Audiovisual Assistance for the Elderly - An Overview of the FEARLESS project*

Rainer Planinc, Martin Kampel and Sebastian Zambanini

Vienna University of Technology, Computer Vision Lab,
Favoritenstr. 9/183-2, A-1040 Vienna, Austria
{rplaninc,kampel,zamba}@caa.tuwien.ac.at
http://www.caa.tuwien.ac.at/cvl

Abstract. This paper gives an overview of the recently granted AAL-JP project FEARLESS which stands for "Fear Elimination As Resolution for Loosing Elderly's Substantial Sorrows". The proposed project aims to reduce elderly's fears within their homes. As elderly potentially refuse or forget to wear any additional sensors to activate alarm calls, FEARLESS will visually and acoustically detect and handle risks by contacting the relatives or care taker organization automatically - without the need of any user intervention. This is done by using only one single type of sensor making the system affordable for everyone. It increases the feeling of safety, reduces fears, enhances the self-efficacy and thus enables elderly to be more active, independent and mobile in today's self-serve society.

Keywords: ambient assisted living; automatic risk detection; elderly;

1 Introduction

Emergency systems for elderly contain at least one sensor (button or accelerometer) which has to be worn or pressed in case of emergency. These emergency call buttons are provided by care taker organizations having the main drawback that no information about an occurred incident prior the button is pressed is available. Moreover, people have to wear these buttons which they tend to forget or even refuse. In case of an emergency and if elderly are able to press the button, they have to tell the operator which kind of incident happened. If the elderly is not able to talk to the operator for any reason, there is no information about the type of incident. This causes false alarms as well as ambulance deployments, although there is no emergency situation at all. To ensure the detection of emergency situations where the elderly is not able to actively raise an alarm (e.g. due to the lost of consciousness), sensors acting autonomously are needed.

Autonomously acting sensors are used in the field of smart homes to fulfill core functions defined in [6]: the control of the system, emergency help, water and energy monitoring, automatic lighting, door surveillance, cooker safety, etc.. Due to various reasons summarized in [2], smart homes are not established yet.

* This work was supported by the European Union under grant AAL 2010-3-020.

One of the reasons mentioned in [2] are the costs: it is easier and less expensive to integrate smart home technology into new buildings than it is for already existing buildings. This results in the demand of a robust system, which can be integrated into existing buildings. Moreover, one of the outcome of the former project MuBisA [17] is that elderly accept technical assistance only if the system is not discernible for third persons (i.e. visitors).

Considering these facts, a computer vision approach is feasible as it is able to overcome the limitations of other sensor types [11]. Furthermore, not only falls can be detected but also other events where help is needed (e.g. fire, flooding,. . .). By the use of a vision based system the detection of emergency situations is done by software, meaning that this system is extendable as only the respective algorithms need to be developed or adopted. A wide variety of computer vision algorithms for different applications exist (e.g. [3, 4, 14, 15]), but there is no "perfect" algorithm for detecting emergencies in elderly's homes yet.

As falls are considered to be a major risk for elderly, there has been done research on automatic fall detection [12]. Not only the fall itself but also the consequences of a fall are a great risk for elderly. Noury et al. [13] have shown that getting help quickly after a fall reduces the risk of death by over 80% and the risk of hospilization by 26%. Hence, FEARLESS is able to provide emergency service – if needed – immediately.

2 Event Detection

The goal of this project is to detect a wide range of risks with a single sensor unit, enhancing mobility and enabling elderly to take active part in the self-serve society by reducing their fears. In order to detect reduced mobility, a long-term tracking of the elderly is considered in the FEARLESS system. Another main focus of this project is the lack of expertise at the supplier side and thus the integration of important parts of the supply chain (i.e. network of electricians and electric shops). To ensure a holistic approach, the project consortium does not only consist of technical members, but also of psychologists, medical scientists and companies being able to transfer the knowledge from research to economy. Furthermore, the integration of end-user organizations ensures that end-user's wishes and needs are considered throughout the whole project.

The FEARLESS system uses cameras equipped with microphones as sensors, allowing for the combined visual and acoustic detection of risks. Furthermore, we make use of a late fusion approach developed during the former project MuBisA, performing analysis of the scene on each camera individually and then combining the individual results to get an overall decision [18], as shown in Figure 1. In contrast to other works (e.g. [1]), our system is not vulnerable to low-quality images, as only some basic information (e.g. silhouettes) are extracted from the image. We define empirical, semantic driven rules using features with fuzzy boundaries introduced in [5] to analyze the scene and make the decisions.

To be able to visually detect risks, the following steps are applied:

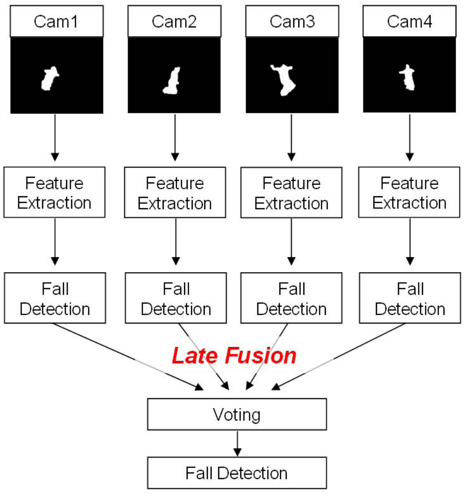

Fig. 1. Late Fusion for multi-camera fall detection [18]

1. **Motion detection:** First, motion detection is performed on the video to segment motion (e.g. the person) from the background. For this purpose, a robust background model has to be established which is able to adapt to changing conditions (e.g. lighting) as well as to reject motion in the background (e.g. a TV). A recently upcoming promising concept for background modeling is boosting [8] which permits the rejection of recurrent motion in the background during run-time without any presumptions. To increase robustness, color information is also exploited for shadow detection [9].
2. **Feature identification:** According to the different risks to be detected, a collection of various features is extracted. Fall detection requires features describing the human posture [4]. Specific actions can be detected by the use of space-time interest points [10]. Other events like smoke can be detected e.g. by using a wavelet transformation or dynamic texture change as feature. A dataset for evaluation of features is provided by [16].
3. **Risk detection:** Different risks (e.g. fall, fire, flooding,) are pre-defined to interpret the features and relate them to the risks by using confidence values. The final decision about the current risk for a single camera is made by a voting step which combines the individual confidences. By the use of multiple cameras, the overall robustness and reliability of the system is increased since the voting neglects individual wrong detections. Moreover, the problem of occlusions (e.g. by furniture) is implicitly solved.

To overcome the limitations of computer vision approaches, in the FEARLESS system risk detection also comprises the processing of audio data from microphones. This allows for combined audiovisual data processing, which is based not only on visual information but on the accompanying audio signal, and enhances the general performance of the risk detection. Due the smaller size and dimensionality, audio data is easier and faster to process than video data. Many of the events to be detected by the FEARLESS system are usually accompanied by characteristic audio sounds that could provide useful information for detection, e.g. the loud sound when accidentally falling on the floor followed by silence. Audio classification is a major application area of pattern recognition, i.e. the scientific discipline whose goal is the automatic classification of data patterns into a number of categories. A multi-stage approach is performed to recognize events from audio data, consisting of the following steps:

1. **Silence elimination:** audio is checked first for a processable signal, in order to prevent a further processing of audio containing silence only. This is achieved by comparing the audio power against a threshold value estimated from a long-term analysis for each microphone.
2. **Feature extraction:** audio signals can be represented in the time domain (time-amplitude representation) or the frequency domain (frequency-magnitude representation). Common features used are the average energy, zero crossing rate or silence ratio in the time domain, and bandwidth, energy distribution or harmonicity in the frequency domain.
3. **Audio pre-classification:** the audio data is classified into some common types of audio such as speech, sounds or noise. This is achieved by either using each feature individually in different classification steps or by using a set of features together as a vector to calculate the closeness of the input to the training sets.
4. **Final audio classification:** based on the output of the previous step, the different audio types are further processed in different ways. For speech recognition, techniques based on Hidden Markov Models [7] are applied as they are currently the most widely used and produce the best recognition performance. For sound recognition, Dynamic Time Warping and Artificial Neural Networks have shown promising results in the past.

3 Conclusion

In this paper an overview of the FEARLESS project was given. Due to elderly's reluctance or forgetfulness to wear any devices, the FEARLESS project is designed to detect different events using audiovisual pattern recognition algorithms automatically. By providing safety, elderly will suffer from less fears and thus being more active in todays self-serve society. To ensure a holistic view, the project consortium consists of partners from different disciplines and thus not only technical aspects are covered. Finally, the integration of the supply chain and also elderly throughout the project is crucial. At a later stage of the project, field tests and pilot phases will be conducted to validate this approach.

References

1. Aghajan, H., Wu, C., and Kleihorst R.: Distributed Vision Networks for Human Pose Analysis. Signal Processing Techniques for Knowledge Extraction and Information Fusion, 181–200 (2008)

2. Aldrich, F.: Smart Homes: Past, Present and Future. In: Harper, R. (ed.) Inside the Smart Home, pp. 17–39, Springer, London (2003)
3. Anderson, D., Keller, J.M., Skubic, M., Chen, X., and He, Z.: Recognizing falls from silhouettes. In: Proc. of the 28th Annual Int. Conf. on Engineering in Med. and Biology Society, pp. 6388–6391 (2006)
4. Anderson, D., Luke, R.H., Keller, J.M., Skubic, M., Rantz, M. and Aud, M.: Linguistic Summarization of Video for Fall Detection Using Voxel Person and Fuzzy Logic. Computer Vision and Image Understanding 113, 80–89 (2009)
5. Doulamis, A.D., Doulamis, N.D. and Kollias, S.D.: A fuzzy video content representation for video summarization and content-based retrieval. Signal Processing, 80(6):1049-1067 (2000)
6. Fisk, M.J.: The implication of smart home technologies. In: Peace, S.M. and Holland, C. (eds.) Inclusive housing in an ageing society: innovative approaches, pp. 101–124, Policy Press, Bristol (2001)
7. Gales, M. and Young, S.: The Application of Hidden Markov Models in Speech Recognition. Foundations and Trends in Signal Processing 1(3), 195–304 (2008)
8. Grabner, H., Leistner, C. and Bischof, H.: Time Dependent On-line Boosting for Robust Background Modeling. In: Proc. of the Int. Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 612–618 (2008)
9. Kampel, M., Wildenauer, H., Blauensteiner, P. and Hanbury, A.: Improved motion segmentation based on shadow detection Electronic Letters on Computer Vision and Image Analysis 6(3), 1–12 (2007)
10. Laptev, I., Caputo, B., Schüldt, B. and Lindeberg, T.: Local velocity-adapted motion events for spatio-temporal recognition. Computer Vision and Image Understanding 108(3), 207–229 (2007)
11. Mihailidis, A. and Carmichael, B. and Boger, J.: The Use of Computer Vision in an Intelligent Environment to Support Aging-in-Place, Safety, and Independence in the Home. IEEE Transactions on Information Technology in Biomed. 3, 238–247 (2004)
12. Noury, N., Fleury, A., Rumeau, P., Bourke, A.K., Laighin G., Rialle, V. and Lundy, J.E.: Fall detection Principles and Methods. In: Proc. of the 29th Int. Conf. of Engineering in Med. and Biology Society, pp. 1663–1666 (2007)
13. Noury, N., Rumeau, P., Bourke, A.K., Laighin, G. and Lundy, J.E.: A Proposal for the Classification and Evaluation of Fall Detectors. BioMedical Engineering and Research 29, 340–349 (2008)
14. Rougier, C., Meunier, J., St-Arnaud, A. and Rousseau, J.: Monocular 3D head tracking to detect falls of elderly people. In: Proc. of the 28th Annual Int. Conf. on Engineering in Med. and Biology Society, pp. 6384–6387 (2006)
15. Rougier, C., Meunier, J., St-Arnaud, A. and Rousseau, J.: Fall Detection from Human Shape and Motion History Using Video Surveillance. In: Proc. of the 21st Int. Conf. on Advanced Information Networking and Applications Workshops, pp. 875–880 (2007)
16. Stöttinger, J., Zambanini, S., Khan, R.U. and Hanbury, A.: FeEval - A Dataset for Evaluation of Spatio-temporal Local Features. In: Proc. of the 20th Int. Conf. on Pattern Recognition, pp. 499–502 (2010)
17. Zambanini, S., Kampel, M. and Machajdik, J.: Computer Vision for an Independent Lifestyle of the Elderly and Disabled - An Overview of the MuBisA Project. In: Proc. of the AALIANCE Conf. (2010)
18. Zambanini, S., Machajdik, J. and Kampel M.: Early versus Late Fusion in a Multiple Camera Network for Fall Detection. In: Proc. of the 34th Annual Workshop of the Austrian Association f. Pattern Recognition, pp. 15–22 (2010)