

Video Surveillance for Behaviour Monitoring in Home Health Care

Ben J.A. Kröse^{1,2}, Tim van Oosterhout¹, Gwenn Englebienne²

¹*Digital Life Centre, Amsterdam University of Applied Science, Amsterdam, the Netherlands
{B.J.A.Krose, T.J.M.van.Oosterhout}@hva.nl*

²*Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands
G.Englebienne@uva.nl*

Introduction

As the number of elderly needing assistance in their daily life increases, sensing systems that monitor behaviour are becoming ever more important. The type of the behaviors that are to be monitored varies enormously, from nutritional intake, physical activities, cognitive activities, fall detection, wandering detection or sleep behaviour. Also the types of sensing systems vary accordingly. Wearable sensors are often used for measuring physical activities and for alarming. However, these sensors require special attention from the user, in terms of battery replacement and compliance. Ambient sensors are more reliable in that sense, but result in another trade-off. Simple sensors such as motion detectors and pressure mats are non-obtrusive but are less accurate in their measurements. Advanced imaging sensors give much more accurate data but come with the problem of privacy. In the Balance-IT project, we specifically studied the applicability of imaging sensors in health care applications, specifically fall detection, wandering detection, remote physical therapy and fitness.

Falls are a major source of injury for elderly people (Gallagher et al., 2001). A lot of research has gone into automated solutions. Various fall detection methods using standard cameras have been presented, where rapid changes in the modelled body shape (Foroughi et al., 2009; Yu et al., 2009) or estimated pose (Liu et al., 2010) or bounding volume (Tao et al., 2005; Anderson et al., 2009) are used to detect falls. A major issue with these approaches is that how much the perceived body shape changes depend both on the camera location and the direction of the fall. This is not an issue in depth cameras, since they measure the absolute size of the body shape, and not the size of the projection of the body onto a 2D plane. Several methods have investigated the use of depth cameras, the Kinect in particular (Rougier et al., 2011; Mastorakis and Makris, 2012). Yet the body pose estimation of 3D sensors may fail dramatically in particular situations, such as in bright sunlight or when the person is partially occluded. In (Josemans et al., 2013), we investigated how combining 2D and 3D environmental cameras improves fall detection.

Wandering detection is a problem for people with dementia. Especially in nursing homes there is a need for automatic detection of wandering behaviour. In the Balance-IT project we focus on camera-based methods for wandering detection. Similarly to fall detection, it requires the precise detection of the patient's location in the environment, but it cannot be detected from short-term features: instead, longer-term tracking of the person is required and the tracks themselves are indicative of wandering. In (Martino-Salzman et al., 1991), wandering is distinguished from efficient ("direct") travel based on the path the patient takes, and is further subdivided into random motion, pacing (walking back and forth between two points) and lapping (walking in circles). (Vuong et al., 2011) directly implemented a rule-based classification algorithm to detect the 4 types of motion ("direct", "pacing", "lapping" or "random") and obtained promising results on a limited data set.

Although the current scientific trend is focused on using computer vision for fall- and wandering detection, such systems are not yet deployed at large scale. Probably a combination of costs, non-robustness and privacy issues play a role. Recent commercial applications also show a trend towards the use of the smartphone for this (Igal et al, 2013). In this paper we present two of the four Balance-IT projects that focused on computer vision: fall detection by RGB-D cameras and wandering detection with off-the-shelf video cameras.

Fall Detection

In practical applications, computer vision systems need to deal with the cameras' limited field of view, changing light conditions, occlusions, and varying projections. Active 3D sensors (RGB-D cameras) such as the Microsoft Kinect, provide a partial solution to both changing light conditions (by being active) and varying projections (by providing depth information), but are still subject to the sensor's limited field of view and to random occlusions. By combining multiple cameras looking at the scene from a different vantage point, we can improve on any single camera's detection results. In particular, the Kinect's pose estimation can fail quite dramatically when the subject is occluded, or when the subject's pose is unusual (e.g., lying down). In (Josemans et al., 2013), we investigated how the data from skeleton tracking could be fused with the input from a conventional camera for improved fall detection.



Figure 1. Skeleton as derived from Kinect depth image. Figure 2. Bounding ellips extracted from foreground segmentation

Our approach is based on combining data from a Kinect camera and a wide angle camera mounted on the ceiling. For the Kinect system we used the skeleton extracted from the depth image, see Figure 1. The 60D vector (20 joints) is reduced by a Principal Component Analysis (PCA). For the overhead image we fit an ellipse to the detected foreground from the conventional camera, as illustrated in Figure 2, and create a feature vector that includes the ratio between the ellipse's width and height and its location within the camera's field of view. First we tested a naïve fusion technique by just concatenating the two feature vectors. The results of the individual and combined features are given in Table 1. As can be seen this naïve fusion approach does not yield improved results over individual features. Our approach, shown in Figure 3, validates the Kinect output with the conventional camera. To do this, we re-project the skeleton found by the Kinect into conventional camera's image. We then compute the overlap between the detected foreground in that camera and the projected skeleton, resulting in a "Skeleton Match Score", and feed this as an additional feature to our classifier. By using a non-linear Support-Vector Machine with RBF kernel, we obtain a classifier that correctly weighs the most informative features, i.e. the skeleton features when the skeleton is validated, and the conventional camera's features when it is not.

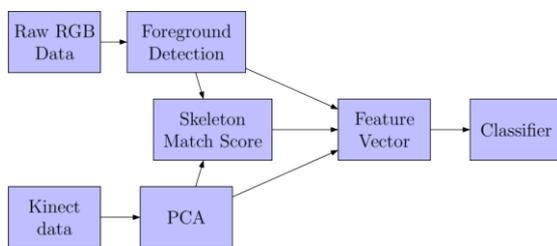


Figure 3: approach showing the skeleton match score

	TP	TN	FP	FN
Skeleton based	38.76	37.04	2.96	1.24
Bounding ellipse	36.28	40.00	0.00	3.72
Using all features	36.84	36.80	3.20	3.16
Using also skeleton match score	39.36	39.48	0.52	0.64

Table 1: Performance of the four methods. TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negatives

We recorded falls of students in our lab. The recorded falls all take place in an area of roughly 3x3 meters, and the direction of the fall is varied. In total, 40 fragments of about 5 minutes were recorded. Each fragment contains mostly general activity with a fall at the end. These falls were manually annotated.

Our results, shown in Table 1, with 4-fold cross-validation on 25 runs using random permutations of a dataset containing 40 falls and 40 non-fall sequences show that this approach substantially outperforms our baselines: the two methods based on a single sensor and the naive fusion method. In fact, only one fall sequence is not well recognized. In this sequence, the subject falls largely outside both cameras' field of view, so that both the skeleton-based and the conventional camera-based features are misleading.

Wandering detection

A second application area in the Balance-IT project is wandering detection. In contrast to fall detection, which focuses on an accurate estimate of the human pose, here we focus on an accurate estimate of the track: firstly to distinguish people with dementia from other people in the home, secondly to raise an alarm if the person with dementia is walking in an unwanted area (near staircase or other persons' apartments). We focused on a single camera.

We implemented person detection with HOG (Histogram of Oriented Gradients) features, which has been successfully applied to pedestrian detection. We found that most methods suffer from false positive detections when presented with a cluttered environment. To solve that, we developed a method that uses knowledge about the spatial structure of the environment. Many false positives appear in the wrong location, e.g. inside walls, or at the wrong scale, e.g. too small for the surroundings or with the feet not touching the ground (Figure 4). Furthermore, the range of searched scales must be limited in order to limit the computational complexity, resulting in missed detections either very close to or far away from the camera.

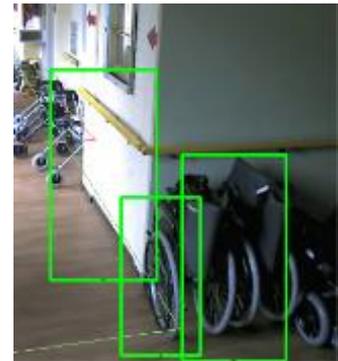


Figure 4. False positives in using HOG features

For an overview of the wandering detection see Figure 5. Using well established camera calibration techniques we obtain parameters for the camera's lens and position. In addition we define one or more areas on the ground that combined become our region of interest. Only this area is searched for the presence of people.

Traditionally a Support Vector Machine (SVM) classifier is trained on the HOG feature vectors of upright people. Then, an image is transformed into HOG features after which the trained feature vector, a window of fixed size, is swept across the transformed image. The process is repeated for transformations of the image at different scales. The result consists of detections across the whole image at different scales. Some position/scale combinations could have been eliminated beforehand considering the expected size of people and other constraints implied by perspective distortion and other environmental knowledge. We improve on this process by limiting the search area and scales using the manually defined region of interest and camera calibration.

For a certain distance to the camera an average person height can be computed in image space. Equidistant points form horizontal lines in the image. Therefore, an area containing several people standing at the same distance from the camera can be contained in a rectangle. If we determine a set of these rectangles, the widths, horizontal and vertical positions of which shall be determined by our region of interest and the height by person height and vertical position, we obtain a constrained set of search locations, all of which will result in people at the correct size and expected locations considering the information contained in the camera calibration and region of interest definition. If we run the HOG feature detector at a fixed scaled determined by the rectangle height for each such rectangle instead of at different scales across the entire image, the result is not only a higher true positive rate but also a lower computation time.

The detections from the method are tracked over time by a single hypothesis multi target tracker which uses Kalman filters for position extrapolation and smoothing. These tracks are classified by a rule based system on the basis of duration, covered distance and maximum speed into a walking or wandering class (pacing, lapping and random walks will all result in a change of these properties). Furthermore, regions within the region of interest determine which areas are safe and for which areas an alert should be generated. With this addition an alert can be generated for a wandering pattern that moves towards an alert region even before the person reaches

it (Figure 6). Additionally alerts can be omitted if necessary for tracks originating from an alert region and moving away from it. We compared our method to a traditional multiscale HOG feature based people detector as implemented in OpenCV 2.4.8. On the same hardware our method ran 7 to 8 times faster while generating less false positives and less false negatives. Our method was not susceptible to certain patterns in the background generating consistent false positives in the compared method, while being more sensitive to people appearing at smaller scales because the search space could include finer scale differences at further distances. Finally, because of these more reliable detections our method caused less split tracks and less tracks consisting entirely or for a significant portion of false detections.

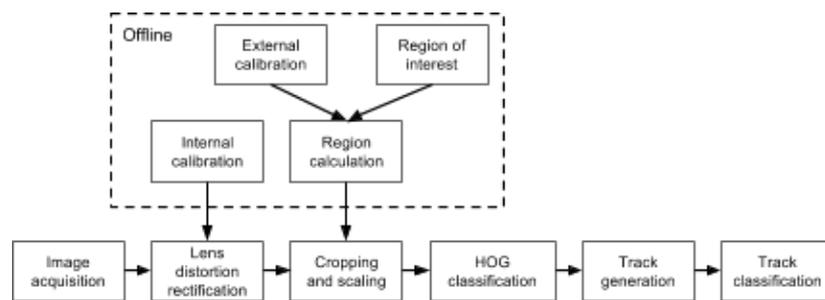


Figure 5: Set-up for the wandering classification

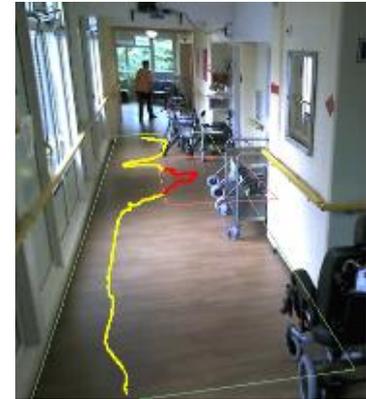


Figure 6: Results of our method

Conclusions and discussion

Automatic monitoring of patients in a nursing home leads to important improvements in the patients' health and quality of life. Proper fall detection leads to improved response times, which in turn lead to dramatic improvements in the prognosis of recovery. Wandering itself can lead to distress, and is an important symptom of patients with dementia. A pragmatic approach is therefore required: for wandering detection fewer sensors can be used, meaning that a larger area can be monitored. For fall detection, which requires more precise pose estimation and where high recognition rates and few false positives are critical, a correct combination of sensors should be applied.

In the Balance-IT project we noticed that camera surveillance is always associated with privacy issues. In doing the experiments and gathering the data we were well aware of this. For the early work (presented in this paper) we used students to play the falls. In later work on fall detection that was carried out in a center for epileptic patients we used an actor (one of the care givers) who was extremely good in mimicking the epileptic attacks. For the wandering detection we had a consent of the board of clients (clientenraad) to make recordings of the people with dementia and of, the head of the nursing institute to record the employees.

Acknowledgement

This work has been funded by the Foundation Innovation Alliance (SIA - Stichting Innovatie Alliantie), in the framework of the Balance-IT project. We thank AMSTA for their cooperation.

References

- Gallagher, B., Corbett, E., Freeman, L., Riddoch-Kennedy, A., Miller, S., Smith, C., Radensky, L. and Zarrow, A. (2001) A Fall Prevention Program for the Home Environment, *Home Care Provider* 6, no. 5, 157–63.
- Gillespie, L. (2004). Preventing falls in elderly people: we need to target interventions at people most likely to benefit from them. *BMJ: British Medical Journal*, 328(7441), 653.

- Igual, R., Medrano, C., & Plaza, I. (2013). Challenges, issues and trends in fall detection systems. *Biomedical engineering online*, 12(1), 66.
- Josemans, W., Englebienne, G., and Kröse, B. J. (2013). Fusion of Color and Depth Camera Data for Robust Fall Detection. *Proceedings of VISAPP* (1) 608-613.
- Liu, C. L., Lee, C. H., & Lin, P. M. (2010). A fall detection system using k-nearest neighbor classifier. *Expert systems with applications*, 37(10), 7174-7181.
- Martino-Saltzman, D., Blasch, B. B., Morris, R. D., & McNeal, L. W. (1991). Travel behavior of nursing home residents perceived as wanderers and nonwanderers. *The Gerontologist*, 31(5), 666-672.
- Mastorakis, G., & Makris, D. (2012). Fall detection system using Kinect's infrared sensor. *Journal of Real-Time Image Processing*, 1-12.
- Rougier, C., Auvinet, E., Rousseau, J., Mignotte, M., & Meunier, J. (2011). Fall detection from depth map video sequences. In *Toward Useful Services for Elderly and People with Disabilities* (pp. 121-128). Springer Berlin Heidelberg.
- Tao, J., Turjo, M., Wong, M. F., Wang, M., & Tan, Y. P. (2005). Fall incidents detection for intelligent video surveillance. *Information, Communications and Signal Processing, 2005 Fifth International Conference on*, 1590-1594.
- Yu, X., Wang, X., Kittipanya-Ngam, P., Eng, H. L., & Cheong, L. F. (2009). Fall detection and alert for ageing-at-home of elderly. In *Ambient Assistive Health and Wellness Management in the Heart of the City* 209-216, Springer Berlin Heidelberg.
- Vuong, N. K., Chan, S., Lau, C. T., & Lau, K. M. (2011). Feasibility study of a real-time wandering detection algorithm for dementia patients. In *Proceedings of the First ACM MobiHoc Workshop on Pervasive Wireless Healthcare* (p. 11). ACM.