The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Wearable and Ambient Sensor Fusion for the Characterisation of Human Motion

Douglas McIlwraith, Julien Pansiot and Guang-Zhong Yang

Abstract— **Home monitoring plays an important role within pervasive healthcare, particularly for monitoring the elderly and patients with chronic disease. For assessing activities of daily living, one of the most challenging problems for research remains that of accurate *transition detection* and *characterisation*. Early detection of a change in these transitions, such as difficulty getting up from a seated position, can be an indicator of further complications which often precede a fall. Such changes can also accompany early stage neurological disorders which can be treated effectively to improve quality of life. In this paper, we present a system for the accurate characterisation of motion based upon the fusion of ambient and wearable sensors. A probabilistic, privacy respectful method for the extraction of detailed 3D posture information is proposed and fusion with an ear-worn accelerometer and gyroscope is discussed. We present results detailing high accuracy in the recognition of complex motions over four subjects.**

I. INTRODUCTION

Postural change is an important factor in the care and rehabilitation of elderly patients. Changes in gait can signal the onset of neurological diseases such as Parkinson's [1], whilst changes in how routine tasks are performed – such as getting out of a chair – can indicate an increase in frailty and a loss of muscle strength which can increase the chance of a fall [2]. By characterising motion within the home it becomes possible to provide a more accurate picture of elderly patients' health and wellbeing.

Unfortunately, there are various difficulties associated with monitoring specific activities, particularly those related to *transitions*. Firstly, these motions represent only a small percentage of daily activity. Secondly, they can occur anywhere within the home and the subject may be oriented in any direction. Finally, a high level of subject detail and context is required. In this paper, we present a motion characterisation framework based on the fusion of both wearable and ambient sensors. We use a probabilistic method for detailed posture extraction which utilises only four cameras and provides a detailed, but privacy respectful, rotationally invariant signature of posture. Data from a wearable accelerometer is fused with optical flow to provide both local and global motion detection which is robust to the relative positioning of camera and subject. We further discuss our method for the fusion of gyroscope and video data which provides both the advantages of detailed local posture features along with a global estimation of posture through an ear mounted gyroscope. The results derived demonstrate that complex motions can be detected and distinguished between accurately.

Douglas McIlwraith is with the Wolfson/MIC Laboratory, Imperial College, London, UK. dm05@doc.ic.ac.uk

Julien Pansiot and Guang-Zhong Yang are with the Institute of Biomedical Engineering, Imperial College, London, UK. {jpansiot,gzy}@doc.ic.ac.uk

II. BACKGROUND

Motion characterisation has been the subject of extensive research – generally separated into one of two areas: sensor based and image based. These can be further understood in terms of model based and feature based systems. Bidargaddi *et al.* [3] adopt a wavelet based approach to extract sit-to-stand transitions from a waist worn accelerometer, using the time taken to transition as an indicator of subject mobility, whilst in [4] a trunk worn accelerometer was used to gather sit-to-stand motion data before the complexity of the acquired signals is determined using a fractal measure – the approach presuming that more able subjects will have a less complex signal. Culhane *et al*. [5] provide a review on the use of accelerometers for elderly care with a particular focus on gait, sit-to-stand transitions and activity monitoring.

Image centred, model based systems presume that motion is subject to particular constraints – defined by the specific model used, and these can be further separated into 2D and 3D variants. Of the former, a *star* approximation of a skeletal whole body posture model has been determined through extremity detection and topology formation from the background segmented centre of mass – analysis of lower extremity movement should then correspond to that of the legs [6]. Whilst 2D approaches have the advantage that few cameras are required, projection of 3D posture into 2D space loses information. Projection is also viewpoint dependent, creating inconsistencies when only a single model of motion is used. More realistic models have been proposed to reconstruct 3D posture [7], [8], however they typically require many synchronised cameras.

Feature based systems also come in both 2D and 3D variants and are characterised by the absence of underlying models of motion. Petkovic *et al.* [9], propose a 2D system where specific features associated with different tennis strokes are extracted from 2D silhouettes of players. Such a system is especially suited to tennis due to the overall consistency of player orientation. Consequently, the existence of particular image features can be associated with individual strokes. Yamato *et al.* [10] utilise a *mesh feature* to characterise the distribution of foreground/background within an image, reducing this to a series of *codewords* for temporal recognition. Whilst 3D feature based systems share some of the problems inherent to 3D model based systems, they preserve much more information regarding the evolution of posture than a 2D approach, since projection onto the

image plane is not performed. Chu *et al.* [11] present a 3D system which reconstructs a 3D visual hull and describes the evolution of hierarchical shape descriptors using multiple Hidden Markov Models (HMMs) and there are various other shape descriptors within the literature which can be used to this effect [12], [13].

III. System Overview

The system introduced in this paper takes a 3D feature based approach and fuses extracted decisions with those obtained using a wearable sensor. We introduce a *probabilistic visual hull* for deriving rotation invariant metrics of posture. The sensor used is the ear mounted e-AR [14]. During training of our algorithm, the accelerometer output is combined with the optical flow from the surrounding vision sensors to provide a viewpoint independent estimate of subject motion. Key postures are then learnt with a HMM and the relative likelihood of these postures over time is used to characterise posture evolution, with the gyroscope utilised to estimate subject attitude.

A. Probabilistic Visual Hull

To create a traditional visual hull, we first collect a set of reference images simultaneously before combining these to generate an estimate of subject posture. Images are acquired using multiple video sensors, whilst an approximate 3D model of the subject can be generated using the *Image-Based Visual Hulls* [15] method. This approach consists of two stages. First, the subject is segmented from the background of each individual reference image, typically using an adaptive statistical background model [16]. Secondly, assuming full calibration, the two-dimensional binary silhouettes are reprojected into 3D using epipolar geometry. The 3D model is finally generated by sampling the intersection of the 3D reprojected silhouettes.

Within the scenario proposed above, a key problem lies in the hard binary segmentation performed early in the 2D image space. If one reference image is erroneously segmented, the error will propagate to the 3D model. To tackle this, we adopt a probabilistic fusion based approach. Previously, Franco and Boyer [17] have presented a solution using Bayesian inference. Instead of labelling the foreground and the background in the image space, this operation is delayed until the 3D reprojection stage. The state of each individual voxel is then inferred from the foreground probabilities of every image. More recently, Jean-Yves Guillemaut *et al* [18] have taken a similar approach by performing simultaneous segmentation and 3D reconstruction.

B. Subject-Centric Shape Descriptor

In order to generate a viewpoint-invariant representation of the probabilistic visual hull, a canonical coordinate system is defined. By definition, the canonical coordinate system must be invariant to the relative position and orientation of the subject with respect to the video sensors, thus a subject-centred coordinate system is used. A Kalman filter [19] tracks the user's head – initialised at the beginning of the video sequence using a naïve detector which assumes the subject is standing. Whilst the motion of the head is not linear and therefore cannot be perfectly modelled by a Kalman filter, its use can help to reduce noise.

Two general options can be taken to ensure a rotation-independent representation of the model. A model-normalised orientation can be determined through the use of Principal Component Analysis (PCA). Such normalisation is sensitive to noise, as the canonical orientation depends on the model sampling quality – alternatively, a rotation-invariant descriptor can be used. For the current application, the 3D descriptor needs to be rotation-invariant in the horizontal plane, but not in the vertical. For this reason, a vertical cylindrical coordinate system centred in the subject's head is used to project the 3D model, as illustrated in Figure 1. The maximal extent of the model is projected on the cylinder, building a cylindrical 2D depth map, which is then composed over 360 degrees to create a rotationally invariant signature.

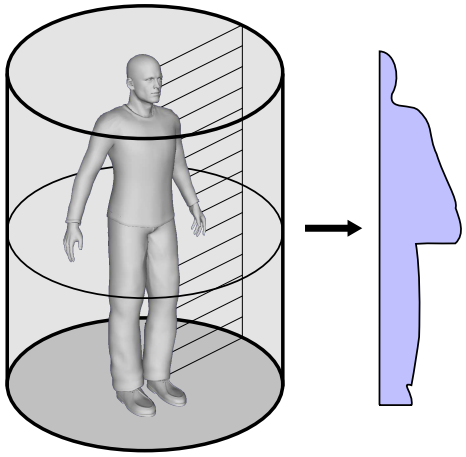

Fig. 1. Cylindrical maximal extent sampling performed on a subject 3D model. Left: 3D human model being sampled. Right: resulting sampling for one given angle. Note that this depth map is composed over all angles to obtain a rotationally invariant signature of pose.

IV. System Operation

Initially, the system needs to be trained to recognise key postures of a given motion along with the associated angular variance. This does not require detailed labelling, as we leverage our motion detection subsystem to automatically select posture subsets. Postures are chosen after periods of motion have been observed, through analysis of fused accelerometer and optical flow data. We assume that after periods of motion there is an increased likelihood that posture has significantly changed. These postures are characterised by training a HMM with Gaussian observed values on the 2D signatures returned by our posture detector. The angle travelled between postures is then recorded using the gyroscope and motion recognition occurs through the analysis of posture detector interaction, and the angular variance during motion. Figure 2 provides an overview of the system.

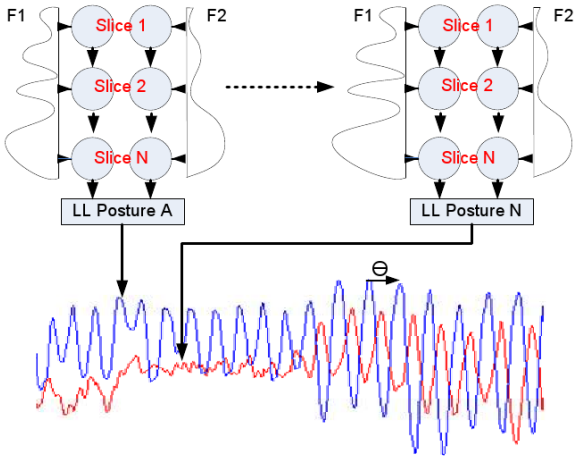

Fig. 2. An overview of the motion recognition system. Multiple features are generated from the projection of the 3D visual hull created using our probabilistic method, Section III-A. For each posture selected through motion analysis, Section IV-B, a 2D posture signature is learnt using a Hidden Markov Model with Gaussian observed values. We utilise both the relationship between confidence in recognition and the angular variance between postures to characterise movements. In the example provided above, we witness the antiphase behaviour of two posture detectors as a motion is performed contiguously eight times toward the end of the data set. The angular distance travelled by the user during a motion is given by θ.

A. Fusion of Accelerometer and Optical Flow Data

In order to provide a viewpoint independent estimate of subject motion, the average unit normalised intensity of the sample optical flow field, M is calculated for each frame, and for each camera using the Horn-Schunck method [20]. N_c is equal to the number of cameras used for posture recognition, R_k is the resolution of the optical flow field sampling for camera k and $OF_k(i,j)$ is the unit normalised optical flow at position (i,j) in camera k's field of vision. Due to the use of normalised optical flow, M is necessarily unit normalised.

$$M = \frac{\sum_{k=1}^{N_c} \sum_{j=1}^{R_k(y)} \sum_{i=1}^{R_k(x)} OF_k(i,j)}{\sum_{k=1}^{N_c} R_k(x) R_k(y)} \quad (1)$$

This is then combined with the unit normalised root mean square acceleration, A, over all three axis collected using the wearable sensor. A_x represents the acceleration obtained on the x axis of the sensor, whilst A_{sat} is the maximal (or saturated) value on all axes of the accelerometer.

$$A = \frac{\sqrt{{A_x}^2 + {A_y}^2 + {A_z}^2}}{\sqrt{3 \times {A_{sat}}^2}} \quad (2)$$

These are combined to create a single motion metric, V, using mixing parameter ω which expresses the relative importance of global/local motion in posture selection;

$$V = (1-\omega)M + (\omega)A \quad (3)$$

To improve noise resilience, V is further processed using a Savitzky Golay [21] smoothing filter, which removes high frequency noise, whilst maintaining key features of the motion distribution. Turning point maxima and minima are then extracted.

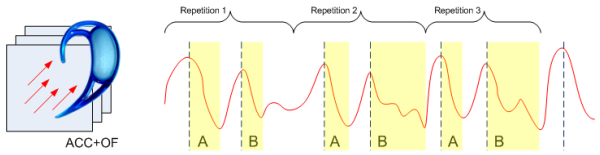

Fig. 3. Demonstrating output for the combined motion detector. Areas marked in yellow are those extracted by our algorithm as corresponding to a change in posture. Signatures returned from the posture detector during these periods are then used to train individual posture recognition detectors – in this case for posture A and posture B. How these detectors interact over time is used along with the angular motion detected by the gyroscope, to characterise the entire motion, see Figure 2.

B. Posture Selection & Recognition

In order to select key postures for recognition during the training phase, we first perform peak detection on V to extract both global and local motion maxima and minima. We assume that the period between maxima and minima characterises a possible new posture, since motion necessarily precedes posture change (although it does not imply it). A key challenge is determining the resolution at which motion detection should operate. Too high and every small motion performed by the subject will trigger the posture recognition subsystem to characterise and learn a new posture, too low and key changes to posture will be overlooked. Furthermore, each motion must be observed in each training instance, with similar characteristics. We demonstrate this problem in Figure 3, which shows the motion trace for three repetitions of the same complex motion. We note firstly that each repetition is characterised by a large maxima, at the start of section A, with a smaller maxima, at the start of section B. For each repetition, these represent the largest, and second largest motion maxima respectively – with the low or decreasing motion period directly after, corresponding to a new posture held by the subject under observation.

Our algorithm for posture selection operates as follows. Firstly, after hand segmentation of the performed activities, all motion maxima are extracted. Secondly, a proportion of these are discarded dependent upon a sensitivity parameter, α, supplied to the algorithm. Each maxima is then ranked by magnitude, with the algorithm ensuring that these maxima occur consistently in the same order. Repetitions deviating from this pattern are discarded, provided no more than a specified fraction of performances are removed from consideration. If this threshold is passed, the sensitivity parameter, α is decreased and the algorithm re-run. The maxima left now represent our estimate of when posture has changed dramatically. In order to mark these regions we select the global minima between the remaining maxima – leaving the yellow areas of Figure 3 for cylindrical feature extraction and posture learning.

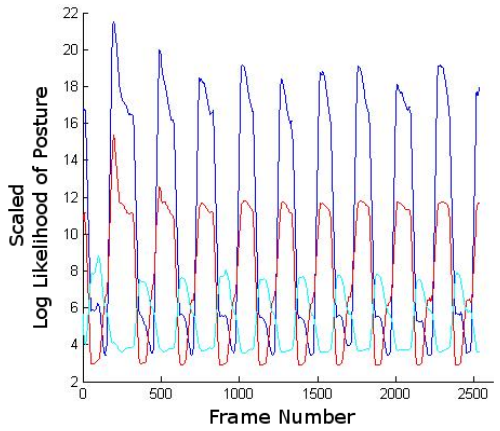

Fig. 4. Illustrating interaction between three posture recognition HMMs trained on the maximal extent feature for the action "pick up object from the floor, place it on a high object." This sample is taken from video in which the action was performed ten times. HMMs were trained only upon the data from the first five attempts, with the three postures selected using the algorithm outlined in Section IV-B. Note, log-likelihood has been artificially scaled and transformed into the positive axis.

C. Posture Evolution

In order to recognise a particular motion, we rely upon the interaction of posture detectors. This encodes the relative similarity between postures in a compact manner. Consider Figure 4. This shows the output from three posture detectors for the action "pick up object from the floor, place it on a high object." Postures were automatically selected using the algorithm outlined in Section IV-B. The data is from a set of ten repetitions of the action, where only the first five were used to train the posture detector HMMs.

It can be seen that there is clear evidence of a repeated signature and thus it appears possible to recognise a rapid transitional motion through posture detector interaction. There is an obvious anti-phase relationship between dark blue and light blue signatures, with the red minima preceeding the dark blue minima on each repetition. In order to characterise this, a further HMM using three observed Gaussians can be utilised to generalise this *temporal* relationship, in the same manner that the postural recognise generalise the *spatial* relationship with the user and their environment.

D. Fusion of Ambient and Wearable Sensing Modalities

In this system, fusion occurs at the *decision level*. After training, cylindrical features are continually exposed to spatial HMMs to determine the likelihood of a given posture, and the output passed to temporal HMMs to extract the likelihood that a given temporal pattern has occurred. This framework exists for each motion within the system, and thus an ensemble of detectors must be composed in order to reach a decision. Further to this a gyroscope subsystem utilises five 3D Gaussians – to characterise the variance on each dimension of the gyroscopes' motion for each action.

Depending upon data alignment, not all vision based classifiers are guaranteed to respond with a likelihood. In this scenario, we assume that since the data is out of range, the possibility of a match with the associated motion is negligible, so this motion is removed from consideration. Our algorithm proceeds by calculating the number of responding classifiers. If greater than one T in Equation 4 is calculated, where $\mathbb{C}_{tr}$ is a set of classifier outputs containing a likelihood entry for each responding classifier. If only a single classifier responds, the vision system is ignored and the gyroscope is used for classification, since there is no possibility to determine comparative confidence between motions.

$$T = \frac{|max(\mathbb{C}_{tr})|}{|max(\mathbb{C}_{tr}) - max(\mathbb{C}_{tr} \backslash max(\mathbb{C}_{tr}))|} \tag{4}$$

In order for the output corresponding to $max(\mathbb{C}_{tr})$ to identify the motion being performed, T in Equation 4 must be minimised (assuming negative log-likelihood). When minimised, the following is assumed. Firstly, that the individual classifier is confident in its selection. This is determined by minimisation of the numerator. Secondly, that there is a suitable amount of discrimination between other classifiers – consider that it would not be possible to discriminate between motions if the output from the respective classifiers was very similar. In order to account for this, the second largest output from a motion classifier is obtained, and subtracted from the largest output – the denominator of Equation 4. To provide a level of both individual classifier confidence and distance from the nearest potential class we permit the vision sensors' decision to be used only if $T < 1$. If this constraint is not met, the decision returned using the gyroscope data is referenced.

V. Experiment Procedure

In order to validate our approach, we provide results that demonstrate the ability of our system to determine several complex motions: 1) Reach for an item on the floor, place this at waist level; 2) Reach for an item above head level, hold this forward at waist level, hold this to the right hand side; 3) Reach for and item on the floor and place this above the head; 4) Dress self with a shirt; 5) Brush teeth. These activities were performed at a fixed location within a single room such that the coverage from each of the four cameras used was favourable.

Subjects were initially asked to perform each of the above activities ten times in a constrained manner, pausing at the steps outlined in Figure 5 – we call these *static* attempts. This allows our motion detection subsystem described in Section IV-B to determine the key sub-postures for these motions. This can be considered a training step for the system. Subjects were then asked to perform the same activities a further ten times in a *fluid* fashion *i.e.* with no requirement to stop.

To test the efficacy of our system, we perform two experiments. Firstly, for each subject and each motion, the first five *static* attempts were used to train the system – including sub-posture selection, posture detector interaction and gyroscope variance clustering. The trained system was then exposed to the remaining five static attempts along with the fluid attempts for that subject. For this experiment, hand

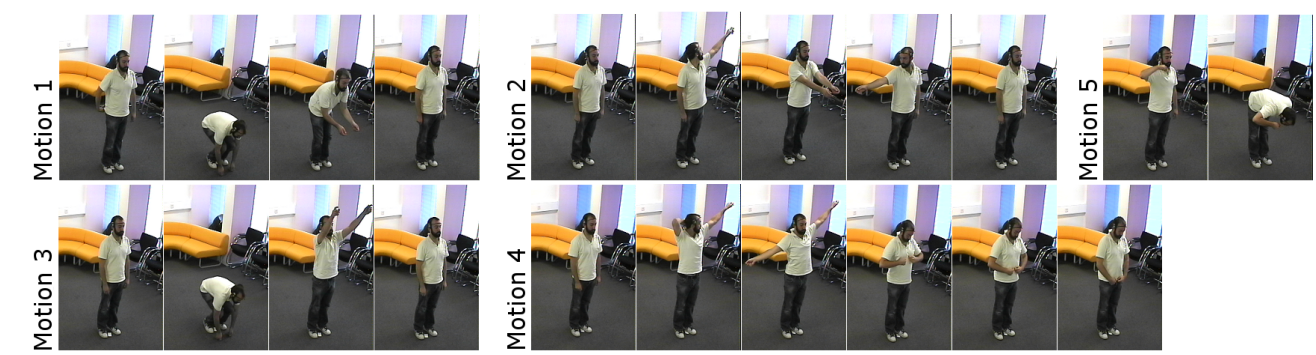

Fig. 5. Illustrating sub-steps of the motions performed during experimentation.

segmentation of the motions was utilised. Secondly, in order to test the ability to automatically select candidate sequences for motion characterisation (auto-segmentation), we train a posture detector on the first five start and finish postures for each motion using cylindrical data up to five frames away. For each experiment, per-subject training and testing was performed, with the average result over four subjects presented. In both experiments, only the maximal extent feature was extracted from the visual hull, Figure 1, and subsequently used for posture recognition.

VI. RESULTS & DISCUSSION

Table I provides the accuracy of the system in motion recognition when exposed to *all* remaining motion attempts, along with the *static* and *fluid* motions alone, for gyroscope only, vision only and the fused system. NA appears as an entry for *Motion 5* since, as there is no difference between *static* and *fluid* attemps, only ten performances were captured. Tables II and III illustrate the accuracy and precision of the segmentation subsystem respectively.

A. Classification Accuracy

From the results derived, we can observe the way in which the motion is performed appears to have an effect on the classification accuracy, regardless of the modality used. *Fluid* motions are almost always distinguished at a lower rate than *static* counterparts. In the case of the gyroscope, this may be explained by an actual change in the variance of the signal – since under static attempts the user will spend a longer time stationary. This is a limitation of the approach that needs to be addressed in subsequent research. In the case of the vision subsystem, this may again be explained with respect to a deformation of the posture likelihood curves in the time dimension. Whilst the temporal HMM should be able to cope with this, the implementation discussed here uses time as a feature in posture recognition, reducing time warping ability.

In general, it appears that the gyroscope and vision subsystem are complementary. *Motion 2* appears well recognised by the gyroscope, but poorly determined by the vision subsystem, whilst the opposite is true of *Motion 5*. After fusion, there is a notable increase in *Motion 3 All*, and *Motion 2 Static*, both of which outperform the results from either the wearable or vision based system alone. In general, fusion offers an overall increase in accuracy of **6.4%** over a vision based system, and **17.2%** over a gyroscope alone.

(%)	1/2s	1s	2s	4s	8s	16s
Motion 1	7.83	12.30	24.35	58.28	73.73	73.73
Motion 2	4.50	13.83	50.23	65.38	74.60	75.80
Motion 3	5.68	10.33	28.95	67.13	74.00	75.25
Motion 4	9.53	12.65	29.30	52.73	74.65	77.65
Motion 5	17.63	24.25	48.90	50.90	55.53	55.53
Average	9.03	14.67	36.35	58.88	70.50	71.59

TABLE II

THE PERCENTAGE OF BOUNDARY POINTS MATCHED AS THE SEARCH AREA IS INCREASED FROM 1/2 SECOND TO 16 SECONDS. AVERAGED OVER 4 SUBJECTS.

B. Segmentation Subsystem

Given the output from the auto-segmentation algorithm, we provide results showing both the accuracy and precision of segmentation, Tables II and III, using a greedy matching algorithm as the allowable distance between true and detected segmentation boundary is increased. In the first table we see that only **9%** of boundaries are recognised to within half a second, but as the distance is increased, this reaches **71.6%** of boundaries with a sixteen second search space. Most notable, however, is that whilst the allowable search space is increased, the majority of matches between detected segments and true boundaries are close, with an average frame distance of 63.04 at sixteen seconds. Since the frame rate is 25 per second, this corresponds to an average distance of around 2.5 seconds – a perfectly acceptable approximation which can be passed to the motion characterisation algorithm as a starting point for a local search.

VII. CONCLUSION

In this paper, we have presented a framework to characterise posture evolution with high accuracy through the fusion of both ambient and wearable sensing modalities. We have validated our approach to sensor fusion and shown preference over each individual sensor alone. A system for

(%)	Gyroscope			Vision			Fusion		
	All	Static	Fluid	All	Static	Fluid	All	Static	Fluid
Motion 1	32.3	50.0	22.5	68.0	80.0	67.1	62.3	75.0	55.2
Motion 2	85.0	80.0	87.5	51.5	55.0	50.0	75.5	**90.0**	68.8
Motion 3	68.3	80.0	62.6	87.2	100.0	97.2	**98.3**	**100.0**	93.8
Motion 4	62.8	80.0	56.3	37.7	35.0	38.3	40.1	55.0	34.6
Motion 5	41.7	NA	NA	100.0	NA	NA	**100.0**	NA	NA

TABLE I

AVERAGE CLASSIFICATION ACCURACY OVER FOUR SUBJECTS. *All* DENOTES THAT THE CLASSIFIER WAS TRAINED ON THE FIRST 5 PERFORMANCES, WITH THE REMAINDER USED FOR TESTING (A MIXTURE OF STATIC AND FLUID PERFORMANCES). *Static* AND *fluid* DENOTE THE SAME CLASSIFIER TRAINING WITH TESTING OCCURRING ON THE REMAINING STATIC AND FLUID ATTEMPTS RESPECTIVELY. WHERE FUSION PERFORMS AS WELL AS, OR BETTER THAN BOTH MODALITIES ALONE, ENTRIES ARE GIVEN IN BOLD TYPE.

(Frames)	1/2s	1s	2s	4s	8s	16s
Motion 1	5.75	8.33	28.80	51.55	67.95	67.95
Motion 2	3.33	9.98	28.63	49.83	57.63	63.29
Motion 3	1.85	8.00	35.72	52.75	58.45	65.63
Motion 4	3.58	5.45	23.35	51.05	72.68	81.55
Motion 5	6.75	10.06	20.25	26.98	36.78	36.78
Average	4.25	8.36	27.35	46.43	58.68	63.04

TABLE III

THE AVERAGE DIFFERENCE (IN FRAMES) BETWEEN TRUE AND DETECTED BOUNDARY POINTS AS THE SEARCH AREA IS INCREASED FROM 1/2 SECOND TO 16 SECONDS. AVERAGED OVER 4 SUBJECTS.

auto-segmentation has also been provided to extract candidate sequences for recognition and learning. The high level of subject detail extracted by cylindrical projection enables complex postures to be detected in a privacy respectful manner – a property which is essential for home monitoring environments. In further work we intend to investigate additional features of cylindrical projection which can be readily considered by spatial HMMs with increased dimensionality. We also intend to further validate our probabilistic approach to convex hull generation. The extent to which this method can provide improved convex hulls with fewer cameras than traditional approaches is of key concern for home monitoring environments. In this vein, we also plan to investigate the quality of motion characterisation with varying degrees of camera coverage *i.e.* at different locations within a room, and within different rooms with varying camera configurations.

REFERENCES

[1] D. B. Calne, B. J. Snow, and C. Lee, "Criteria for diagnosing parkinson's disease," *Annals of Neurology*, vol. 32, no. 1, 1992.

[2] L. A. Lipsitz, P. V. Jonsson, M. M. Kelley, and J. S. Koestner, "Causes and correlates of recurrent falls in ambulatory frail elderly," *Journal of Gerontology*, vol. 46, no. 4, pp. 114–122, 1990.

[3] N. Bidargaddi, A. Sarela, J. Boyle, V. Cheung, M. Karunanithi, L. Klingbei, C. Yelland, and L. Gray, "Wavelet based approach for posture transition estimation using a waist worn accelerometer," in *Proceedings of the 29th Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2007, pp. 1884–1887.

[4] R. Ganea, A. Paraschiv-Ionescu, A. Salarian, C. Büla, E. Martin, S. Rochat, C. Hoskovec, C. Piot-Ziegler, and K. Aminian, "Kinematics and dynamic complexity of postural transitions in frail elderly subjects," in *Proceedings of the 29th Conference of the IEEE Engineering in Medicine and Biology Society*, 2007.

[5] K. M. Culhane, M. O'Connor, D. Lyons, and G. M. Lyons, "Accelerometers in rehabilitation medicine for older adults," *Age and Ageing*, vol. 34, no. 6, pp. 556–560, 2005.

[6] H. Fujiyoshi and A. J. Lipton, "Real-time human motion analysis by image skeletonization," in *Proceedings of the 4th IEEE Workshop on Applications of Computer Vision*, 1998, p. 15.

[7] I. C. Chang and C.-L. Huang, "The model-based human body motion analysis system," *Image and Vision Computing*, vol. 18, no. 14, pp. 1067 – 1083, 2000.

[8] P.-B. Wieber, F. Billet, L. Boissieux, and R. Pissard Gibollet, "The HuMAnS toolbox, a homogenous framework for motion capture, analysis and simulation," in *International Symposium on the 3D Analysis of Human Movement*, Valenciennes France, 2006.

[9] M. Petkovic, W. Jonker, and Z. Zivkovic, "Recognizing strokes in tennis videos using hidden Markov models," in *Proceedings of the IASTED International Conference on Visualization, Imaging and Image Processing (VIIP 2001)*, 2001, pp. 512–516.

[10] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden markov model," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun 1992, pp. 379–385.

[11] C.-W. Chu and I. Cohen, "Posture and gesture recognition using 3D body shapes decomposition," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, p. 69.

[12] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Proceedings of the Eurographics Symposium on Geometry Processing*, 2003.

[13] J.-S. Pan, H. Luo, Z.-M. Lu, and J.-C. H. Chang, "A new 3D shape descriptor based on rotation," in *Proceedings of the 6th International Conference on Intelligent Systems Design and Applications*, 2006.

[14] B. Lo, L. Atallah, O. Aziz, M. Elhelw, A. Darzi, and G.-Z. Yang, "Real-time pervasive monitoring for post-operative care," in *IFMBE proceedings of the 4th International Workshop on Wearable and Implantable Body Sensor Networks*, March 2007, pp. 122–127.

[15] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan, "Image-based visual hulls," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 2000, pp. 369–374.

[16] D.-S. Lee, "Effective gaussian mixture learning for video background subtraction," *PAMI*, vol. 27, no. 5, pp. 827–832, 2005.

[17] J.-S. Franco and E. Boyer, "Fusion of multiview silhouette cues using a space occupancy grid," in *ICCV*, vol. 2, 2005, pp. 1747–1753.

[18] J.-Y. Guillemaut, J. Kilner, and A. Hilton, "Robust graph-cut scene segmentation and reconstruction for free-viewpoint video of complex dynamic scenes," in *ICCV*, 2009, pp. 809–816.

[19] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[20] B. K. Horn and B. G. Schunck, "Determining optical flow," Massachusetts Institute of Technology, Cambridge, MA, USA, Tech. Rep. AIM-572, 1980.

[21] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.