UBIQUITOUS SENSING FOR POSTURE/BEHAVIOR ANALYSIS

Jeffrey L Wang, Benny L Lo, and Guang-Zhong Yang
Imperial College London, The United Kingdom

ABSTRACT

Advances in body sensor networks require a combined strategy for ambient and wearable sensing. The purpose of this paper is to develop a pervasive visual sensing technique for automated human behavior analysis. A set of vision cues for posture classification is proposed so that adverse events such as unbalanced gait or fall can be detected. The method starts with the extraction of human blobs and then personal metrics that lead to behavior profiling. This provides an un-obtrusive monitoring environment that complements the current development of body sensor networks.

Keywords: posture estimation, ambient sensing, behavior profiling, computer vision, ubiquitous sensing

INTRODUCTION

Increasing demand on public healthcare service due to the aging population has become a major problem in developed countries. Growing efforts have been directed towards alternative ways of providing quality care for the elderly at home. Thus far, a number of clinical gait analysis studies have shown that changes in posture and gait, especially for the elderly or patients with chronic illness, can indicate the onset of an adverse event or the worsening of an existing problem (1,2). For example, changes in gait can be associated with early signs of neurological abnormalities linked to several types of non-Alzheimer's dementias (3). Unstable gait can be a major factor contributing to falls, which can be fatal (6). In parallel with the advances in ubiquitous computing techniques, extensive research is being carried out in using sensor networks for home care environments (3). However, their general lack of richer movement information about human behaviors means that precursors to certain adverse events cannot be monitored.

Emerging technologies in computer vision and pervasive sensing make feasible the development of an automated system for human behavior analysis. Currently, vision-based human motion analysis (12) is one of the most active research topics, which mainly addresses detection, tracking and recognition of people, and more generally, the understanding of human activity patterns from image sequences. In contrast to the sensor based homecare systems, researchers are actively pursuing the use of low-cost noninvasive computer vision techniques for monitoring and assessing daily activities of the occupants. For example, Tao and Hu proposed a visual tracking system for home-based rehabilitation to help stroke patients to recover and improve their mobility (7), Nait-Charif *et al.* presented a simple vision system set up in a supportive home environment for activity summarisation and fall detection (4), and Gao *et al*. proposed a technique of fusing motion segmentation with tracking for eating activity analysis of patients in a nursing home (5).

The purpose of this paper is to describe the UbiSense framework can be used along with body sensor networks in a homecare environment. It aims to provide an unobtrusive domestic health monitoring system for the elderly by using computer vision techniques for detecting changes in postures, gait and activities of the occupants. The proposed posture estimation and change detection methods are based on fusing multiple visual cues such as projection histogram, radial shape description and elliptical fitting. The method adopts the concept of "*from blobs to personal metrics to behaviour profiling.*"

HUMAN BLOB DETECTION

To extract human blobs from the background in each frame, a statistical background subtraction method based on Gaussian models $N(u,\sigma^2)$ in the RGB color space similar to (8) is used. Scene initialization is achieved over a period of frames such that the mean and standard deviation of the color distribution for each pixel is derived. This can be used to compare with colors of the associated pixels in current frame to produce a mask that is considered as a region of interest for further processing, i.e.,

$$M^t_{x,y} = \bigcup\{| f^t_{x,y} - u^t_{x,y} | \geq \lambda\sigma^t_{x,y}\}$$

where $f^t_{x,y}$ denotes the value of each color component R , G or B at (x, y) at time t . To take into account changes in the background, the background model is adapted progressively. The images resulting from background subtraction are first filtered with a 3-by-3 median filter for noise and distortion removal. Morphological operations are then applied to smooth out the region boundaries and fill holes due to erroneous segmentation in regions of low chromatic content or low texture. The connected component analysis is finally applied to extract a single highly compact connected region, as shown in Fig. 1.

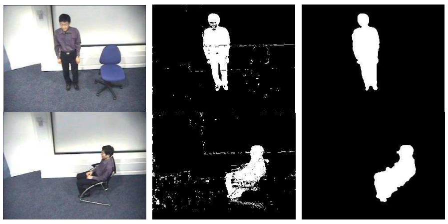

Fig. 1. Original images (*left column*), binary masks (*middle column*), and the results after post-processing (*right column*)

BLOB METRICS

Since the spatial distributions of the pixels embody sufficient information for estimating human postures, it is possible to extract features directly from those detected binary foregrounds. To this end, three different simple representations are used to describe human blobs.

***Radial shape representation*:** For reducing redundancy and keeping spatial relationships of the foreground pixels, contours are used to describe each shape in a radial representation as outlined in Fig. 2. To this end, the silhouette boundary is first extracted, followed by determining of the shape centroid (x_c, y_c). The boundary is then unwrapped into a distance representation $s=[r_1,r_2,\cdots,r_i,\cdots,r_N]^T$ consisting of all ordered radial distances between the centroid and each boundary pixel (x_i, y_i) (9), which is then normalised with respect to the amplitude and length.

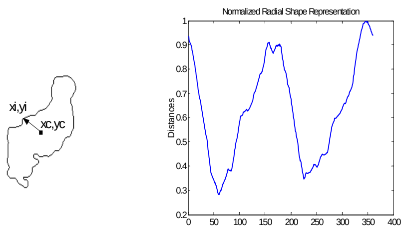

Fig. 2. Normalised radial distance signals for sitting

***Best-fit ellipse*:** To extract the principal axis of the binary foreground object C, elliptical fitting based on the moment operators is applied, as shown in Fig. 3. An ellipse is defined in a 5D parameter space as $[\bar{x},\bar{y},\phi,a,b]$, where $(\bar{x},\bar{y})$ is the central position of the ellipse. The orientation, ϕ, is defined as the angle of axis of the least moment of inertia, and it can be easily computed by the central moments

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

$$\phi = \frac{1}{2} tan^{-1}\left(\frac{2u_{11}}{u_{20}-u_{02}}\right)$$

The lengths of the major axis, a, and the minor axis, b, can be computed through the relationships between them and the least and greatest moments of inertia C.

Fig. 3. Some examples of ellipse fitting for different postures

Projection histograms: For posture analysis, the projection histogram (10) is applied. The method is particularly useful when background subtraction performs well. The histogram is defined as the total numbers of foreground pixels projected along horizontal or vertical directions. Histogram normalisation is required to obtain scale-stable results, as shown in Fig. 4.

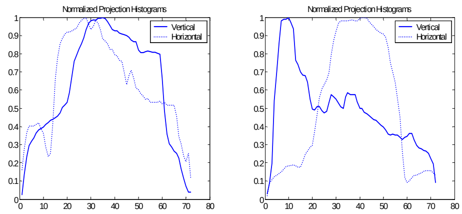

Fig. 4. Normalised projection histograms for standing (*left*), and lying down (*right*)

POSTURE ESTIMATION

It is evident that there are apparent variations between different postures, so the above silhouette representations are directly used to perform separate posture estimation. For radial shape representation, Procrustes shape analysis that caters for changes in rotations, position, and scale is used. Given a set of n shapes u_i, their mean can be found by computing the matrix

$$S_u = \sum_{i=1}^{n} (\mathrm{u}_i \mathrm{u}_i^*)/(\mathrm{u}_i^* \mathrm{u}_i)$$

The Procrustes mean shape is the dominant eigenvector that corresponds to the greatest eigenvalue of S_{u}. Results show that for the elliptical representation, orientation and aspect ratio allows robust distinction between postures. To incorporate the normalised projection histogram, templates for each posture are computed by using a set of detected silhouettes and the sum of absolute differences can be used to estimate the most similar postures.

$$S_{ph} = \sum_h \sum_v |H_h^i - PH_h| + |V_v^i - PV_v|$$

where H_h^i and V_v^i, PH_h and PV_v are the horizontal and vertical projection histograms of the ith main postures and the detected silhouette, respectively.

Let T represent a test posture and T_i represent the i^{th} reference posture in terms of certain silhouette representation. Blob metrics is used to classify a given posture to one of three main postures by minimizing the similarity between it and all reference patterns

$$c = \arg\min_{i} d(T, T_i)$$

where *d* is any one of the similarity measures computed for three different representations.

BEHAVIOR PROFILING

To analyze posture changes, accumulated change scores based on appearance information are used to represent the degree of change from the exemplar posture according to their posture variances. Under this framework, if the change scores are beyond the range of normal posture obtained from training, they will be considered as changed. Appearance is not necessarily preserved across different sequences. It is hard to represent small changes in postures, because they can mistakenly be judged to belong to certain normal postures (*e.g.*, a hand waves left and right before the body). Considering the fact that appearance is sensitive to background subtraction, optical flow is used as an alternative to detect subtle changes from certain body parts such as head, hand and leg. Features based on pixel-wise optical flow (11) are the most natural way to capture motion independent of appearance. Any residual motion within the spatio-temporal volume is due to the relative motion of different body parts. PCA-based eigenspace transformation is applied to the amplitude of regional flow vectors. Optical flows of different posture changes with respect to the reference are clustered in eigenspace for comparison, as shown in Fig. 7.

Since an observed activity can be regarded as a measurement vector over the temporal axis, an aggregation of the visual cues can be used for determining, as well as predicting, the corresponding posture in serial time frames, thus providing information about the human activity. An example is shown in Fig. 8.

EXPERIMENTAL RESULTS

A UbiSense node with SAMSUNG SCC-641 was used to capture human body motion. Only three common postures are considered, *i.e.*, standing, sitting and lying down. The images were captured at a rate of 15 frames per second in AVI format and finally decoded into 24-bit JPG files with a resolution of 320×240. The training data includes subtle random motion for each posture. In contrast to the training data, during classification, the subject can move more freely so as to demonstrate the practical value of the proposed processing framework, and some example images are illustrated in Fig. 5.

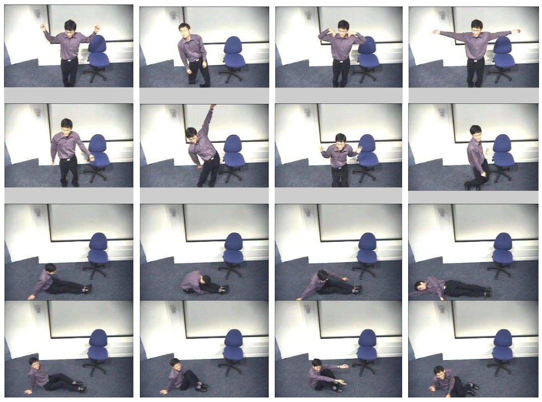

Fig. 5. Posture changes: standing (*top two rows*), and lying down (*bottom two rows*)

To evaluate the robustness of posture estimation, classification of the sequence with posture changes was performed according to the minimum similarity value. It can be shown that for all postures, both ellipse fitting (100%) and projection histogram (average 99%) algorithms perform better than radial shape representation (average 55%). In general, ellipse fitting is the most robust because it focuses more on the whole body shape rather than changes in individual limbs. It is also evident that the projection histogram is sensitive to errors in the binary shape. However, the radial shape description is particularly most sensitive to shape variations associated with posture changes, while this unique characteristic enables it perform better in experiments of change detection of postures (accuracy of above 85% compared with ground truth data).

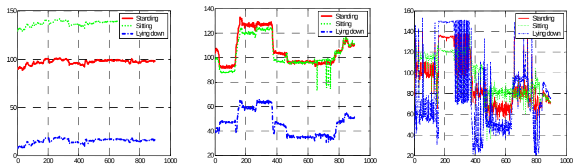

Fig. 6. Posture estimation on changed lying down sequences (similarity vs. frame no): ellipse (*left*), projection histogram (*middle*) and radial description (*right*)

Although appearance information can be used to judge the majority of postures with relatively bigger changes, it is more important to discriminate between the change positions and especially between changes of interest (*e.g.*, possibly associated with physical changes). Appearance is not sufficient to detect local subtle changes in postures, and therefore optical flow analysis is used to avoid this problem. Fig. 7 outlines some example images of the standing posture with small head movements towards different directions, and their corresponding results by using the eigenspace analysis. The results are promising, in particular for determining which type of change is occurring.

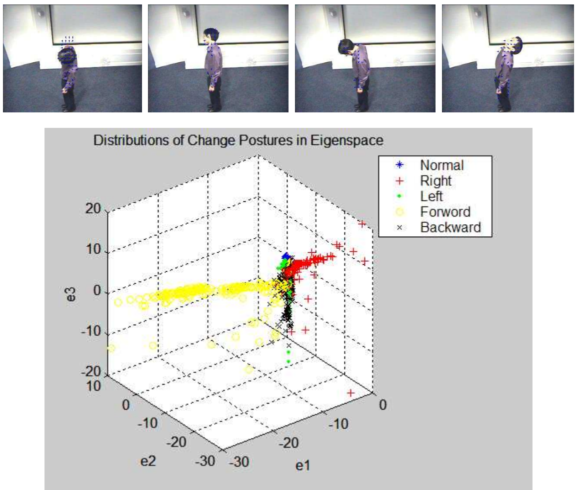

Fig. 7. Subtle posture changes (*top*) and clustering results based on optical flow (*bottom*)

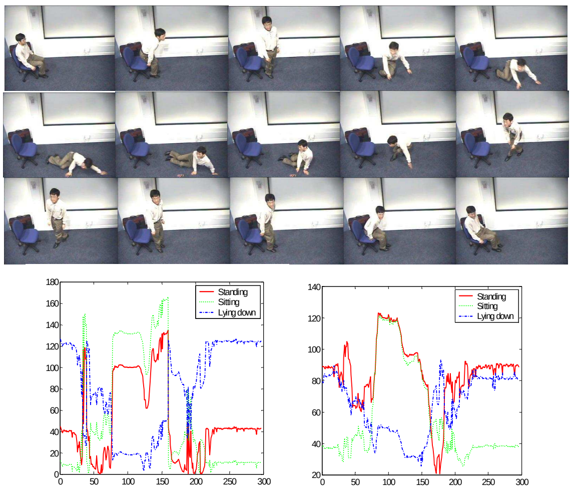

Fig. 8. Posture estimation and activity reasoning. Some images in an activity sequence (*top*) and results of posture determination using elliptical fitting and projection histogram (*bottom*)

Activities analyzed include entering and exiting the monitored area, walking, and falling. In particular, extra video sequences collected on different days were used. They include continuous changes of postures, *e.g.*, from sitting, standing, lying down, standing, walking, to sitting again, as shown in Fig. 8. The purpose of this data is to simultaneously evaluate posture estimation, activity interpretation, and sensitivity analysis with respect to both different clothes and scenes. It is evident that ellipse fitting still performs correctly, while radial description performs the worst. The incorrect frames for the projection histogram and radial description are mainly due to very poor segmentation results. Those results are probably due to sudden and big motion changes between consecutive frames, especially during falling down and during sitting and standing.

DISSCUSSIONS AND CONCLUSIONS

This paper outlines a set of computer vision methods for human posture and behavior analysis. The method starts with the extraction of blobs, then personal metrics and behavior profiling. Unlike previous studies, the proposed method only uses frame-by-frame posture estimation for activity reasoning, thus avoiding the use of explicit complex tracking. Although the proposed posture analysis method performs well, it remains to be qualitative and view-dependent. Our extensive experiments on classifying both postures and views simultaneously have shown that an accuracy of 65% can be obtained with the current system. By the use of view-independent features, however, it is expected that significant improvement can be achieved.

REFERENCES

1. Hausdorff, J.M., Rios, D.A., and Edelberg, H.K. 2001. Archives in Physical Medicine and Rehabilitation, 82, 1050-1056

2. Verghese, J., Lipton, R. *et al*. 2002. The New England Journal of Medicine, 347(22), 1761-1768

3. Korhonen, Parkka, J., Van Gils, M. 2003. IEEE Engineering in Medicine and Biology Magazine, May/June, pp. 66-73

4. Nait-Charif, H., McKenna, S. 2004. Intl. Conf. on Pattern Recognition

5. Gao, J., Wactlar, H., Hauptmann, A. 2004. Intl. Conf. Automatic Face and Gesture Recognition, 699-704

6. Rubenstein L., Trueblood, P. 2004. Annals of Long-Term Care: Clinical Care and Aging, 12(2), 39-45

7. Tao, Y., Hu, H. 2003. Chinese Automation & Computing Society Conf. in the UK, 343-448

8. Wren, C., Azarbayejani, A., Darrell, T., Pentland, A. 1997. IEEE Trans. Pattern Analysis and Machine Intelligence, 19(7), 780-785

9. Fujiyoshi H., Lipton, A. 1998. Workshop on Applications of Computer Vision

10. Haritaoglu, I., Harwood, D., Davis, L. 2000. IEEE Trans. Pattern Analysis and Machine Intelligence, 22 (8), 809-830.

11. Horn, B., Schunck, B. 1981. Artificial Intelligence, 17, 185-203

12. Wang, L., Hu, W.M., Tan, T. N. 2003. Pattern Recognition, 36 (3), 585-601