Visual Fall Detection in Home Environments

J. Spehr, M. Gövercin, S. Winkelbach, E. Steinhagen-Thiessen, and F. M. Wahl

***Abstract*— One important aim of intelligent environments is to allow elderly people to stay as long as possible in their home environment. Fall detection is an essential part of intelligent environments since falls are one of the most distinctive factors that threaten the independence of elderly people. The fall detection system proposed in this paper is based on cameras that are equipped with wide angle lenses which can acquire a complete view of the living space. Our approach uses a novel background subtraction technique to detect the human body in the camera image. The novel technique is especially designed for home environments and can adapt even to strong changes in the scene (illumination changes, open/closed doors) very fast. The orientation of the detected human body is then analyzed to detect falls. First results of our feasibility study show that the visual fall detection system is usable in detecting falls in home like environments of adults and elderly people.**

I. Introduction

FALLS are one of the most important factors that threaten the independence of elderly in our civilization. Thirty per cent of individuals aged 65 and older fall at least one time per annum. This number rises looking at individuals >80 up to 80% p.a. In long term care setting, 50% of habitants experience at least one fall each year [1]. For 20% of elderly, a fall results in high-maintenance and admission to a nursing home. More than 20% of falls result in an injury including a high rate of fractures and even a significant mortality [2]. The cost of fall related injuries for individuals older than 65 y. crosses € 1 billion in Germany and $12.6 billion in USA [3]. Recent developments in Information and Communication Technologies (ICT), especially miniaturization and decreased costs, make new technologies suitable in the prevention and detection of falls and fall related injuries in home environments for a greater audience.

Intelligent environments like smart-homes are one of the key technologies that will provide efficient and cost-effective solutions. One of the main challenges in intelligent environments is the right choice of the sensor technology. Sensors used in intelligent environments vary from simple binary sensors [4] to more sophisticated devices like ultrasonic sensors [5] or accelerometer sensors [6]. Most of these technologies suffer from the following problems: a large number of sensors is needed, active components have to be installed, or the occupant has to wear a device. Vision-based sensors, like standard cameras, are a promising alternative due to their non-intrusive technique, their low-price and the availability of image processing hardware.

West et al. introduce a vision-based activity recognition system [7], where the image of a ceiling mounted camera is divided into polygonal regions, so called "virtual sensors". Each virtual sensor defines a device in the image like the stove or fridge. Thus, activities such as being near, passing or using a device can be detected. However, other important situations like falls remain undetected.

A similar camera setup was used by Nait-Charif et al. to detect falls [8]. The image is divided into "inactivity zones" like the bed region and "entry zones" like the door. Falls are defined as inactivities in a "non-inactivity zone". Therefore, general anomalies especially those in an "inactivity zone" cannot be detected. Furthermore, these proposals use standard foreground-background segmentation, which is inappropriate for long-term applications in real world environments.

The following method overcomes limitations of previous approaches. Our system is based on one or more cameras that are equipped with wide angle lenses and can acquire the complete living space. The person's pose is recognized within a camera image using a novel hybrid background subtraction technique, which is specially designed for home environments. Our approach uses standard background subtraction in the normalized YUV color space to detect a person. The background model is efficiently updated using temporal differencing and high-level knowledge. Thus even strong changes in the scene (illumination changes, open/closed doors, changed furniture) can be updated very fast. Finally, the orientation of the detected human body is analyzed and used to detect falls.

The rest of the paper is organized as follows. First, we introduce our methods in Sec. II and III. Here, standard background subtraction is compared to hybrid background subtraction. Furthermore, we present our novel fall detection approach in Sec. III. We conclude with experimental results and a general discussion in Sec. IV and V.

Manuscript received May 07, 2008

J. Spehr, S. Winkelbach and F. M. Wahl are with the Institute for Robotics and Process Control of the Technische Universität Braunschweig, 38106 Braunschweig, Germany (e-mail: {J.Spehr, S.Winkelbach, F.Wahl}@tu-bs.de).

M. Gövercin, E. Steinhagen-Thiessen are with the research group Geriatrie of the Charité Universitätsmedizin Berlin, Germany (e-mail: {Mehmet.Goevercin, Elisabeth.Steinhagen-Thiessen}@charite.de)

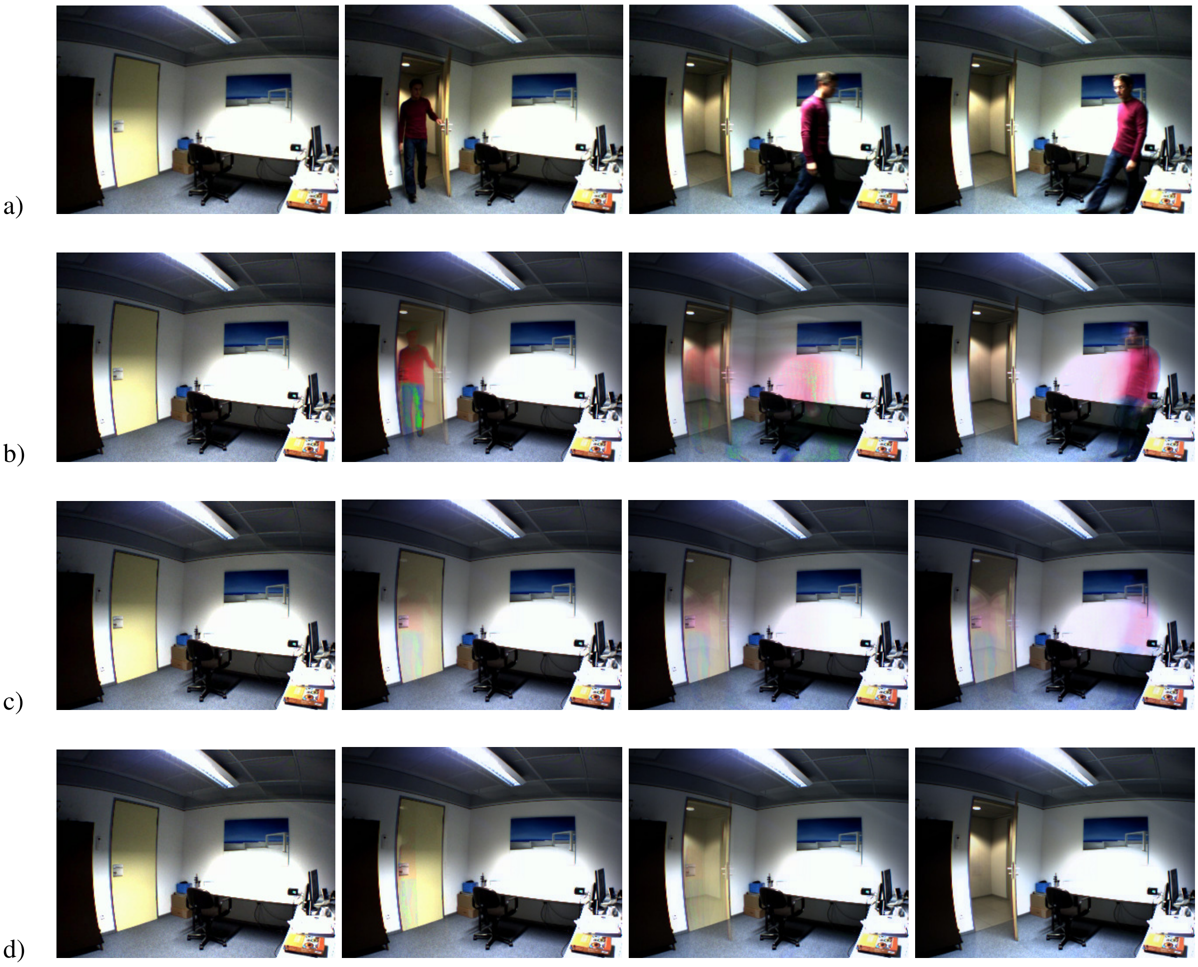

Fig. 1. Image sequence: Person entering a room. This sequence demonstrates two tasks of robust background subtraction. First, the background model has to be updated on the left image side since the door changes from closed to opened. Simultaneously, the updating scheme should avoid adaption on the right image side due to the motionless person. First column at time step 0, second at time step 33, third at time step 66 and fourth at time step 99; The first row shows the input images, the second and third shows the reference image calculated by an adaptive approach with $\alpha = 0.9$ resp. $\alpha = 0.995$. The fourth row shows the reference image updated by the proposed hybrid scheme. As can be seen in time step 99 the door region is correctly updated and the motionless foreground isn't adapted to the background model.

II. RECOGNITION OF THE HUMAN BODY

A. *Standard Background Subtraction*

In general recognition of a human body in an image is very challenging. However, in the case of a static camera and scene, background subtraction (BS) techniques can be applied. Consider an image sequence $I_0, I_1, \ldots, I_t$, where $I_t(x, y)$ is the value of point (x, y) in *RGB* color space and R_t is an image of the static scene without persons, also referred to as background model. Often statistical informations like the variance of pixel values are also included in the background model [9].

Background subtraction provides a foreground mask M_t using the image I_t and the background model R_t like follows:

$$M_t = \begin{cases} 1 & if \ \left| d(R_t, I_t) \right| > threshold \\ 0 & else \end{cases}$$

where d denotes a distance function. To reduce the influence of shadows, the distance function works in the normalized YUV-color space [10]. First, the *RGB* color values are transformed into the YUV-color space. Then, the chrominance information is normalized by the brightness, $U_n = U / Y$, and $V_n = V / Y$.

The main problem of BS is to estimate the background model R_t. Especially in home environments the background model has to be continuously updated due to changes in the scene like illumination changes, open/closed doors. Standard approaches uses e.g. median filter over time, average over time, or simple adaptive filter to update the background model. Adaptive approaches use the update scheme

$$R_t(x, y) = \alpha R_{t-1}(x, y) + (1 - \alpha) I_t(x, y)$$

where α controls the adaptive time. For low values ($\alpha \approx 0.5$) the reference image is adapted very fast to changes in the scene, resp. for high values ($\alpha \approx 0.99$) very slow. However, these approaches assume that the foreground object (i.e. the person) is continuously moving.

Fig. 2. Images captured with a camera mounted at the ceiling of a room. The camera is equipped with a wide angle lens that can acquire a complete view of the living space with one image. Image (a) shows a subject standing in the room. The orientation of the human body points to the image center. Image (b) shows a person lying on the floor. The orientation of the human body does not point to the image center and thus indicates a fall.

A motionless person quickly becomes part of the background model. Fig. 1 (b) and (c) shows the reference image results for $\alpha = 0.9$ and $\alpha = 0.995$.

B. Hybrid Background Subtraction

Recent approaches suggest using high-level information to solve the problem of updating the background model and avoiding adaptation of motionless foreground objects. In [11] explicit knowledge of different object categories is added to improve BS. Other approaches like [12] go one step further and integrate high-level information of a blob tracking system into the background updating process. Unfortunately, tracking systems cannot guarantee to avoid loss of the tracked object. In this paper we propose a hybrid BS technique, which overcomes this limitation. The aim of our proposal is a fast and accurate update of the background model; simultaneously we want to avoid an adaptation of motionless foreground objects. In order to accomplish these objectives we have to know the mask M_t, since this mask defines those pixels that are element of the foreground and thus should not be updated. This may be considered as chicken and egg problem. Because for updating the background model we need to know the foreground mask, but to get the foreground mask we need to have the updated background model.

We propose an estimation of the foreground mask D_t using temporal differencing and high-level knowledge. The main advantage of temporal differencing is, that it only uses the two last images I_t and I_{t-1} to define the foreground mask.

$$D_t = \begin{cases} 1 & \text{if } \left| d\left(I_{t-1}, I_t\right) \right| > threshold \\ 0 & else \end{cases}$$

In general, temporal differencing of a moving object results in a silhouette-like foreground mask. Using high-level knowledge about the appearance of the object, this silhouette-like foreground mask can be used to estimate the real foreground mask M_t. The aim is to reconstruct the foreground mask M_t based on the temporal differencing result and high-level knowledge about the human body model.

Due to the computational cost of a sophisticated human body model we approximate the pose of the human body by a 2d-gaussian function:

$$N_t(x, y) = \exp\left(-\frac{1}{2} (\vec{x} - \vec{\mu}_t)^T K_t^{-1} (\vec{x} - \vec{\mu}_t) \right),$$

with $\vec{x} = (x, y)$, $\vec{\mu}_t$ corresponds to the center of the human body and K_t is the covariance matrix. The value N_t is the probability, that the point is element of the foreground model, ranging from *0* to *1*, where *0* denotes that the point belongs to the background and *1* that it belongs to the foreground. We can simply approximate the center $\vec{\mu}_t$ and the covariance K_t at time step *t* by

$$\vec{\mu}_t = E[\vec{x}]$$
$$K_t = E\left[(\vec{x} - \vec{\mu}_t)(\vec{x} - \vec{\mu}_t)^T \right],$$

where *E* denotes the expectation of all $\vec{x}$ being element of the foreground mask D_t.

Using these model informations we are now able to update the background model in the following manner:

$$R_t(x, y) = \alpha N_t(x, y) R_{t-1}(x, y) + (1 - \alpha N_t(x, y)) I_t(x, y)$$

Fig. 1 (d) shows the results of the algorithm applied to an

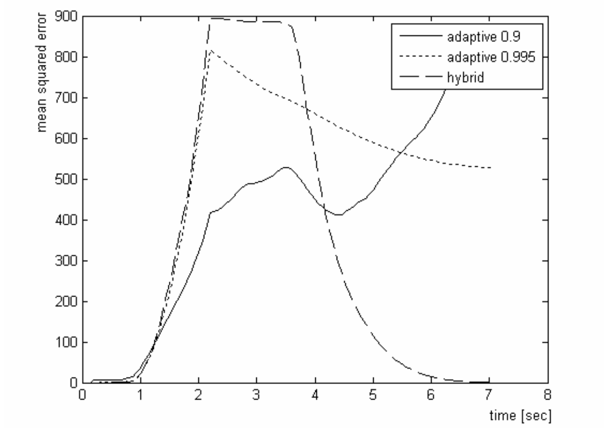

Fig. 3. Mean squared errors (MSE) of the reference images. The reference images are generated by the different update approaches for the image sequence shown in Fig. 1. The adaptive approaches update the motionless foreground object and cause a high MSE after 4.5sec.

image sequence. Since the updating uses the advantages of background subtraction and temporal differencing in a hybrid manner, we will refer to this technique in the following as hybrid background subtraction.

III. DETECTION OF FALLS

Recognition of the human body in an image sequence is a very complex task due to the many degrees of freedom of the human body. Furthermore, we have to deal with the following problems: resolution of the human body in the image is low, on-line recognition is required, arbitrary clothing and variable light conditions. A detailed overview over commonly used human motion capture systems can be found in [13] and [14].

We are using cameras with wide angle lenses that can acquire a complete view of the living space with just one image (See Fig. 2). The cameras are mounted at the ceiling of the room with vertical oriented optical axes, so that the body axes of a standing person is parallel to the optical axes of the camera. A consequence of this setup is that the orientation of a standing human body will always point to the center of the image since the image center is some kind of a vanishing point. From this it follows that the orientation of the human body is an important information to distinguish a standing and a lying person.

The orientation of the human body is calculated with central moments that are defined as:

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q M(x, y)$$

where $\bar{x}$ and $\bar{y}$ represents the center of mass. With the central moments of second order the orientation can be determined as:

$$\gamma = \frac{1}{2} \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right).$$

As mentioned before the orientation of a standing person should point to the center of the image (c_x, c_y). Thus, the difference

$$\delta = \left| \gamma - \arctan\left(\frac{\bar{y} - c_y}{\bar{x} - c_x} \right) \right|$$

can be used to detected lying persons.

IV. RESULTS

In our feasibility study we investigate different setups concerning age and appearance (e.g. dark and bright clothes). We tested the proposed approach with a single camera, which was mounted in the center of the ceiling in our home like environment 'living lab'. The implementation of the proposed fall detection is running on an AMD 2.2 GHz computer with 33.7fps at a resolution of 320×240. We are using hybrid background subtraction to handle difficult environment conditions. Fig. 3 shows the mean squared error (MSE) of the different update approaches for the image sequence shown in Fig.1. Between 1.2sec and 3.2sec the person is entering the room. The ground truth is manually generated. As can be seen in the diagram only hybrid background subtraction is able to adapt to the new background model. The adaptive approaches update the motionless foreground object and cause a high MSE at 7 sec.

In order to evaluate our fall detection system Fig. 4 shows the development of δ, where a fall occurs after 17 sec. As can be seen in the diagram, the increasing difference angle δ indicates the fall. We found empirically that a difference angle of $\delta \geq 28°$ can be used to detect falls resp. lying persons.

Using this threshold the fall detection system has a false positive rate $\alpha = 0.31$ and a false negative rate $\beta = 0.22$. False positives are mainly caused in situations, where the persons are lying in bed or sitting on sofa. Since we are detecting lying persons as a consequence of a fall, we cannot distinguish between these situations. A solution to this problem could be manually masking these regions in the image and neglecting falls occurring in the mask regions.

False negatives depend on the fall direction. Since we can only detected falls with $\delta \geq 28°$, falls with a smaller difference angle remain undetected. This is a general problem of a monocular camera setup and cannot be solved without using more than one camera.

V. CONCLUSION

The results of the feasibility study show that the visual fall detection system is usable in detecting falls in home like environments of adults and elderly people. The visual fall detection system is able to detect falls even under difficult environment conditions due to the novel hybrid background subtraction technique that overcomes limitations of previous approaches. Our method takes advantage of high-level information in order to estimate the foreground mask and update the reference image.

Further investigation is needed to give answers to the

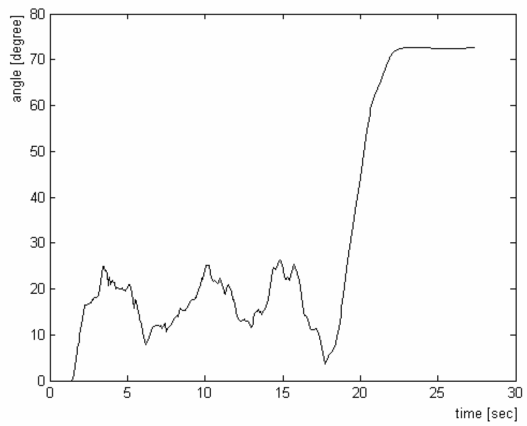

Fig. 4. Development of the difference angle δ in an image sequence of a going person. After 17 sec. a fall occurs. In the diagram this is indicated by the increasing graph.

question of usability in home environments and effectiveness of preventing falls including cost efficiency.

In future work we will expand the tests of our approach on image sequences acquired in real home environments. Especially long-term analyses will hopefully demonstrate the significance of computer vision as a new key technology for intelligent environments and home care systems.

References

[1] D. Oliver, F. Daly,F. C. Martin, M. E. McMurdo. Risk factors and risk assessment tools for falls in hospital in-patients: a systematic review. *Age Ageing*. Mar 2004, 33(2), pp. 122-130.

[2] W. von Renteln-Kruse, T. Krause. "Fall events in geriatric hospital in-patients. Results of prospective recording over a 3 year period". *Z Gerontol Geriatr.* Feb 2004; 37(1), pp. 9-14.

[3] M. B. King, M. E. Tinetti. "Falls in community-dwelling older persons," *J. Am. Geriatr. Soc.,* Oct 1995; 43(10):1146-1154.

[4] E. M. Tapia, T. S. S. Intille, K. Larson, "Activity recognition in the home setting using simple and ubiquitous sensors," *Proceedings of Pervasive (2004),* pp. 158-175

[5] J. Russo, A. Sukajo, A. Helal, R. Davenport, W . C. Mann, "Smart Wave – Intelligent meal preparation system to help older people live independently*," International Conference on Smart Homes and Health Telematics* , 2005, pp. 122 - 135

[6] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, B. G. Celler,, „Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring," *IEEE Transaction. Inf. Technol. Biomed.* ,10 (1), 2006, pp. 156-167

[7] G. West, C. Newman, S. Greenhill, "Using a Camera to Implement Virtual Sensors in a Smart House," *From Smart Homes to Smart Care: International Conference on Smart Homes and Health Telematics*, 15 (1) , 2005, pp. 83-90.

[8] H. Nait-Charif, S. J. McKenna, "Activity Summarisation and Fall Detection in a Supportive Home Environment,," *Pattern Anal. Appl. 7 (4)* , 2004, pp. 386-401.

[9] C. Stauffer, W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 1999, 246-252

[10] C. R. Wren, A. Azarbayejani, T. Darrell, A. Pentland, „ Pfinder: Real-Time Tracking of the Human Body*," IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7)*, 1997, pp. 780-785.

[11] R. Cucchiara, C.Grana, M. Picardi, A. Prati, „Detection Moving Objects, Ghosts, and Shadows in Video Streams*," IEEE Transactions on Pattern Analysis and Machine Intelligence 25* (10), 2003, pp. 1337-1342.

[12] L. Taycher, J.W. Fischer III, T. Darrall,, „Combining object and feature dynamics in probabilistic tracking*," Computer Vision and Image Understanding 108*, 2007, pp.243-260.

[13] T.B. Moeslund, E. Granum, "A Survey of Computer Vision-Based Human Motion Capture," International Journal of Computer Vision and Image Understanding 81(3), 2001.

[14] T.B. Moeslund, A. Hilton, V. Krüger, "A Survey of Advances in Visoin-Based Human Motion Capture and Analysis", *International Journal of Computer Vision and Image Understanding 104(2-3)*, 2006.