available domain knowledge. Since we know the cameras are static, background subtraction can be applied to find the moving foreground objects, including the person. Likewise, one can build on domain knowledge to design simple yet robust fall features, such as the aspect ratio [1, 2, 13, 27] or the speed of the head of a person [18, 12] (exploiting the fact that the head remains mostly unoccluded). These can be combined in a low dimensional representation and presented to a classifier, with limited risk of overfitting. Background subtraction has been used by many systems (e.g. [1, 2, 5, 12, 13, 16, 23, 27]). However, in many cases, it is assumed that this results in an accurate silhouette of the person, based on which the pose can be determined (e.g. [1, 2, 5, 16]). This is usually not the case for our real life data. Due to the

Fig. 5. Examples of the output of a state-of-the-art person detector [28]. Upper left panel: Successful detection of human pose. Upper right and lower panels: Failed to detect pose of the person.(Green : head, yellow : torso, violet : left arm, light blue : right arm, red : left leg, dark blue : right leg)

low image quality as well as problems with overexposure, occlusions or changing illumination conditions, background subtraction (even after shadow removal) only gives a rough idea of where the person might be. Also the fact that older persons often stay seated at the same place over long periods of time does not help in this respect.

In conclusion, methods exploiting relatively low-level cues (e.g. [10, 18, 19, 23]) seem most promising in a real life context. They are robust, fast to compute, and relatively generic (no need for retraining or calibration for each new camera setup). More complex schemes can then be added as further verification or to corroborate the results, if applicable.

4 METHODS

Our fall detection algorithm consists of four main parts: video acquisition, person detection, fall detection and alarm generation (see Figure 6). The video is first converted to gray level images. This way there is no need to alter the processing if we switch to near-infrared at night. The alarm generation is not implemented at this stage. The next sections explain the person detection, features for fall detection and fall detector in further detail.

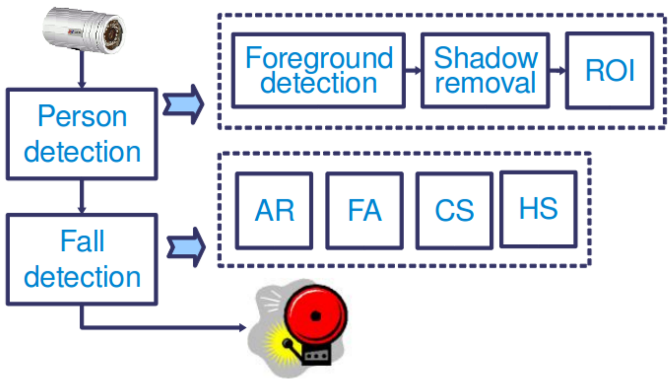

Fig. 6. Overview of the system (ROI: region of interest detection; Different fall features: Aspect Ratio (AR), Fall Angle (FA), Speed of center of gravity (CS), Head Speed (HS)

4.1 Person Detection

Foreground Detection We first needed to segment out the foreground. For this we used a background subtraction technique based on an approximate median filter [11]. 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, as any background subtraction method using a dynamic background, the foreground is influencing the background. 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. Not updating the model within the detected ROI (see below) is not a solution, since a background model that is not updated over a longer time is also not representative anymore due to changes in lighting conditions.

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 detected foreground region larger than it should be. To remove this shadow, we used the property that a shadow only changes the intensity of the pixel while the texture of the covered region does not change [6]. As a result, the texture of the shadow is correlated with the corresponding texture of the background image. Jacques and Jung describe in [9] the usage of the cross correlation (CC) to see how good the detected foreground pixels match the background pixels. In case the cross correlation 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

Fig. 7. Extraction of fall features: purple: bounding box, white: bounding ellipse, green: center of gravity, blue: head position (The black box is for privacy reasons)

a threshold for the cross correlation of 0.98 together with a 5×5 neighborhood gives a good result. These values were also used in our experiments.

ROI Detection The next step in our algorithm was the determination of a region of interest (ROI). We first used an erosion/dilation step on all foreground pixels. Next, we applied a connected components analysis to determine the foreground objects. The largest object in the foreground was selected and considered to correspond to the person. As noted earlier, selecting the largest foreground object is prone to errors, since furniture or walking aids may move as well. A better choice is to rely on a tracker. However, this was left as future work. To minimize noise and interference, the object had to be larger than a certain threshold. In our case, a minimum of 17500 pixels gave the best performance. From this object we started to extract the features to detect a fall.

4.2 Fall Detection Features

Using the person, we extracted four features to detect a fall, including: aspect ratio (AR) [1, 2, 13, 27], fall angle (FA) [19, 27], center speed (CS) [19] and head speed (HS) [5, 12] (see Figure 7). These features have been designed based on domain knowledge, i.e. in such a way that they capture relevant information to discriminate falls from other actions, while at the same time being robust to inaccuracies in the person detection. These are also the most widely used in the literature, as explained in Section 3.

Aspect ratio The aspect ratio is calculated as the ratio of the width of the bounding box (BB) around the foreground object and its height. A low aspect ratio represents an upright person, while a high aspect ratio might point to a person lying down.

Fall Angle 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 direction. A person that is standing, has an angle close to 90 degrees. A small angle represents a person lying down (if seen from a side-view). We defined the fall angle as the change in angle over a fixed timespan (2 seconds in our experiments). A large fall angle can indicate a fall.

Center speed and head speed 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 used two fall features related to speed, center speed and head speed. Center speed is the speed of the center of gravity of the foreground object. This center of gravity has the advantage that it is rather stable. Small changes in appearance of the person give only small changes in the center of gravity. But an occlusion of the lower body, which happens frequently, causes the center of gravity to move upwards. The head, on the other hand, is visible in most non-fall actions. In [5] Foroughi *et al.* define the head as the highest point of the object. Here we used the highest end of the main axis of the bounding ellipse as head position. The speed itself was then defined as the amount of pixels that the point had shifted between two adjacent frames in the video divided by the time between these two frames.

4.3 Fall Detection with SVM

Given that the features defined in the previous section are based on domain knowledge, each of them can be used as a basic fall detector simply by choosing an appropriate threshold (as done e.g. in [27]). However, better results can be obtained if they are merged, and a single classifier combining the different cues is learned. In this section we propose a Support Vector Machine (SVM) [26] based fall detector which classifies a time slot (by its features) either as a fall or as another event.

As noted earlier, the 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 [17]. In the SVM learning objective errors for the minority class were multiplied by w while majority errors were multiplied by $1 - w$. How we determined w, is explained later.

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

The hyper-parameters used in this paper are (a) the weight w, (b) the regularization parameter of the SVM and (c) the Radial Basis Function (RBF) kernel bandwidth. These were 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 was 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 were computed on different data partitionings.

Additionally, feature selection was 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) were compared in terms of AUC. Next, the best feature (corresponding to the best SVM model) was retained and combined with each of the remaining features in a bivariate SVM model. The best feature set was retained and the procedure was repeated to find the best feature set with incremented cardinality. Note that features were standardized to have zero mean and unit standard deviation.

5 RESULTS

As mentioned before, we acquired an extensive dataset. To validate the algorithm, we used for each of the 24 falls, the camera on which the person is best visible. From this video, we selected a fragment of 20 minutes with the fall occurring in the last two minutes of the video. Our current system does not use the post-fall information (i.e., the person lying on the floor). Each video was divided in non-overlapping time slots of two minutes long. For each time slot, the fall features were extracted and the maximum values during that time slot were used for further analysis (max pooling). In total this resulted in 240 epochs, of which 24 are labeled as a fall. In a real system, the choice of the cameras could be dealt with using a voting mechanism. The extraction of the different fall features was executed on a pc with an Intel Core2 Quad Core Q9650 CPU running at 3 GHz. The algorithm was implemented in C++ using OpenCV. We can run four threads with different video, each processing around eight frames per second.

Given our four 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. [1] Table 1 lists the averaged AUC scores and the corresponding standard deviations for SVM models based on different feature sets. Figure 8 and Figure 9 respectively present the ROC and Precision Recall curves of the four best performing SVM models (measured in terms of AUC). It can be observed that the combination of aspect ratio and head speed is to be preferred. Using this feature set SVM outputs an averaged operating point with a recall of $0.9(\pm 0.2)$ and a precision of $0.26(\pm 0.07)$. Another observation is that the fall angle performs significantly lower than the other features.

Considering Figure 9, we noticed that the precision quickly drops when increasing the recall. This behavior can be explained by looking at Figure 10, that represents the distribution of the data when considering features aspect ratio and

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

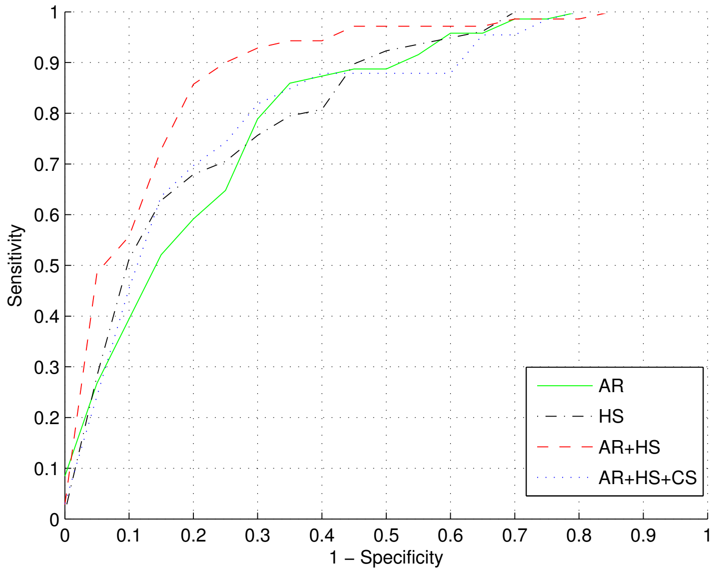

Fig. 8. ROC Curve

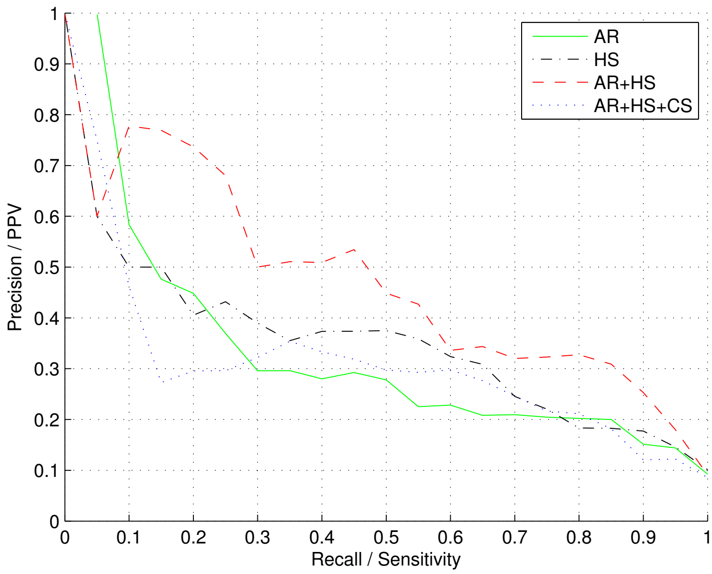

Fig. 9. Precision Recall graph

Table 1. Fall detection results

Feature set	AUC
$\{AR\}$	0.88(±0.06)
$\{FA\}$	0.53(±0.09)
$\{CS\}$	0.84(±0.05)
$\{HS\}$	0.87(±0.05)
$\{AR, HS\}$	0.91(±0.06)
$\{AR, HS, FA\}$	0.90(±0.05)
$\{AR, HS, FA, CS\}$	0.86(±0.06)

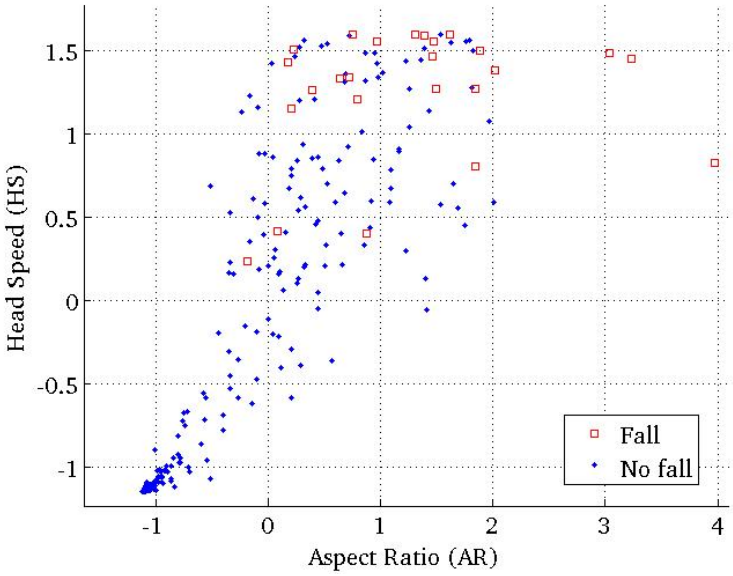

Fig. 10. Class distribution using normalized aspect ratio and head speed

head speed. Here we can see that there are quite a number of non-falls that are close to the falls. Closer visual inspection revealed that 90% of these have 4 main causes. In 25% of the cases 2 persons were present in the room. In 20% of the cases another foreground object had almost the same size as the person. In both cases, the system often switched to the other person or object, resulting in large motions and changes in aspect ratio. In 25% of the cases, the person's image was split in 2 blobs which were almost the same size. Situations where such an event occurs included: over-illumination, the person wearing a shirt that is similar to the background or the person starting to be integrated in the background by the background update. This often resulted in a deviating aspect ratio as well as large motions as the system jumps back and forth between the different parts. Finally, in 20% of the cases there was interference of a ghost figure or moved furniture.

6 DISCUSSION

Comparing our results with those reported in the literature [13][18], we have a similar or higher detection rate, but a higher false alarm rate. Two out of the three undetected falls started and ended outside of the view of the camera. This was e.g. the case when the older person was taking something out of the closet. The door was occluding the person at the start of the fall. During the fall the person was visible just for a very short time, before tumbling in the bathroom. A better placing or additional cameras can solve this. The higher false alarm rate can be explained by the challenging nature of our dataset, 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 an acceptable level 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 expected from the use of 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. This is the first step that we will investigate further. 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 a mixture of Gaussians to model the background showed no improvement on first sight. A means to detect a person in the foreground, like for example the person detector of Felzenszwalb *et al.* [3] can also reduce erroneous foreground objects. This detector is only trained for standing persons (both whole body and upper body), but it can still help as a verification every now and then. Alternatively, an articulated pose estimator such as [28] may be used as well. In Figure 5 it did not perform well. However, given a good initialization based on foreground detection, it may be useful.

Additional improvements may be possible by adding other fall features (e.g. posture or other appearance-based approaches), integrating information of several cameras or other sensors and especially by integrating the post-fall information.

In our tests, we used the camera on which the person was best visible. In a real system, this choice has to be made automatically. A voting mechanism, that uses the information how certain the system is that a fall occurred, can be implemented for this. This knowledge of the certainty of the fall can also be used to determine the needed action.

To reduce the annoyance of the false alarms, it is also possible to use an alarming chain. A possible fall could first be presented to the resident itself, if he doesn't react, a further escalation to different levels of caregivers can be executed.

7 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 have given 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 recall of 0.9($\pm$0.2) and a precision of 0.26($\pm$0.07). This calls for further research into more discriminative fall features, as well as better foreground detection algorithms, including tracking and person detection.

ACKNOWLEDGMENTS

This work is funded by the FWO via project G039811N: "Monitoring van gedrag en ongebruikelijke menselijke activiteit met meerdere camera's", by the IWT via TETRA project 80150 "Fallcam: Detection of fall in older persons with a camera system." and by the EU via ERASME (FP7) project IWT 100404 "AMACS: Automatic Monitoring of Activities using Contactless Sensors." The authors like to thank the persons who participated in the research by giving their permission to be monitored during several months.

References

1. Anderson, D., Keller, J., Skubic, M., Chen, X., He, Z.: Recognizing falls from silhouettes. In: Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE. pp. 6388 –6391 (SEPTEMBER 2006)
2. Cucchiara, R., Prati, A., Vezzani, R.: An intelligent surveillance system for dangerous situation detection in home environments. Intelligenza Artificiale 1(1), 11–15 (2004)
3. Felzenszwalb, P., Mcallester, D., Ramanan, D.: A discriminatively trained, multiscale, deformable part model. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Anchorage, Alaska, June 2008. (JUNE 2008)
4. Fleming, J., Brayne, C.: Inability to get up after falling, subsequent time on floor, and summoning help: prospective cohort study in people over 90. British Medicine Journal 337(v17 1), 2227 (2008)
5. Foroughi, H., Aski, B., Pourreza, H.: Intelligent video surveillance for monitoring fall detection of elderly in home environments. In: Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on. pp. 219–224 (2008)
6. Grest, D., Frahm, J., Koch, R.: A color similarity measure for robust shadow removal in real-time. Vision, Modeling and Vizualization (2003)

7. Haentjens, P., Lamraski, G., Boonen, S.: Costs and consequences of hip fracture occurrence in old age: An economic perspective. Disability and Rehabilitation 27(18-19), 1129–1141 (2005)
8. Hartholt, K.A., van der Velde, N., Looman, C.W.N., van Lieshout, E.M.M., Panneman, M.J.M., van Beeck, E.F., Patka, P., van der Cammen, T.J.M.: Trends in fall-related hospital admissions in older persons in the netherlands. Arch Intern Med 170(10), 905–911 (2010)
9. Jacques, J.C.S., Jung, C.R.: Background subtraction and shadow detection in grayscale video sequences. The XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI05) (2005)
10. Lee, T., Mihailidis, A.: An intelligent emergency response system: preliminary development and testing of automated fall detection. Journal of Telemedicine and Telecare 11(4), 194–198 (2005)
11. McFarlane, N.J.B., Schofield, C.P.: Segmentation and tracking of piglets in images. Machine Vision and Applications 8(3), 187–193 (MAY 1995)
12. Miao, Y., Naqvi, S., Chambers, J.: Fall detection in the elderly by head tracking. In: Statistical Signal Processing, 2009. SSP '09. IEEE/SP 15th Workshop on. pp. 357 –360 (SEPTEMBER 2009)
13. Miaou, S.G., Sung, P., Huang, C.: A customized human fall detection system using omni-camera images and personal information. Distributed Diagnosis and Home Healthcare pp. 39–42 (2006)
14. Milisen, K., Detroch, E., Bellens, K., Braes, T., Dierickx, K., Smeulders, W., Teughels, S., Dejaeger, E., Boonen, S., Pelemans, W.: Falls among community-dwelling elderly: a pilot study of prevalence, circumstances and consequences in flanders. Tijdschr Gerontol Geriatr 35(1), 15–20 (2004)
15. Nait-Charif, H., McKenna, S.J.: Activity summarisation and fall detection in a supportive home environment. In: ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 4. pp. 323–326. IEEE Computer Society, Washington, DC, USA (2004)
16. Nater, F., Grabner, H., Van Gool, L.: Visual abnormal event detection for prolonged independent living. In: International mobile health (mHealth) workshop (2010)
17. Osuna, E., Freund, R., Girosi, F.: Support vector machines: Training and applications. AI Memo 1602, Massachusetts Institute of Technology (1997)
18. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Monocular 3d head tracking to detect falls of elderly people. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. (2006)
19. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: 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. vol. 2, pp. 875–880 (2007)
20. Scuffham, P., Chaplin, S., Legood, R.: Incidence and costs of unintentional falls in older people in the United Kingdom. J Epidemiol Community Health 57(9), 740–744 (Sep 2003)
21. SeniorWatch: Fall detector: Case study of european ist seniorwatch project. Tech. rep., SeniorWatch (2001)
22. Syngelakis, E., Collomosse, J.: A bag of features approach to ambient fall detection for domestic elder-care. In: Proc. Intl. Symp. on Ambient Technologies (AMBIENT 2011) (2011)

23. Thome, N., Miguet, S., Ambellouis, S.: A real-time, multiview fall detection system: A lhmm-based approach. Circuits and Systems for Video Technology, IEEE Transactions on 18(11), 1522–1532 (2008)
24. Tinetti, M.E.: Preventing falls in elderly persons. New England Journal of Medicine 348(1), 42–49 (2003)
25. Toreyin, B.U., Dedeoglu, Y., Cetin, A.E.: Hmm based falling person detection using both audio and video. In: Computer Vision in Human-Computer Interaction, Proceedings, Lecture Notes in Computer Science, vol. 3766, pp. 211–220. Springer-Verlag Berlin (2005)
26. Vapnik, V.: Statistical learning theory. Wiley, New York (1998)
27. Willems, J., Debard, G., Vanrumste, B., Goedemé, T.: A video-based algorithm for elderly fall detection. Medical Physics and Biomedical Engineering World Congress, WC2009 (2009)
28. Yang, Y., Ramanan, D.: Articulated pose estimation with flexible mixtures-of-parts. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. pp. 1385 –1392 (JUNE 2011)