Detecting Falls at Homes Using a Network of Low-Resolution Cameras

Sebastian Zambanini, Jana Machajdik and Martin Kampel

Abstract— **In a smart home system, a camera-based fall detector at elderly homes leads to immediate alarming and helping. In this paper we propose an approach for the detection of falls based on multiple cameras. Based on semantic driven features, fall detection is done in 3D and fuzzy logic is used to estimate confidence values for different human postures as well as for the incidence of a fall/no fall. Emphasis is given on simplicity, low computational effort and fast processing. Therefore, based on an evaluation on 73 test sequences, we show the applicability of the method for videos with low spatial resolution and frame rate.**

I. Introduction

Currently, in the European Union about 30 % of people older than 65 live alone [1], with an upward trend. Smart homes bear the potential to improve the life quality of elderly and disabled by supporting them in their daily routines and fulfilling their special needs. As part of a smart home system for elderly, the permanent monitoring of the inhabitants is of high value, e.g. to automatically detect cases of emergency. Within the MuBisA project[1], monitoring is achieved by a network of digital cameras, providing both flexibility and expandability: using just one sensor type, a vast amount of events (e.g. falls, fire, flooding...) can be detected with the appropriate computer vision techniques. Moreover, compared to the prevalently used emergency system which includes mobile devices worn by the elderly, camera-based monitoring is completely passive and thus eliminates the shortcomings of this system, e.g. the need for human activation in case of an accident or the risk of forgetting to wear the device.

Falls at homes are one of the major risks for elderly and an immediate alarming and helping is essential to reduce the rate of morbidity and mortality [2]. In this paper we present a fall detection system based on a network of cameras. The system uses inexpensive low-resolution cameras, in order to make the system flexible and affordable for the elderly in the future.

Camera-based fall detection approaches proposed in the past work either by modeling the temporal characteristics of the fall action itself or by detecting falls explicitly from human posture and motion speed on a frame-by-frame basis. In the former type of methods, parametric models like Hidden Markov Models are trained using simple features, e.g. projection histograms [3] or the aspect ratio of the bounding box surrounding the detected human [4]. However, the applicability of these methods in real-life scenarios is limited due to the high diversity of fall actions and the high number of different negative actions which the system should not classify as fall. The second type of methods basically measures two types of features: the human posture and the motion speed. The underlying assumption is that a fall is characterized by a transition from a vertical to a horizontal posture with an unusually increased speed, i.e. to discern falls from normal actions like sitting on a chair or lying on a bed. In this manner, in the past various features have been used for camera-based fall detection, including the aspect ratio of the bounding box [5] or orientation of a fitted ellipse [6] for posture recognition and head tracking [7] or change rate of the human's centroid [8] for motion speed. Apart from the features used, the methods also differ in the way how the final decision is derived from the features. Besides parametric classifiers like Neural Networks [9], primarily empirically determined rules are applied [7], [8], [10]. In order to reduce false alarms, a final verification step can be performed which measures if the person was able to move and stand up again in a given period of time.

Our method basically follows the approach of Anderson et al. [10]: initially, the human silhouette detected in the different cameras is used to obtain a 3D reconstruction in voxel space. Features extracted from this rough reconstruction of the human are then finally used to reason about a fall. Although in our work we rely on a 3D reconstruction as in [10], we partly use different kinds of features and a substantially simpler decision process. We mainly contribute to their work by thoroughly evaluating the method using a comprehensive set of test sequences and by showing that the method can be applied to videos with both low spatial resolution and low frame rate. Using this setup, fall detection can be achieved in real-time without the need for powerful and expensive hardware.

II. Methodology

In our methodology we focus on simplicity, low computational effort and therefore fast processing without the need of high-end hardware. These design goals render, for instance, sophisticated model-based approaches for posture recognition infeasible. Therefore, posture recognition is kept simple and estimates basically the general orientation of the human body, i.e. standing/vertical or lying/horizontal. For this purpose, detected motion in calibrated cameras is fused to obtain a 3D voxel reconstruction of the human. Features are extracted from voxel space and combined to confidence values for different posture states and for the occurrence of

Manuscript received August 24, 2010. This work was supported by the Austrian Research Promotion Agency (FFG) under grant 819862 (MuBisA)

S. Zambanini, J. Machajdik and M. Kampel are with the Computer Vision Lab, Institute of Computer-Aided Automation, Vienna University of Technology, A-1040 Vienna, Austria {zamba,jana,kampel}@caa.tuwien.ac.at

[1] http://www.cogvis.at/mubisa/

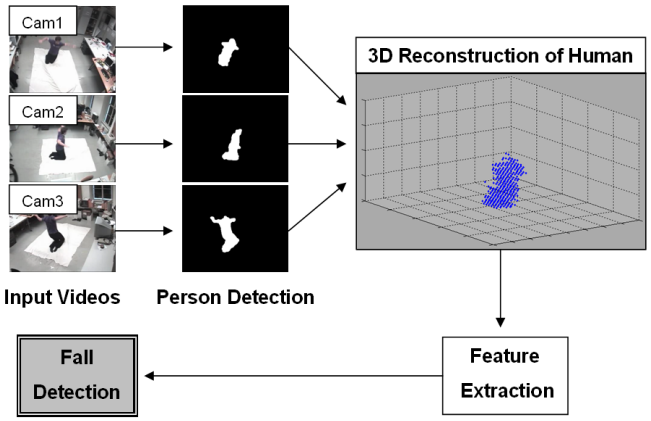

Fig. 1. The workflow of the presented fall detection method.

a fall using fuzzy logic [11]. Fig. 1 exemplarily shows the method's workflow for a setup consisting of three cameras, divided into the steps of person detection (Section II-A), 3D reconstruction (Section II-B), feature extraction (Section II-C) and estimation of posture and fall confidence values (Section II-D)

A. Person Detection

Segmentation of the person from the background is the first step in our fall detection process. In the current state, person detection is kept simple and a more sophisticated person detection will be part of future work. We apply simple background subtraction with a slowly adapting background model to detect motion [12]. To remove noise from the motion image we make use of several morphological operations. Since the system is designed for elderly living alone, we simply choose the largest connected component to mark the region representing the person. Detection of multiple persons at the same time is not considered, as automatic fall detection and alarming is assumed to be unnecessary when more than one person is present. The result of the human detection procedure is a mask with a rough silhouette of the human in each camera frame, i.e. a set of silhouette pixels $\mathcal{P}_{c,i}$, where c is the camera index and i is the frame index. These silhouette pixels serve as input for the 3D human reconstruction.

B. 3D Reconstruction of Human

For the 3D reconstruction Shape-from-Silhouette [13] (also known as visual hull reconstruction) is used, since we are able to apply this technique directly to the binary motion images from calibrated cameras and the achieved rough reconstruction is sufficient for our task of rough posture estimation, i.e. to differentiate between a lying and a standing posture. From all camera views c we have to find the intersection of the visual rays going through the points in $\mathcal{P}_{c,i}$. In order to keep the processing time within reasonable limits, a preprocessing step is applied which constructs a voxel list $L_c(m,n)$ for all image points (m,n) and all cameras c. The voxel list $L_c(m,n)$ stores all voxels $v(x,y,z)$ in the scene that are intersected by the visual ray going through the image point (m,n) in the c-th camera. Once this voxel list has been build, every camera c defines a set of voxels $\mathcal{V}_{c,i} = \cup L_c(m,n)$ for all (m,n) in $\mathcal{P}_{c,i}$. The reconstruction $\mathcal{V}_i$ is finally obtained by an intersection test, i.e. $\mathcal{V}_i = \cap \mathcal{V}_{c,i}$ for all c.

C. Feature Extraction

We use a set of straightforward semantic driven features which is inspired by previous works [5], [6], [8], [10] and chosen based on empirical experiments. We discern between the intra-frame features which are computed within each frame and focus on describing the character of the object, i.e. the posture, and an inter-frame feature which expresses the character of the change that happens between consecutive frames.

In particular, the following features are extracted at every frame with index i from the set of voxels $\mathcal{V}_i$ representing the person:

- Intra-frame features
  - **Bounding Box Aspect Ratio** (B_i): The height of the bounding box surrounding the person divided by the mean of both its widths.
  - **Orientation** (O_i): The orientation of the major axis of the ellipse fitted to the person, specified as the angle between the major axis and the groundplane.
  - **Axis Ratio** (A_i): The ratio between the lengths of the longest axis and the second longest axis of the ellipse fitted to the person.
- Inter-frame feature
  - **Motion Speed** (M_i): The relative number of new motion voxels in the current frame compared to the previous frame: $M_i = |\mathcal{V}_i \backslash (\mathcal{V}_i \cap \mathcal{V}_{i-1})| \,/\, |\mathcal{V}_i|$.

D. Fuzzy-Based Estimation of Posture and Fall Confidence Values

In conformity with Anderson et al. [10], we define three posture states in which the person may reside: *"standing"*, *"in between"* and *"lying"*. Sets of primarily empirically determined fuzzy thresholds in the form of trapezoidal functions are assembled to interpret the intra-frame features and relate them to the postures. Thus, each feature value results in a confidence value in the range $[0,1]$ on each posture, where the confidences of one feature sum up to 1 for all postures. These are then combined to assign a confidence value for each posture which is determined by a weighted sum of all feature confidences. The membership functions for the orientation O_i are exemplarily shown in Fig. 2.

From the computed confidence values for the different postures, for every frame a confidence value for a fall event

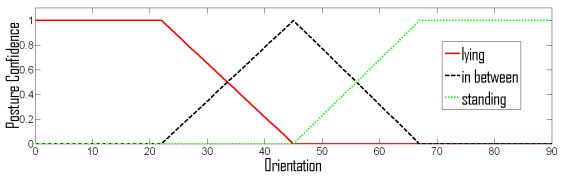

Fig. 2. Membership functions for the three postures and the intra-frame feature O_i.

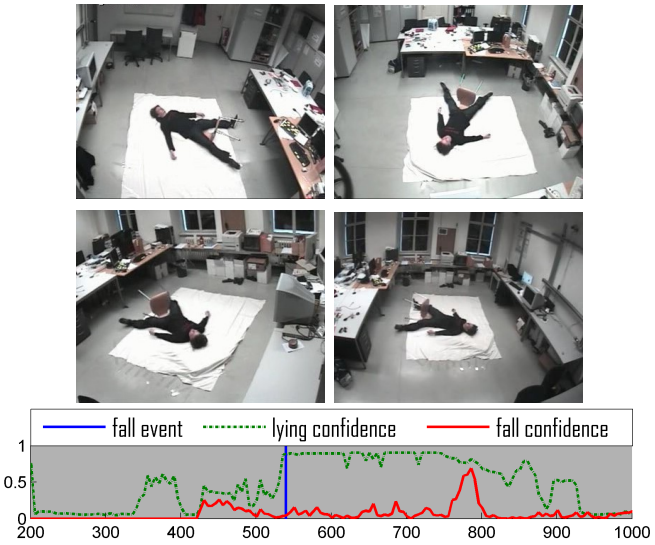

Fig. 3. Confidence values for the posture *"lying"* and a fall event plotted over time for a test sequence showing a person falling from a chair.

is computed. Therefore, we combine the intra- and inter-frame features with the assumption that a fall is defined by a relatively high motion speed, followed by a period with a *"lying"* posture. Thus, the confidence for a fall event at frame i is computed as the motion speed M_i multiplied by the confidence values for the posture *"lying"* for the next k frames.

Fig. 3 shows the estimated confidences for the posture *"lying"* and a fall over time for a given test sequence acquired from four cameras. The particular sequence shows a simulated fall from a chair. The fall occurs approximately at frame number 540, thus a peak in the fall confidence can be spotted at frame number 790, i.e. 250 frames later (please note that this "delay" is caused by $k = 250$).

E. Anonymization of Video Data

Since digital cameras serve as sensors for our system, the protection of privacy is a major concern. As we detect falls in real-time, no visual data has to be stored in general. However, if a fall occurs, the system makes an anonymous snapshot of the scene. The snapshot only shows the person silhouette and an edge image of the environment. An example is shown in Fig. 4.

III. Experiments

In order to thoroughly evaluate our fall detection method, test sequences were acquired that follow the scenarios described by Noury et al. [14]. Hence, a testset consisting of various types of falls as well as various types of normal actions was created. A complete list is given in Table I. Four cameras with a resolution of 288×352 and frame rate of 25 fps were placed in a room at a height of approx. 2.5 meters. The four camera views are shown in Fig. 3. Five different actors simulated the scenarios resulting in a total of 43 positive (falls) and 30 negative sequences (no falls).

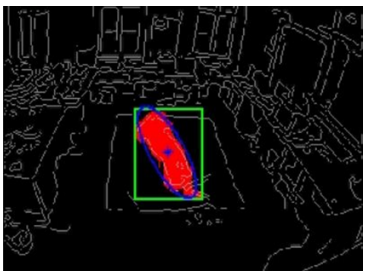

Fig. 4. Anonymized snapshot of a fall event.

TABLE I

Acquired test scenarios with corresponding number of videos in the testset.

Category	Name	Outcome	#
Backward fall	Ending sitting	Positive	4
	Ending lying	Positive	4
	Ending in lateral position	Positive	3
	With recovery	Negative	4
Forward fall	On the knees	Negative	6
	Ending lying flat	Positive	11
	With recovery	Negative	5
Lateral fall	Ending lying flat	Positive	13
	With recovery	Negative	1
Fall from a chair	Ending lying flat	Positive	8
Syncope	Vertical slipping against a wall finishing in sitting position	Negative	2
Neutral	To sit down on a chair then to stand up	Negative	4
	To lie down then to rise up	Negative	2
	To walk around	Negative	1
	To bend down, catch something up on the floor, then to rise up	Negative	2
	To cough or sneeze	Negative	3

In contrast to the definition given in [14], we consider falls ending on the knees as negative instances which the system should not detect as fall. The reason is that in this case people are whether still able to move, i.e. they would stand up, or would consequently lie down and thus the alarm would be initiated.

For the given testset, the parameter k defining the considered time period of the *"lying"* posture for fall detection (see Section II-D) was set to 10 seconds. Please note that in a real scenario this parameter has to be set to a higher value. In our simulated falls the lying periods are considerably shorter than they would be in case of a real fall event, for obvious reasons.

Since our method results in confidence values for a fall event in every tested frame, we report its sensitivity and specificity in the form of ROC curves. For generation of the ROC curve, true positives and false positives were counted as the number of positive and negative sequences, respectively, where a fall confidence above the threshold could be found. In order to evaluate the influence of the videos' spatial and temporal resolution, we successively reduced the resolutions and tested each resolution on the whole dataset. Spatial resolution was successively halved from 288×352 down to 9×11. For temporal resolution, frame rates of 5, 2.5, 1.25 and 0.5 frames per second (fps) were tested. The results are shown in Fig. 5.

It can be seen that the proposed method shows similar

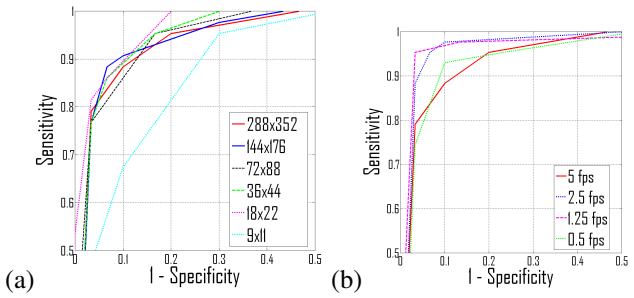

Fig. 5. ROC curves for different (a) spatial and (b) temporal resolutions of the test sequences.

performance for spatial resolutions down to 18×22, and significantly drops at a resolution of 9×11. The areas under curve (AUCs) of the resolutions 288×352, 18×22 and 9×11 are 0.956, 0.974 and 0.910, respectively. Temporal resolution analysis reveals slightly better performance at a lower frame rate of 1.25 fps (AUC= 0.974) compared to 5 fps (AUC= 0.953). The reason for this enhancement is that at a lower frame rate the motion speed becomes a more robust feature, since the motion speed measurements are more reliable for longer time intervals.

The results show that the discriminative power of the chosen features is high enough to correctly classify the majority of the sequences and the extraction of features is stable even at low spatial resolutions and frame rates. Inspection of the results on particular sequences reveals that false classifications are mainly caused by the imperfect person detection. For instance, sitting actions are likely classified as falls since the chair moved by the person heavily interferes with the simple person detection. Reliable person detection is essential and will be part of future work. False negatives are primarily caused by a partially short lying period of the test person after the simulated fall. As for this evaluation the considered time period was set to 10 seconds, rising up before this period leads to lower fall confidences.

Another conclusion from the tests is that the motion speed during a fall is a helpful but limited feature. An increased motion speed during a transition from a vertical to a horizontal posture is a strong clue that a fall has happened. However, a specific "minimum" motion speed for a fall can not be identified, and therefore a fall detector can not rely on motion speed only. According to caretakers, this is even more critical for elderly who can possibly fall with very low speed.

IV. Conclusions and Outlook

We have presented an approach for elderly fall detection in a network of cameras with low spatial and temporal resolutions. Due to the decreased amount of data to be acquired and processed, the system is able to work on low-cost cameras in real-time.

In the absence of real fall data, tests have to be performed by actors simulating the falls in lab conditions in a preferably realistic way. Although for this reason the evaluation data can not be seen as "perfect", we followed well-defined scenarios and tried to capture the large diversity of fall actions and normal activity at home. For this given test data, the experimental results have shown the general applicability of the approach. However, there is lot of space for improvements in the future. As human silhouettes serve as input for our fall detection system, robust person detection and tracking is crucial and will be further investigated [15]. Our system can also be extended towards a more sophisticated reasoning, e.g. to detect falls that do not end in a characteristic lying posture. Thus, more powerful rules will be defined in cooperation with caretakers and health organizations which are able to cope with fall events in real conditions.

In the future, prototype installations will show the real challenges of the various environments and life styles of the elderly (overfilled flats, pets, dementia, active life style (e.g. exercising), visitors, etc.). Arguably, the manual or automatic definition of inactivity zones [16] will be necessary to make the system more robust against normal sitting and lying actions. Since a conclusion from this paper is the possible use of cheap cameras (low spatial and temporal resolution), future research will also determine detailed hardware specifications which allow a reliable fall detection at home. This includes also the number of cameras needed and their optimal positioning, in order to give an estimate of the overall costs of the system.

References

[1] *The Life of Women and Men in Europe : A Statistical Portrait*. Eurostat, 2008 edition, 2008.

[2] D. Wild, U.S. Nayak, and B. Isaacs. How dangerous are falls in old people at home? *Br Med J*, 282(6260):266–268, 1981.

[3] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani. Probabilistic posture classification for human-behavior analysis. *SMC-A*, 35(1):42–54, 2005.

[4] D. Anderson, J.M. Keller, M. Skubic, X. Chen, and Z. He. Recognizing falls from silhouettes. In *Proc. of EMBS*, pp. 6388–6391, 2006.

[5] J. Tao, M. Turjo, M.F. Wong, M. Wang, and Y.P. Tan. Fall incidents detection for intelligent video surveillance. In *Proc. of ICICS*, pp. 1590–1594, 2005.

[6] N. Thome, S. Miguet, and S. Ambellouis. A Real-Time, Multiview Fall Detection System: A LHMM-Based Approach. *IEEE TCSVT*, 18(11):1522–1532, 2008.

[7] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Monocular 3D head tracking to detect falls of elderly people. In *Proc. of EMBS*, pp. 6384–6387, 2006.

[8] C.W. Lin, Z.H. Ling, Y.C. Chang, and C.J. Kuo. Compressed-domain Fall Incident Detection for Intelligent Homecare. *VLSISP*, 49(3):393–408, 2007.

[9] C. Huang, E. Chen, and P. Chung. Fall detection using modular neural networks with back-projected optical flow. *BME*, 19(6):415–424, 2007.

[10] D. Anderson, R.H. Luke, J.M. Keller, M. Skubic, M. Rantz, and M. Aud. Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *CVIU*, 113(1):80–89, 2009.

[11] L.A. Zadeh. Fuzzy sets. *Information and control*, 8(3):338-353, 1965.

[12] M. Piccardi. Background subtraction techniques: a review. *Proc. of IEEE SMC*, pp. 3099–3104, 2004.

[13] C.R. Dyer. Volumetric scene reconstruction from multiple views. *Foundations of Image Understanding*, pp. 469–489, 2001.

[14] N. Noury, A. Fleury, P. Rumeau, AK Bourke, GO Laighin, V. Rialle, and JE Lundy. Fall detection–Principles and methods. In *Proc. of EMBS*, pp. 1663–1666, 2007.

[15] R. Poppe. Vision-based human motion analysis: An overview. *CVIU*, 108(1-2):4-18, 2007.

[16] H. Nait-Charif and S.J. McKenna. Activity summarisation and fall detection in a supportive home environment. In *Proc. of ICPR*, volume 4, pp. 323–326, 2004.