Computer Vision for an Independent Lifestyle of the Elderly - An Overview of the MuBisA Project

Sebastian Zambanini*, Martin Kampel and Jana Machajdik

Pattern Recognition and Image Processing Group, Vienna University of Technology, Favoritenstr. 9/1832, A-1040 Vienna, Austria

E-Mails: zamba@prip.tuwien.ac.at (S. Z.); kampel@prip.tuwien.ac.at (M. K.); jana@prip.tuwien.ac.at (J. M.)

* Tel.: +43 1 58801-18366; Fax: +43 1 58801-18392

Abstract: Many elderly people are in need of care. However, the care in special facilities is expensive and a 24/7 support is not always possible, nor is it always needed. To allow the elderly to live in their own home longer, while giving them the needed prospect of security and the availability of help in moments when it is really needed, home assistance systems are of high value. The MuBisA project develops a closed system for the automated event detection and the communication with mobile devices. It is based on the idea of smart homes, but uses state-of-the-art 2D and 3D computer vision techniques. Cameras will be connected to a central computer which will perform the processing and automatic analysis of the image sequences. In case of a detected emergency (e.g. the person has fallen) an alert signal will be sent to a designated assistant or a call center. The system provides a mobile, configurable event notification for elderly, attendants and relatives and guarantees high security through the integration of the Austrian citizen card. In this paper an overview of the project is given. Additionally, initial results for camera-based fall detection are reported.

Keywords: Smart Home Healthcare; Video Analysis; Mobile Alerting; Fall Detection

1. Introduction

In the European Union about 30 % of people older than 60 live alone [1]. For these people, collapsing at home is one of the major risks and an immediate alarming and helping is essential to reduce the rate of morbidity and mortality [2]. Research found that half of those patients with a “long lie” (i.e. those remaining on the ground for more than one hour after a fall) died within six months of the fall, even if there was no direct physical injury (death was usually a complication such as bronchopneumonia, dehydration or hypothermia). Even if the fall didn’t cause any physical injury, the impact on the psyche should not be underestimated. In many cases such accidents represent the first step to loosing independence for elderly people.

The goal of the MuBisA project is a reliable and automated computer vision system to enable an independent lifestyle for the elderly and disabled. In contrast to other projects, the system relies solely on computer vision techniques. The daily lives of the people involved will not be affected and it overcomes many problems of existing systems. Main goal is the robust fall detection. Since the detection set-up consists basically of IP based cameras and a central storage and calculation server, the system is open and flexible towards other applications: Fire, smoke and water detection and assistance for medication can be introduced in a second step.

This paper will give a comprehensive overview of the MuBisA project. In the following sections, the consortium and the technical setup will be described. Additionally, research results for camera-based fall detection will be summarized. This includes a description of the state-of-the-art, the proposal of different fusion strategies for multi-camera fall detection as well as the reporting of preliminary results on simulated fall sequences.

2. Consortium

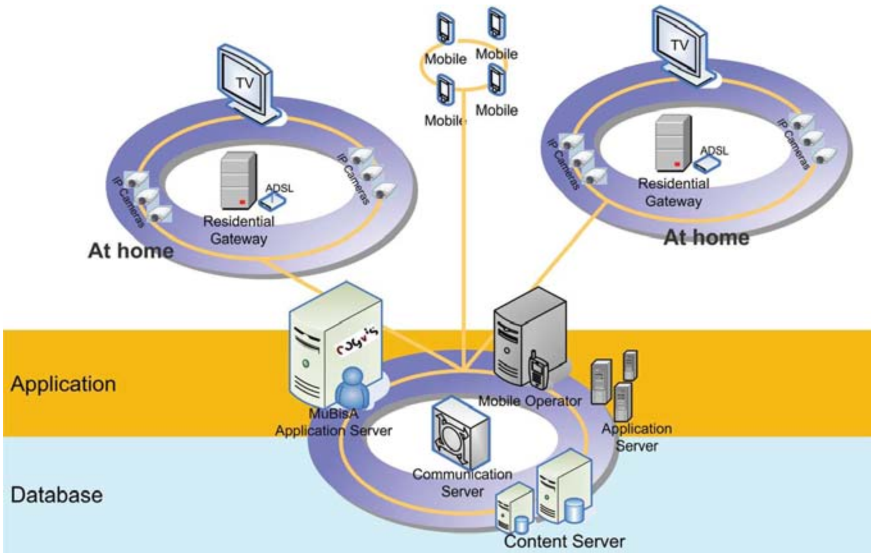
The consortium of the MuBisA project consists of the following partners:

- **CogVis Software and Consulting GmbH**, the project coordinator and responsible for the implementation and installation of the video analysis software.
- **Pattern Recognition and Image Processing Group, Vienna University of Technology**, responsible for providing proper scientific input into the development of the video analysis software by the investigation of state-of-the-art methods and the creation of innovative algorithms.
- **e-nnovation IT-Systeme GmbH,** brings in experience in IT solutions for the health sector and is responsible for the communications part of the project.
- **Medical University of Vienna**, provides the necessary medical know-how to the project.
- **Samariterbund,** non-profit aid organization for emergency medical service and ambulance service.
- **Fonds Soziales Wien,** organization in charge of home care in Vienna.

3. System Overview

The setup of the overall system is illustrated in Figure 1. Low cost web-cameras will be installed in all the rooms of the home of an elderly or disabled person. In each room, there will be 1-4 cameras, depending on the room setting. The purpose of using multiple cameras is mainly to avoid occlusions of the person in the video sequences and also to allow 3D-based analysis algorithms. To avoid complicated wiring in the consumers flat it is possible to use wireless web-cams which are nowadays widely available.

Figure 1. System overview of the MuBisA project.



All cameras of one home will be connected to a central computer which will perform the processing of the image sequences and event detection. Depending on the concrete hardware, it is possible that some parts of the preprocessing (like simple motion detection) will be done in the cameras themselves. The processing of video data and the fall detection will be done fully automatically, so that no privacy issues are challenged and the consumer does not feel being watched.

In case of an emergency an alert signal (e.g. an SMS) will be sent to a designated assistant or a call center. When receiving a fall alarm, the assistant can call the fallen person and ask them if they need instant help or if they can resolve the situation themselves (e.g. if it was a false alarm, if the fallen person is able to stand up by himself/herself or has assistance nearby, e.g. visiting family). A way of letting the user know that the system has “seen” him/her fall is also considered, since the fallen person needs reassurance that help is on the way. Additionally, a way of “clearing the system” is considered, e.g. for situations when the elderly have visitors, such as lively grand children who may throw themselves on the ground and activate the system unnecessarily.

4. Camera-Based Fall Detection

Today’s state of the art event detection in assisted living is an alarm button on a wristband. Such an approach has many drawbacks: it needs to be worn all the time (and in reality this is rarely the case) and it needs human activation in case of an accident – which is not possible in all cases, e.g. loss of consciousness – or, if a mechanical fall detection sensor is included in the band, the false alarm rate is very high.

Cameras are passive, multifunctional sensors with a wide spectrum of possible usage. They are non-intrusive, relatively reliable and can be used 24/7. Using video signals, the system’s functions are open to updates or adding of new analysis methods (like e.g. fire/smoke detection) without changing the hardware. Furthermore, the retrospective analysis of summary statistics, tracking trajectories and - where appropriate - image data, could be used to provide valuable insight into behavior and health of a person.

4.1 Related Work

In the last five years a growing interest and number of publications for camera-based fall detection has been shown [3]. A general classification of proposed methods can be made by whether a fall is detected by modeling the fall action itself or by a frame-by-frame classification using different features measuring human posture and motion velocity. In the former type of methods parametric models like Hidden Markov Models (HMMs) are applied [4-6]. 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 latter type of methods basically measures two types of features: the human posture and the motion velocity. The underlying assumption is that a fall is characterized by a transition from a vertical to a horizontal posture with an unusually increased velocity, 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 [7] or orientation of a fitted ellipse [8] for posture recognition and head tracking [9] or change rate of the human's centroid [10] for motion velocity. The final decision is derived from the features using parametric classifiers like Neural Networks [11] or empirically determined rules [8-10, 12]. 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.

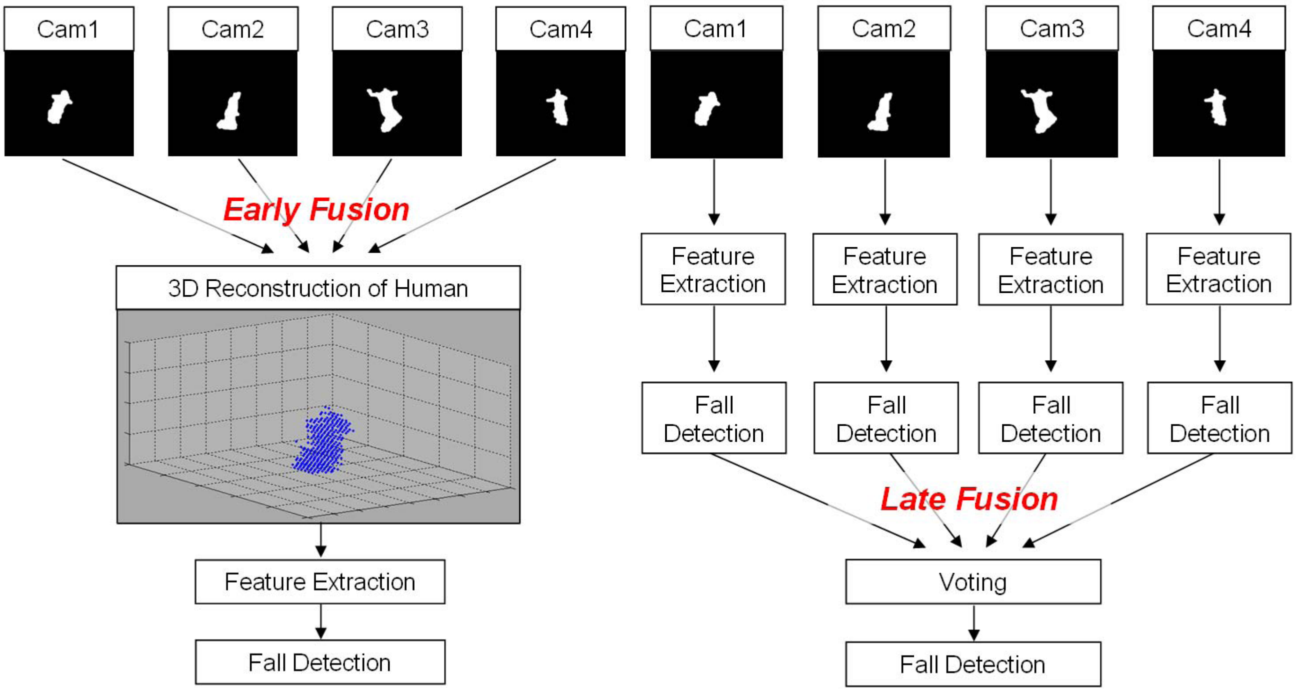
4.2 Fusion of Multiple Cameras

In the context of ambient assisted living, i.e. detecting falls at elderly homes, the use of a multiple camera network is inevitable. Multiple cameras allow for the monitoring of multiple rooms and the resolving of occlusions. Furthermore, the features for posture and motion velocity used in related work are highly dependent on the viewpoint of the camera, e.g. consider a fall along the optical axis of the camera versus a fall perpendicular to it. Therefore, in a real-life scenario robustness is highly increased when multiple cameras are used.

When using multiple cameras, a key question is when the fusion of the video data streams is performed in the overall detection process. Basically, two different detection schemes can be applied, called *early fusion* and *late fusion*. Both schemes are illustrated in Figure 2.

In the early fusion scheme, detected motion is fused together from calibrated cameras to obtain a 3D reconstruction of the human. This scheme offers a robust estimation of the human posture and thus view-invariant features for fall detection. However, the drawback of 3D reconstruction is that it needs camera calibration and demands higher computational effort which restrains the required real-time processing of the data.

Figure 2. Fall detection with early and late fusion of the detected motion in four camera views.



In the late fusion scheme, feature extraction and fall detection is performed individually in each camera. In a final voting step the individual decisions are fused together to an overall decision. This scheme tries to overcome the drawback of view-dependence of the extracted features by a well-adapted fusion strategy that needs no camera calibration and less computational effort.

Since the final system has to be as cheap as possible to be affordable for the elderly, simplicity, low computational effort and therefore fast processing without the need of high-end hardware are essential design goals. Therefore, for both the early and late fusion approach, posture recognition is kept simple and estimates basically the general orientation of the human body, i.e. standing/vertical or lying/horizontal. This is achieved by combining the extracted features to confidence values for different posture states using fuzzy logic [13]. In the late fusion approach, fuzzy logic is also used to fuse the confidence values of the various cameras to a final estimation of the state and for final fall detection.

We implement a collection of straightforward semantic driven features:

- **Bounding Box Aspect Ratio**: The height of the bounding box surrounding the human divided by its width.
- **Orientation**: The orientation of the major axis of the ellipse fitted to the human.
- **Axis Ratio:** The ratio between the lengths of the major axis and the minor axis of the ellipse fitted to the human.

- **Motion Speed:** The relative number of new motion pixels/voxels in the current frame compared to the previous frame.

Generally, 3 posture states, in which the human may reside, are defined: *"standing"*, *"in between"* and *"lying"*. Sets of primarily empirically determined fuzzy thresholds in the form of trapezoidal functions are assembled to interpret the 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.

4.3 Early and Late Fusion

In the early fusion approach, a 3D reconstruction is computed from the set of motion pixels. For this purpose Shape-from-Silhouette [14] is used, since we are able to apply this technique directly to the 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 the computed confidence values for the different postures, for every frame a confidence value for a fall event is computed. We assume that a fall is represented 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 multiplied by the confidence values for the posture "lying" for the next k frames.

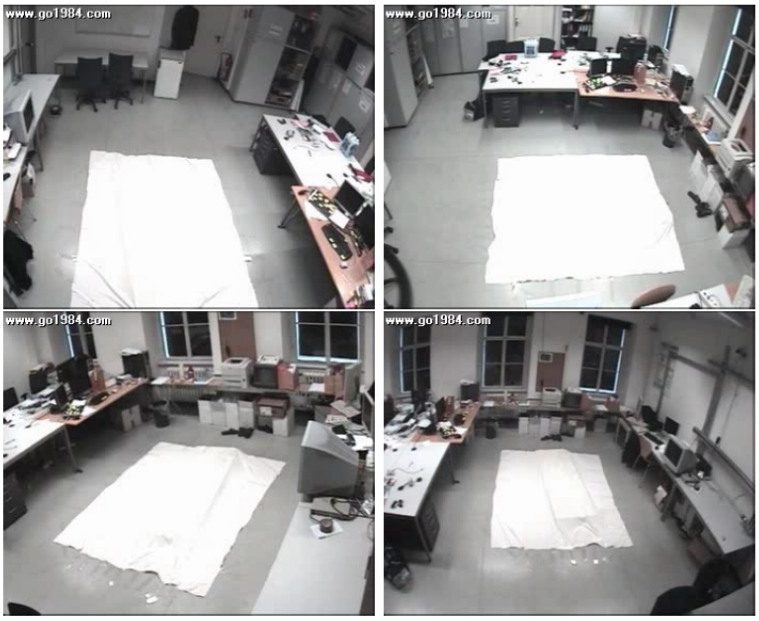
In the late fusion approach, the outputs of all cameras are combined to generate an overall decision. In our work, this is done by averaging the fall confidences from all the cameras. However, there are cases when only one camera actually "sees" the fall. This can happen when the person falls in the direction of the optical axis of some of the cameras. In such a case, cameras that are positioned along that axis will not recognize the fall at all, whereas a camera that is positioned in a perpendicular direction will have a clear view of the scene. A similar situation could arise in the presence of occlusions. Our solution in such a case is that if the confidence of the alarm is very high (above a defined threshold) even in a single camera, this one camera gets "the right to over-vote the other" and the overall fall confidence is determined by this camera only.

4. Experiments

In order to thoroughly evaluate our fall detection method, test sequences were acquired that follow the scenarios described by Noury et al. [15]. Hence, a challenging testset consisting of various types of falls (forward and backward falls, falls from chairs etc.) as well as various types of normal actions (picking something up, sitting down etc.) was created. Four cameras with a resolution of 288 x 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 Figure 3. Five different actors simulated the scenarios resulting in a total of 43 positive (falls) and 30 negative sequences (no falls). In contrast to the definition given in [15], 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.

Tests were performed on the overall dataset using the early fusion approach as well as the late fusion approach. Initial tests turned out that the whole temporal resolution is not needed in our case, hence a reduced frame rate of 5 fps was used.

Figure 3. The four camera views of the test sequences.



The final results of both methods are listed in Table 1. Early fusion reaches a sensitivity of 97.7 % and specificity of 86.7 %, whereas late fusion reaches 83.7 % and 76.7 %, respectively.

The results show that the discriminative power of the chosen features is high enough to correctly classify the majority of the sequences using our fuzzy-based estimation of fall confidence values. Inspection of the results reveals that misclassifications are mainly caused by the imperfect motion detection. For instance, sitting actions are likely classified as falls since the chair moved by the human heavily interferes with the simple human detection. Early fusion is naturally less vulnerable to this type of error since outliers are implicitly removed in the 3D reconstruction step. Therefore, we expect that the late fusion's performance will gain on the early fusion's performance, once a more sophisticated human detection is used. A second cause for misclassifications is the limited overlap in the field-of-view of the cameras, as actions which are not totally visible in all four cameras renders the feature extraction and 3D reconstruction less robust.

Table 1. Results of early and late fusion on the given data.

		Correct	False / Missed
Early Fusion	Alarm	42	1
	No Alarm	26	4
Late Fusion	Alarm	36	7
	No Alarm	23	7

4. Conclusions

We have given an overview of the MuBisA project which is aiming at an powerful and flexible alarming system for elderly homes based on automatic video analysis. In this context, we have proposed and compared two methods for the detection of falls using multiple cameras. In the early fusion method a 3D reconstruction is obtained using calibrated cameras, whereas in the late fusion method every camera detects falls individually and a final voting step leads to the overall decision.

In the future, prototype installations will show the real challenges of the very 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. Special care has to be taken towards the arrangement of the cameras in the rooms, especially for a late fusion approach since the derived features are not invariant to the viewing direction. Future research will also go into this direction, e.g. to determine the necessary number of cameras and their optimal placing.

Acknowledgements

This work has been funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) and the Austrian Research Promotion Agency (FFG).

References and Notes

1. The Life of Women and Men in Europe : A Statistical Portrait. *Eurostat* **2002**.
2. D. Wild, U.S. Nayak, and B. Isaacs. How dangerous are falls in old people at home? *Br Med J*, **1981**, 282, 266–268.
3. J. Willems, G. Debard, B. Bonroy, B. Vanrumste, and T. Goedemé. How to detect human fall in video? An overview. *International Conference on Positioning and Context-Awareness*, **2009,** to be published.

4. D. Anderson, J.M. Keller, M. Skubic, X. Chen, and Z. He. Recognizing falls from silhouettes. *Proc. of International Conference of the Engineering in Medicine and Biology Society*, **2006**, 6388–6391.
5. R. Cucchiara, C. Grana, A. Prati, and R. Vezzani. Probabilistic posture classification for human behavior analysis. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, **2005**, 35(1), 42–54.
6. B.U. Töreyin, Y. Dedeoglu, and A.E. Cetin. HMM based falling person detection using both audio and video. *Lecture Notes in Computer Science*, **2005**, 3766, 211–220.
7. J. Tao, M. Turjo, M.F. Wong, M. Wang, and Y.P. Tan. Fall incidents detection for intelligent video surveillance. *Proc. of Fifth International Conference on Information, Communications and Signal Processing*, **2005**, 1590–1594.
8. C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Fall detection from human shape and motion history using video surveillance. *Proc. of 21st International Conference on Advanced Information Networking and Applications Workshops*, **2007**, 2, 875–880.
9. C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Monocular 3D head tracking to detect falls of elderly people. *Proc. of International Conference of the Engineering in Medicine and Biology Society*, **2006,** 6384–6387.
10. C.W. Lin, Z.H. Ling, Y.C. Chang, and C.J. Kuo. Compressed-domain Fall Incident Detection for Intelligent Homecare. *Journal of VLSI Signal Processing*, **2007**, 49(3), 393–408.
11. C. Huang, E. Chen, and P. Chung. Fall detection using modular neural networks with backprojected optical flow. *Biomedical Engineering: Applications, Basis and Communications*, **2007**, 19(6), 415–424.
12. 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. *Computer Vision and Image Understanding*, **2009**, 113(1), 80–89.
13. L.A. Zadeh. Fuzzy sets. *Information and control*, **1965**, 8(3), 338–353.
14. C.R. Dyer. Volumetric scene reconstruction from multiple views. *Foundations of Image Understanding*, **2001**, 469–489.
15. N. Noury, A. Fleury, P. Rumeau, AK Bourke, GO Laighin, V. Rialle, and JE Lundy. Fall detection–Principles and methods. *Proc. of International Conference of the Engineering in Medicine and Biology Society*, **2007,** 1663–1666.
16. H. Nait-Charif and S.J. McKenna. Activity summarisation and fall detection in a supportive home environment. *Proc. of International Conference on Proceedings of the Pattern Recognition*, **2004**, 4, 323–326.