Activity Summarisation and Fall Detection in a Supportive Home Environment

Hammadi Nait-Charif and Stephen J. McKenna
Division of Applied Computing, University of Dundee, Dundee DD1 4HN, Scotland
{hammadi, stephen}@computing.dundee.ac.uk

Abstract

Automatic semantic summarisation of human activity and detection of unusual inactivity are useful goals for a vision system operating in a supportive home environment. Learned models of spatial context are used in conjunction with a tracker to achieve these goals. The tracker uses a coarse ellipse model and a particle filter to cope with cluttered scenes with multiple sources of illumination. Summarisation in terms of semantic regions is demonstrated using acted scenes through automatic recovery of the instructions given to the actor. The use of 'unusual inactivity' detection as a cue for fall detection is also demonstrated.

1. Introduction

Home environments able to monitor automatically the activities of their occupants can help extend independent, quality living and reduce healthcare costs [1, 4, 9]. In particular, patterns of inactivity can be used to make inferences about health and to help detect falls. It is important to note that the significance of inactivity changes with context. A person lying on a sofa, as she often does, is probably only resting. In contrast, a person lying on the floor where she has not previously lain may have fallen and require assistance. The method presented here enables inactivity outside usual zones of inactivity (e.g. chairs, beds) to be detected. When combined with body pose and motion infomation this should provide a useful cue for fall detection. In addition, a human-readable description of activity in terms of semantic regions provides a useful summary of behaviour.

2. Supportive Home Environments

Sensors in current supportive home environments often have relatively narrow functionality, e.g. passive infra-red sensors, pressure pads and fridge door sensors enable room occupancy, presence in an area or use of a fridge to be monitored [1, 4, 6]. Worn fall detectors are often not worn when returning home, during housekeeping tasks prone to cause false alarms, or when uncomfortable [7, 13]. Embedded sensors, in contrast, have the advantage of ensuring compliance within the home, although multiple sensors are required. User requirements, acceptability and privacy issues surrounding the use of computer vision for home monitoring are being explored using a novel drama-based methodology described elsewhere [9]. Interpretive aims for such vision systems range in complexity from monitoring room occupancy to detecting falls and performing analyses of activity patterns. Reduced mobility can be predictive of a fall and has other health implications. Inactivity detection can, in a context-dependent way, indirectly indicate ill-health or a fall. Activity patterns and significant changes in daily or weekly patterns might also be detected. Resulting information could be used as part of an alarm system (potentially detailing the nature of the alarm event) as well as for prediction and thus prevention of falls through risk assessment. Summarisation of activity in human-readable form enables retrospective analysis, providing insights into behaviour and health to care providers and researchers.

3. Experimental Scenario

Older peoples' homes are often cluttered with furniture brought from former, larger homes. The position and orientation of the cameras has been chosen to minimise occlusion of the person by furniture. The set-up investigated here uses ceiling-mounted, wide-angle cameras with vertically-oriented optical axes. Standard, as opposed to infra-red, cameras have been employed; high resolution infra-red sensing remains relatively expensive. (Preliminary work has been performed using low-resolution infra-red sensing for fall detection [5, 14]).

Figure 1 shows a scene used to illustrate the method. Strong perspective effects due to the wide-angle lens are apparent. Four semantic regions are labelled: two entrances to the room (H and R), a chair with a telephone beside it (C), and a sofa (S). A total of 97 sequences were acquired at $30Hz$ with a resolution of 480×360 pixels (46755 frames, 26 minutes). Aquisition was over two days of changeable weather. The scene contained multiple light sources (windows and indoor lighting) and no attempt was made to control the extent of lighting changes and cast shadows.

Figure 1. Salient regions: a sofa (S), a chair (C), the hall door (H) and the rear door (R).

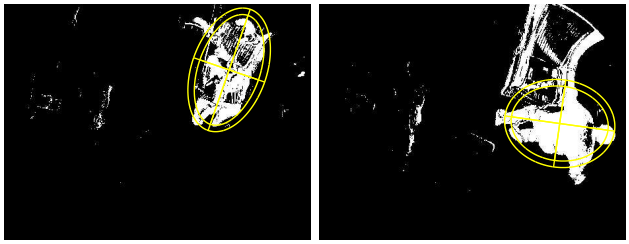

Figure 2. Examples of the strongest particle overlaid on pixel foreground images. The interior ellipse represents the state estimate.

4. Overhead Tracking

A few systems for overhead person tracking have been proposed previously for different applications. For example, the KidsRoom used overhead tracking as a component in a perceptual, interactive play-space [3]. In a home environment, the lighting and layout are far less constrained. Furthermore, the clothing and body postures that will be encountered are highly variable and it cannot be assumed that the articulated structure of the body will be apparent. It is likely to be very difficult to construct detailed statistical shape models that capture the range of variation in such a way that enables the unusual poses associated with events such as falls to be tracked and detected. Instead, the person's position in the image along with a coarse representation of their shape and orientation in the image were tracked using an ellipse so that the state at time t was $\mathbf{e}_t = (x_t, y_t, \psi_t, s_t, e_t)$ where (x_t, y_t) is the ellipse centre and the other parameters are orientation, scale and eccentricity respectively. The authors believe that this representation of a person is rich enough to support recognition of relevant actions and events such as falling, lying down, sitting and standing. It is also coarse enough to enable a wide range of body poses and clothing to be tracked.

Several authors have tracked objects and people using

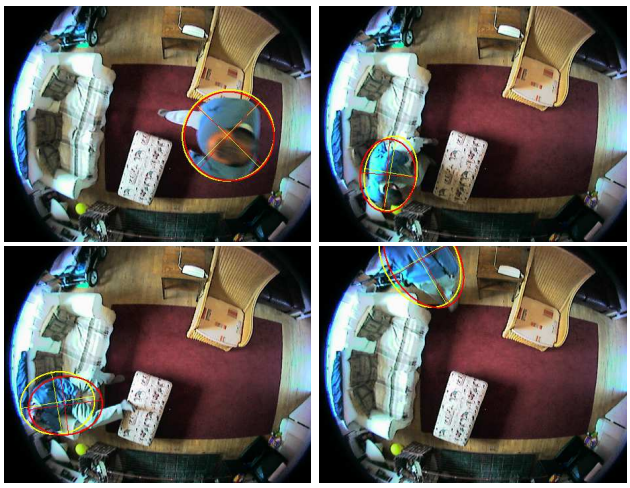

Figure 3. Example ellipse estimates.

either ellipses or Gaussian 'blobs' with elliptical isoprobability contours in image space (e.g. [2, 8, 12, 15]). Measurements made when tracking with an elliptical contour model assume that the ellipse provides a reasonably accurate 2D shape model and that image features such as edges will therefore lie close to the contour [2, 12]. In the case of overhead tracking, body shape is highly deformable and poorly modelled by an ellipse. A spatial Gaussian distribution can be effective even when the object is not elliptical but it is not robust to clutter.

The tracker used here employs a particle filter (Iterated Likelihood Weighting [12]) with image evidence provided using an adaptive background model with shadow detection [10]. Hypothesised ellipses were scored using a function that provided some robustness to noisy background cues (e.g. due to shadows or motion of other objects such as cushions) and highly non-elliptical poses (e.g. outstretched arms). Specifically, pixels exterior to the ellipse and within N pixels of the ellipse contour were considered to constitute an adjacent annular region (see Figure 2). The score was computed such that an ellipse hypothesis was penalised for having pixels in this adjacent region that were likely to be foreground and for having pixels in the ellipse interior that were likely to be background. The nonparametric representation of the state density enables tracking through ambiguous situations and is important when dealing with strong shadows and clutter. Figure 3 shows typical estimates obtained during tracking. Two ellipses are displayed: the strongest particle and the mean.

The tracker provided trajectories in the 5D ellipse parameter space. These trajectories were temporally smoothed using a moving average filter and the person's speed in the image plane was estimated at each point. Smoothed ellipse centre trajectories and speeds were subsequently used to provide a compact representation of the person's global mo-

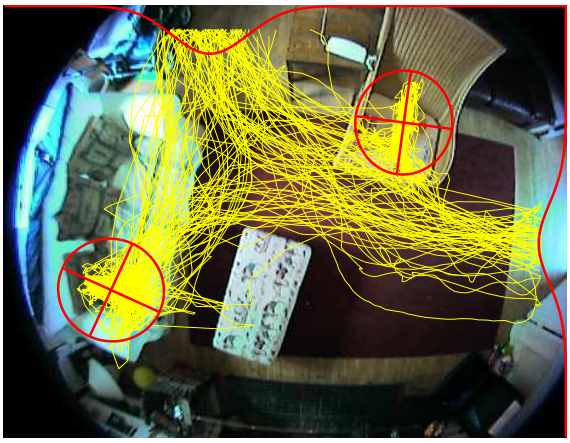

Figure 4. Smoothed trajectories, inactivity zones and entry zones.

tion. The remaining parameters provided pose information but were not used further here. Trajectories were extracted and represented directly in the image plane. The use of a ground-plane constraint was inappropriate because the distance from the person's 'centre' to the floor was large relative to the camera distance and varied significantly with body pose. The camera was uncalibrated and the person was tracked without performing image rectification.

5. Activity Recognition

Within a room in a home, there will typically be a few places in which an occupant spends most of her time while in that room. A living room, for example, contains chairs and sofas and the occupant might even have a favourite seat in which she invariably sits to watch television, read or sleep. Such places will be referred to as *inactivity zones* to indicate that occupancy of such a zone tends to involve little global motion of the person. A room will have a fixed set of entrances which also serve as exits. A place in which entry and exit occurs will be referred to as an *entry zone*. Typical use of a room involves entering followed by visits to one or more *inactivity zones* and finally exiting the room. Of course other activities may occur but these tend to be more highly variable and transient. It is proposed that a useful, compact, semantic representation of behaviour in this context can be provided by temporal segmentation of sensor data into time spent (i) entering via *entry zones*, (ii) inactive in *inactivity zones*, (iii) transitioning between *zones*, and (iv) exiting via *entry zones*. In order to achieve the goals of unusual inactivity detection and behaviour summarisation, a model of spatial context was learned. This was done using MAP estimation of Gaussian mixture models to automatically identify and characterise inactivity zones and entry zones [11]. The learned models of spatial context can be used to automatically temporally segment trajectories and to detect unusual inactivity. The Gaussian components in the mixtures correspond to 2D inactivity zones and 1D entry zones (see Figure 4). Each Gaussian PDF, $p(\mathbf{x}_t|k)$, provides a model for the spatial extent of a zone, k. Entry zones can be used to focus tracker initialisation and to semantically label points of entry and exit. When a person's speed drops to an extent that indicates inactivity, the inactivity zone PDFs provide a way of checking whether the inactivity is occurring in a known inactivity zone. A simple algorithm was used to decide when the person was inactive in a known inactivity zone. Speed, s_t, at each time-step was estimated using a finite difference over a 40-frame temporal window and $p(\mathbf{x}_t|k)/(s_t + 0.1)$ was tested against a threshold.

In order to demonstrate empirically an ability to summarise activity and detect unusual inactivity, an actor was instructed to perform a series of activities in the room designed to emulate aspects of the way an older person might use such a room. Instructions were given in terms of the four regions: H, R, C and S. For example, "enter through the hall door, sit on the sofa and then exit through the rear door" (HSR) or "enter through the hall door, sit and use the telephone, sit on the sofa and then exit through the hall door" (HCSH). Table 1 summarises the 13 classes of sequence acquired, classified according to the instructions given to the actor. The "Fall" class contained sequences in which the actor was instructed to simulate a fall. (There are obvious barriers to obtaining a video data set of older people falling in reality). Sequences were allocated at random to training and test sets such that half of the examples in each class were reserved for testing. Figure 4 shows smoothed ellipse centre trajectories obtained from the training data. Tracking errors occurred in 3 of the 97 sequences. Entry/exit and inactivity points from the training data were used to learn spatial context models. The Bayesian learning method used [11] resulted in model zones that exhibited a one-to-one correspondence with the semantic regions referred to in the natural language instructions given to the actor.

Automatic summarisation of the sequences compared well with the actor's instructions. The annotations given in Table 1 were all recovered correctly with the exception of two HSH sequences which were incorrectly annotated as HSSH. This was due to the person leaning forward and the algorithm therefore labelling them as temporarily leaving and then reentering the S inactivity zone. Figure 5 shows some example trajectories with the current ellipse overlaid. The trajectories here are colour-coded to indicate the temporal segmentation obtained. In each of these cases, the trajectory was correctly segmented into transitions between zones, inactivity within a known inactivity zone and inactivity while not in a known inactivity zone.

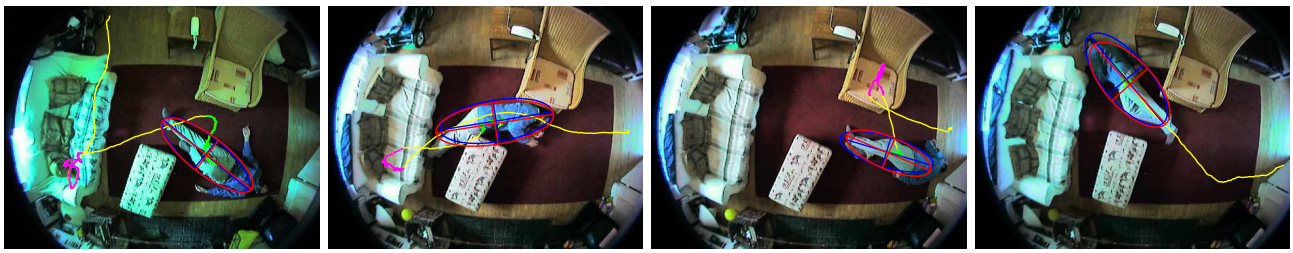

Figure 5. Segmented trajectories and detected unusual inactivity.

Sequence annotation	No. of examples	Average duration (frames)	Tracking errors
RH	11	103	0
RSR	7	504	0
RSH	4	470	0
RCR	7	514	1
RCH	4	561	0
HR	11	119	1
HSR	4	536	0
HSH	16	506	0
HSCH	5	998	0
HCR	4	619	0
HCH	11	672	1
HCSH	4	1150	0
Fall	9	513	0

Table 1. Annotated sequences.

6. Conclusions

A method was demonstrated in a supportive home environment for providing human-readable summarisation of activity and detection of unusual inactivity. High-level activity summarisation and context-dependent inactivity detection are also important in other applications. The former provides an efficient coding for storage and retrieval. The latter is useful in monitoring and surveillance.

Passive fall detection has been identified as a priority for supportive home environments for older people. The methods presented here have gone some way to providing useful cues for fall detection. In future work it is planned to combine these cues (unusual inactivity) with dynamic models of falling. Visual environmental factors such as lighting levels and room layout changes can be significant for many older people with poor vision and are implicated in falls. Automatic detection of such changes is also under investigation.

References

[1] N. M. Barnes, N. H. Edwards, D. A. D. Rose, and P. Garner. Lifestyle monitoring: technology for supported independence. *IEE Computing and Control Engineering Journal*, pages 169–174, August 1998.

[2] S. Birchfield. Elliptical head tracking using intensity gradients and color histograms. In *CVPR*, Santa Barbara, 1998.

[3] A. Bobick, S. Intille, J. Davis, F. Baird, C. Pinhanez, L. Campbell, Y. Ivanov, A. Schutte, and A. Wilson. The kidsroom: A perceptually-based interactive and immersive story environment. *Presence*, 8(4):367–391, 1999.

[4] S. Bonner. Assisted interactive dwelling house: Edinvar housing association smart technology demonstrator and evaluation site. In *Improving the Quality of Life for the European Citizen (TIDE)*, pages 396–400, 1998.

[5] P. A. Bromiley, P. Courtney, and N. A. Thacker. Design of a visual system for detecting natural events by the use of an independent visual estimate: A human fall detector. In *Empirical Evaluation Methods in Computer Vision*. 2002.

[6] M. Chan, H. Bocquet, E. Campo, and J. Pous. Remote monitoring system to measure indoors mobility and transfer of the elderly. In *Improving the Quality of Life for the European Citizen (TIDE)*, pages 379–383, 1998.

[7] K. Doughty. Fall prevention and management strategies based on intelligent detection, monitoring and assessment. In *New Technologies in Medicine for the Elderly*, Charing Cross Hospital, London, November 2000.

[8] F. Liu, X. Lin, S. Z. Li, and Y. Shi. Multi-modal face tracking using Bayesian network. In *Proc. IEEE Int. Workshop Analysis & Modeling of Faces & Gestures*, Nice, 2003.

[9] F. Marquis-Faulkes, S. J. McKenna, P. Gregor, and A. F. Newell. Gathering the requirements for a fall monitor using drama and video with older people. *Technology and Disability*, 2004. In Press.

[10] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler. Tracking groups of people. *Computer Vision and Image Understanding*, 80(1):42–56, 2000.

[11] S. J. McKenna and H. Nait Charif. Learning spatial context from tracking using penalised likelihood estimation. In *International Conference on Pattern Recognition*, 2004.

[12] H. Nait-Charif and S. J. McKenna. Tracking poorly modelled motion using particle filters with iterated likelihood weighting. In *ACCV*, pages 156–161, Korea, 2004.

[13] SeniorWatch. Fall detector. Technical report, Case Study of European IST Seniorwatch Project, IST-1999-29086, 2001.

[14] A. Sixsmith and N. Johnson. SIMBAD: Smart inactivity monitor using array-based detector. *Gerontechnology*, 2(1):110–110, 2002.

[15] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. PAMI*, 19(7):780–785, 1997.