Visual abnormal event detection for prolonged independent living

Fabian Nater[1] Helmut Grabner[1] Luc Van Gool[1,2]

[1]Computer Vision Laboratory
ETH Zurich, Switzerland
{fnater,grabner,vangool}@vision.ee.ethz.ch

[2]ESAT - PSI / IBBT
K. U. Leuven, Belgium
luc.vangool@esat.kuleuven.be

Abstract—**In this paper, we apply visual surveillance techniques to in-house abnormal event detection, where (elderly) persons are monitored in order to assure their well-being. By means of a camera installed in a living room, we are able to spot not only if the person falls, but also other, more subtle abnormal events. We summarize two existing methods and compare them on two video sequences. The approaches both appear to perform well and additionally permit a semantic reasoning on the nature of the (abnormal) human behavior in the scene.**

Keywords-visual surveillance; independent living; abnormal behavior detection; human tracking;

I. INTRODUCTION

The elderly part of the population is growing constantly [1]. Society has to come up with products and services to assist this age group in diverse everyday tasks. For example, to extend the possibilities for independent living, systems have been proposed that raise an alarm in suspicious cases. In this context, fall detection is an important task. Solutions include simple push-the-button devices and automatic accelerometer-based, wearable systems. An overview of such fall detection techniques is given in [12]. Most of these systems have to be worn and require batteries that have to be charged.

Vision based approaches have the advantage of monitoring from a remote location (*c.f.* Fig. 1). For visual fall detection, rule-based methods have been established (*e.g.* [3], [9]). They perform well in predefined cases, for example a rapid change in the orientation of the foreground object [13], but lack general applicability. For a more detailed human behavior analysis, Cucchiara *et al.* [5] use probabilistic posture classification to detect a fall. In these approaches, the act of falling is modeled explicitly. There however are other suspicious situations, which a vision system could also detect, such as the presence of an intruder or the client limping. Visual surveillance tends to focus on abnormal event detection in general (*c.f.* [6] for a survey). Rather than modeling anomalies, many systems detect them as outliers to previously trained models of normality, (*e.g.* [14], [2]).

In our work, we analyze the motion and actions of people in their homes and detect outliers to pre-trained models. We recently proposed two such methods. The first one exploits a fixed model of normality based on a set of pre-trained, supervised human body trackers [11], the second one builds its own model in an unsupervised manner and can update itself incrementally [10]. As we will show, they share the paradigm of detecting abnormalities from hierarchical disagreements.

Fig. 1. Visual surveillance: detection of abnormal events

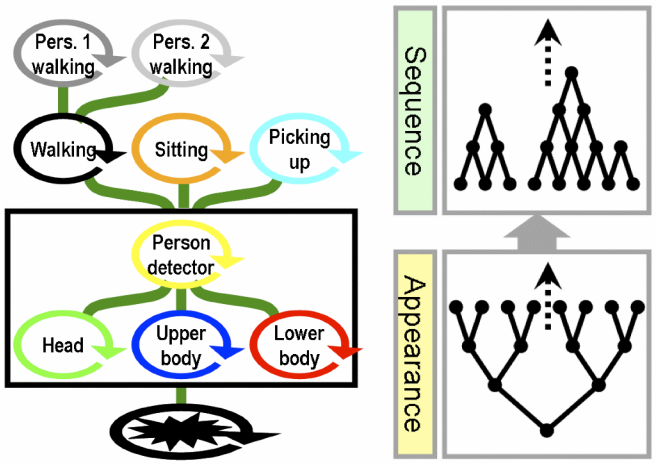
This paper compares the two methods with respect to the application in peoples homes. After giving an overview of the two methods (Sec. II), we show their adequacy and usefulness in a series of experiments (Sec. III). A detailed discussion will highlight benefits and drawbacks of both methods.

II. ABNORMAL EVENT DETECTION IN HIERARCHIES

We outline two methods for the detection of abnormal events. They are both based on an underlying model of normality, which is established from a set of training images. At runtime, when the system is active, unseen data is compared to this model and outliers are spotted. If the outliers persist, an abnormal event is detected and reported (*e.g.* in order to attract an operators attention).

Both models share their hierarchical organization. At each level, a less specific parent tracker or detector instance has more specific children, which use stronger expectations as prior information. The set of more specific children enumerate the possible subpatterns for the less specific parent. For instance, if a person has been detected, that person could be walking, sitting, or picking something up. The enumeration of 'normal' subpatterns depends on what the system has been trained for or has observed, respectively. Abnormality detection is in both cases based on this hierarchy. If none of the more specific nodes picks up on the data, but their less specific parent does, this is a sign that something abnormal is happening. For

This work is funded by the EU Integrated Project DIRAC (IST-027787).



(a) supervised, fixed (b) unsupervised, incremental

Fig. 2. (a) The implemented *tracker tree* with increasingly informed trackers for increasing levels. Each black circle depicts one tracker, a foreground object tracker is placed at the root node. (b) Overview of the unsupervised approach composed of two self-learnt hierarchies, encoding the appearances and the sequence of appearances respectively.

instance, a person is detected, but is not performing any of his normal actions.

A. *Supervised modelling with tracker trees* (TT)

The use of 'tracker trees' was our first approach to abnormality detection in hierarchies [11]. The idea is to arrange a set of trackers in a tree-like structure. Each tracker incorporates a certain amount of information about normality. Trackers further up in the tree have been trained for quite a narrow set of actions – *e.g.* specific to the walking style of one person - whereas trackers closer to the root node are able to track a broad variety of motions. The trackers are learnt in a supervised offline procedure, and the specificity of each tracker determines where the developer puts it in the tree. The currently implemented tracker tree for elderly care applications is visualized in Fig. 2(a).

A simple blob tracker is used [4] at the root. This tracker would cling on to whatever moves in the scene. One level up, a person detector is found [7]. Together with it, detectors for different body parts have been trained (legs, upper body, head-shoulders). Further up, different action-specific trackers are placed (walking, sitting, picking up). They rely on pre-trained low dimensional models (*c.f.* [11] for details). The same way, two person-specific trackers are learnt, which are the children of the generic walking node and which enumerate the different people the system knows about. As described earlier, the trackers are endowed with stronger and stronger knowledge about the (normal) world as one moves away from the root. All the different trackers operate autonomously, *i.e.* a tracker does not need the outcome of any other.

At runtime, unusual events are detected if a level can deal well with an event (can explain it with the available trackers), whereas none of the relevant trackers at the immediately higher level can. This is motivated by the fact that some tracker that uses more pertinent knowledge should be more robust and therefore explain the (normal) situation better. If however none of the more informed trackers can deal with the data, but the less informed one can, then this is a sign that something unusual is going on. From the location in the tree where this happens, a semantic interpretation on the nature of the abnormality is deduced. For instance, if none of the normal action specific trackers does well, but the person detector still works, this might be an indication of an unusual action like limping. Similarly, a fall or an intruder can be detected. If a body part (inside the black box in Fig. 2(a)) is missing, but a person is still detected, no tracker trained on an action defined by that part is expected to be active and no alarm is raised. This is useful for example when a person is walking behind an occluding object.

B. *Unsupervised hierarchical behavior analysis* (HBA)

A different method, as proposed in [10], learns a model of normal human behavior in a completely unsupervised manner. This model consists of two hierarchical representations arranged in a cascade, as illustrated in Fig. 2(b) and is designed such that an update is possible. This model of normality can thus be extended over time, which is in fact a crucial feature of a practical system. Obviously, not every normal situation can be learnt prior to the installation, and the concept of normality might as well change over time.

The first hierarchy in the model encodes human appearances and is built by a top-down process. Input features are clustered (*k*-means) in a hierarchical procedure, such that clusters are more specific on increasing layers. The root node has to describe all training features, whereas leaf node clusters only contain similar data and are more precise. Datapoints which fall into leaf node clusters are attributed a corresponding symbol. The second hierarchy explains sequences of appearances (*i.e.* actions or behavioral patterns) and is built by a bottom-up analysis, inspired by [8]. Temporally adjacent symbols are combined to basic level micro-actions, the combination of low-level micro-actions constitute higher-level micro-actions. In this hierarchy, a longer micro-action implies a more normal behavior.

This structure is used as a model of normality to which unseen data is compared. The person is tracked and the appearance is analyzed in the first hierarchy. Abnormal appearances are detected if the observation is matched to a cluster on one hierarchical layer, but is outlier to all its connected, more specific clusters. Similarly, the actions are analyzed in the second hierarchy.

III. EXPERIMENTS

We evaluate our two methods on two video sequences, recorded in a living-room environment (Data available from `www.vision.ee.ethz.ch/fnater`). One single person is monitored and abnormal events are spotted. The test footage contains in total about 1800 images, which were recorded with a static camera at 15 frames per second in *VGA* resolution. These sequences contain diverse 'every-day' actions such as walking, walking behind occluding objects, sitting on different chairs, picking up small objects, *etc*. but also have abnormal events, *e.g.* the person jumps over the sofa, falls, limps, waves heavily or an intruder enters the room. The images are background subtracted and silhouettes serve as input features.

The supervised tracker tree method (*TT*, *c.f.* Sec. II-A) was trained on approximately 3000 images. Actions were segmented manually for training. The results for the test sequences are presented in the first two lines of Fig. 3(a) and (b). The bounding boxes of the active trackers are displayed in each frame, employing the color code of Fig. 2(a). If a hierarchical disagreement occurs, an abnormal event is detected and the entire image is framed in red.

The unsupervised hierarchical behavior analysis (*HBA*, *c.f.* Sec. II-B) is initially trained with a video of approximately 7000 images containing normal actions, which are performed repeatedly and updated later on. No annotation is provided. The results for this method are presented in the lower part of Fig. 3(a) and (b). The bounding box indicates the tracked location of the person, its color shows how normal the appearance is and the black bar on the left encodes the normality of the performed action. The entire frame is marked red if the person cannot be tracked, *i.e.* the observed appearance does not match any learned cluster in the appearance hierarchy.

In the following we emphasize strengths and shortcomings of the two approaches with examples of the two sequences.

- Usual actions, such as walking sitting, picking up an object from the floor are incorporated in the models of both methods and can therefore be tracked successfully. Walking behind occluding objects is also handled in both cases.
- A fall is always recognized as an abnormal event.
- The person limps after the fall (seq. 1, frame 6). This is recognized by *TT*, since the walking tracker does not apply, but all body parts are observed. In the case of *HBA*, the abnormality is indicated by a small action bar, which means that this action has not been registered before. The appearances however remain normal (green bounding box).
- If an intruder enters the room (seq. 1, frame 9), this is noticed by the *TT* thanks to its specialized trackers. The *HBA* does not recognize that event.
- In sequence 2 we see the benefits of *HBA*. Its underlying model of normality has meanwhile been updated and lying on the sofa is now known (seq. 2, frame 6). Since no automatic update is possible for *TT*, this is abnormal here.
- In addition, due to the larger number of hierarchical levels in *HBA*, normality can be distinguished to different degrees. This is for example useful in case of heavy waving (seq. 2, frame 10), which the *TT* does not detect (lower body is observed normally).

In general, one can notice, that both approaches perform well for the detection of abnormal events in assisted living scenarios. Both also robustly track the person and hence, a combined tracking and action reasoning is performed.

The main advantage of the supervised tracker tree lies clearly in the fact, that the role of each tracker is known precisely, also in the case when no abnormality is occurring. If for example the sitting tracker is active, one can be sure that the person is sitting. In an unsupervised system on the other hand, this information is not provided, since it is established without human interaction. The cost of a manually built system is however its fixed model, once it is in action, human effort would be required to increase its capacities. It thus can only detect what it has been tuned for. This in fact limits its application in rooms of different persons, it would have to be trained specifically every time.

On the other hand, an unsupervised, self-learning system is much more flexible, since it is trained automatically and can even adapt itself over time to the actions it observes frequently. From the intrinsic hierarchical structure, a more subtle reasoning can be performed and 'slightly' abnormal events can be detected as well. Yet, the interpretation of the results is not always so obvious, due to the lack of human annotation.

IV. Conclusion

In this paper, we have shown the applicability of vision based human tracking methods for the use in assisted living scenarios, where elderly people are monitored by an automated system. We briefly outlined two recently proposed methods, and qualitatively evaluated and compared them on two video sequences. They both perform well, have however complementary advantages. While one method is supervised and allows for precise reasoning, the second one is unsupervised and incremental, and therefore requires no human interaction and can adapt to different situations. An ideal system would of course incorporate the advantages of both methods, a mutual integration is therefore envisaged and currently investigated.

References

[1] `en.wikipedia.org/wiki/Population_ageing`, 22/04/2010.

[2] A. Adam, E. Rivlin, I. Shimshoni and D. Reinitz. Robust Real-Time Unusual event detection using multiple fixed-location monitors. *PAMI,* 30(3):555-560, 2008.

[3] 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.

[4] G. R. Bradski. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal,* (Q2), 1998.

[5] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani. Probabilistic posture classification for human-behavior analysis. *Trans. on Systems, Man, and Cybernetics*, 35(1):42-54, 2005.

[6] H. M. Dee and S. A. Velastin. How close are we to solving the problem of automated visual surveillance? *Machine Vision and Applications*, 19(5-6):329-343, 2008.

[7] P. Felzenszwalb, D. Mcallester, D. Ramanan. A discriminatively trained, multiscale, deformable part model. In *Proc. CVPR*, 2008.

[8] S. Fidler, G. Berginc, A. Leonardis. Hierarchical statistical learning of generic parts of object structure. In *Proc. CVPR*, 2006.

[9] A. H. Nasution and S. Emmanuel. Intelligent video surveillance for monitoring elderly in home environments. In *IEEE Workshop on Multimedia Signal Processing*, 2007.

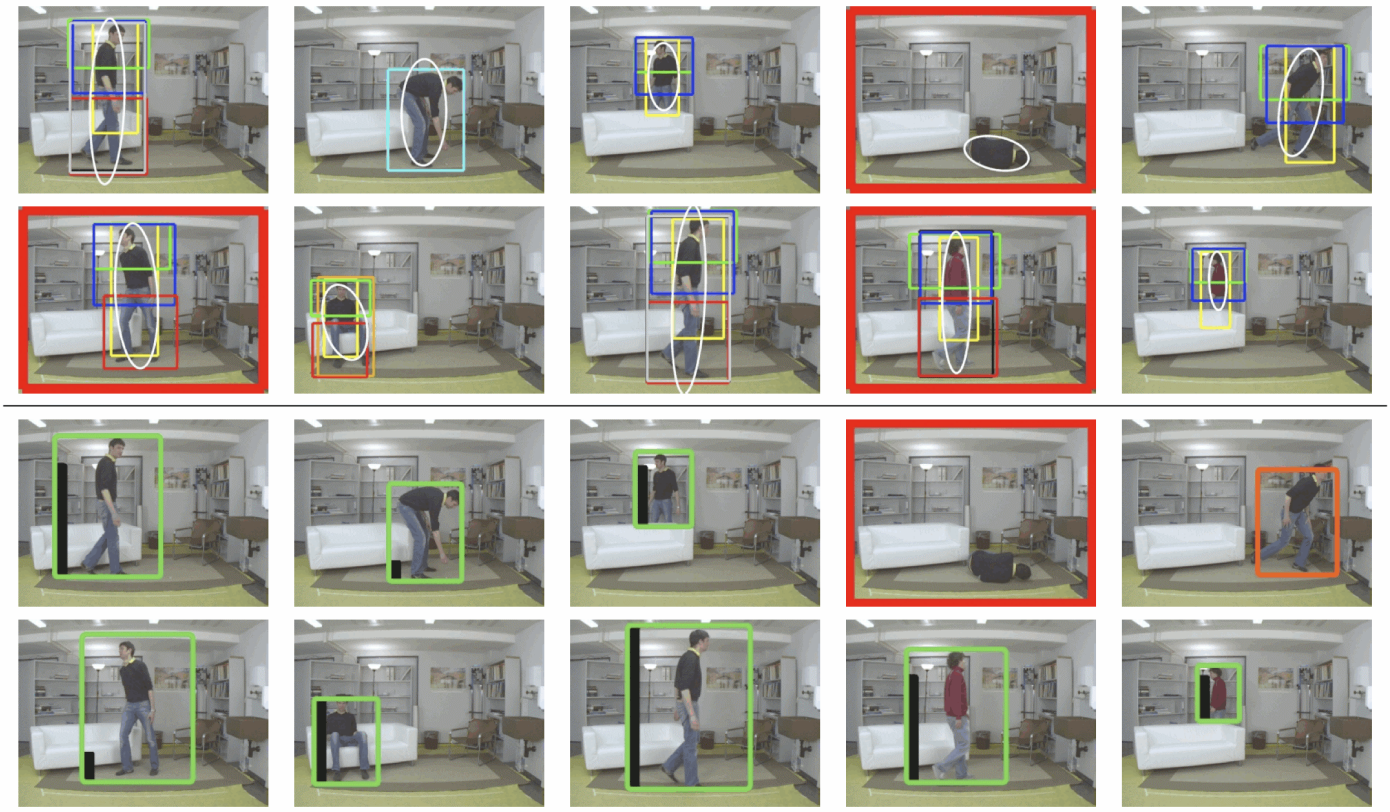
[10] F. Nater, H. Grabner and L. Van Gool. Exploiting simple hierarchies for unsupervised human behavior analysis. In *Proc. CVPR*, 2010.

[11] F. Nater, H. Grabner, T. Jaeggli and L. Van Gool. Tracker trees for unusual event detection. In *ICCV WS on Visual Surveillance*, 2009.

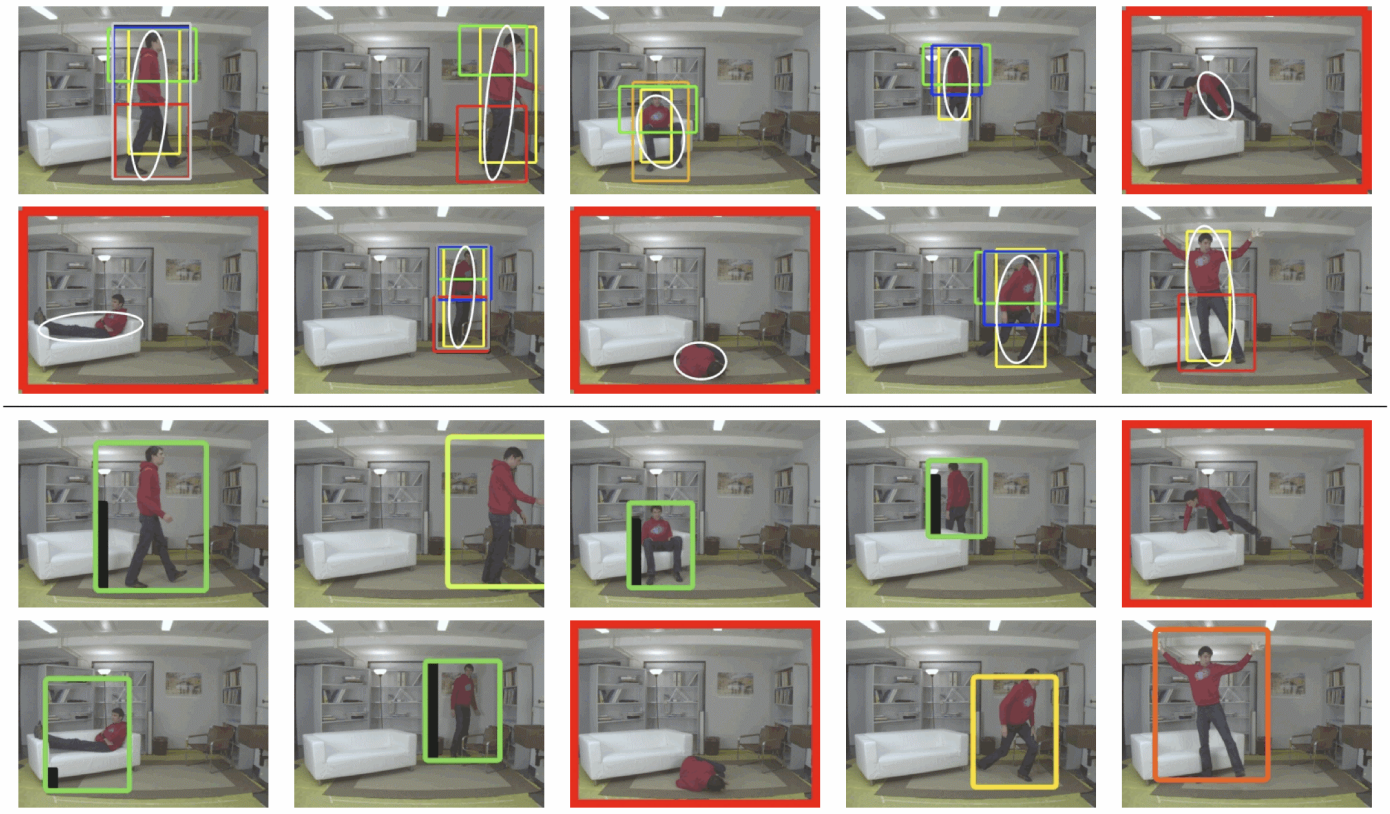
[12] N. Noury, A. Fleury, P. Rumeau, A. K. Bourke, G. O. Laighin, V. Rialle and J. E. Lundy. Fall detection - principles and methods. In *IEEE Engineering in Medicine and Biology Society*, 2007.

[13] C. Rougier, J. Meunier, A. St-Arnaud, J. Rousseau. Fall detection from human shape and motion history using video surveillance. I*n Advanced Information Networking and Applications Workshop*, 2007.

[14] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *PAMI*, 22(8):747-757, 2000.



(a) Results for test sequence 1



(b) Results for test sequence 2

Fig. 3. Ten selected frames of two test sequences, they are displayed in order to visualize the results of both approaches on the same frames. The results of the supervised tracker tree method (*TT*) are in the top rows, the active trackers are displayed and the color code of Fig. 2(a) is employed. In the bottom rows the results for the unsupervised hierarchical behavior analysis (*HBA*) are displayed, the bounding box color indicates the normality of appearance, the length of the black bar on the left encodes the normality of action. If an abnormality is spotted, the entire frame is marked red. Both methods are able to spot abnormal events in independent living scenarios, see text for a discussion. Videos are available from `www.vision.ee.ethz.ch/fnater/`.