Stochastic Temporal Models of Human Activities

Michael Walter † , Shaogang Gong ‡ and Alexandra Psarrou †
†Harrow School of Computer Science,University of Westminster
Harrow HA1 3TP, UK {zeoec,psarroa}@wmin.ac.uk
‡Department of Computer Science Queen Mary and Westfield College
London E1 4NS, UK sgg@dcs.qmw.ac.uk

Abstract

Human activities are characterised by the spatio-temporal structure of their motion pattern. Such structures are probabilistic and often rather ambiguous. Modelling such spatio-temporal structures as static templates can be very sensitive to noise and cannot capture variations in observation measurements caused by different subjects performing the same act. In this paper we introduce the concept of modelling temporal structures by statistical dynamic systems using first-order Markov process descriptions. Prior knowledge is learned from training sequences and recognition is performed through continuous propagation of density distributions. Taking current observations into account to temporarily augment the learned prior leads to more accurate recognition with less computational costs.

1 Introduction

Human activities are characterised by the spatio-temporal structure of their motion pattern. The analysis of such structures offers to impart a richer understanding of a dynamic world than that available from the analysis of static scenes.

These structures are intrinsically probabilistic and often ambiguous. In general, they can be treated as *temporal trajectories* in a high-dimensional feature space representing closely correlated measurements on visual observations. For example, the spatio-temporal structure of a simple behaviour such as walking towards a telephone could be represented by the trajectory of an observation vector given by the mean position and displacement of the human body (Figure 1). Given that human behaviours and activities can be modelled by structures of temporal trajectories in a high-dimensional feature space, behaviour recognition can then be performed by matching these trajectories. Based on this general concept, recognition of gestures, for example, can be treated as the problem of matching "holistic shape" templates in a spatio-temporal feature space [1, 2]. Unfortunately, modelling temporal structures as static templates can be sensitive to noise and ambiguities. Other factors contributing to the spatio-temporal structure include (i) covariance in observation measurements, (ii) nonlinear temporal scaling, and (iii) ambiguities in temporal segmentation of a behaviour. To address such problems, the temporal structures could be modelled as stochastic processes under which salient phases of the structure are modelled as states and prior knowledge on both state distributions and observation covariances is learned from training examples. Predicting state transitions then provides more robust means to cope with time scaling and avoids the need for determining the starting and ending points of behaviours. An example of this approach is the use of Hidden Markov Models.

Hidden Markov Models (HMMs) are widely used for modelling temporal structures. They have been applied to speech recognition [3], visual focus of attention [4], learning object movement and behaviour models [5, 6] and more recently gesture recognition [7, 8]. Hidden Markov states are usually selected so as to capture the locations along observation trajectories where measurements undergo significant change. Prior knowledge in the form of state transition probabilities and conditional observation covariances are estimated from examples. However, the main disadvantages of HMMs are that (i) they can be used to estimate the probabilities for only one model at a time and (ii) they can only give an estimate of the final probability for each model.

More recently the CONditional DENSity propagATION (CONDENSATION) algorithm was proposed by Isard and Blake [9, 10]. Instead of modelling observation probabilities conditional to a finite set of discrete states, a set of probabilities for different models is continuously propagated over time. For gesture recognition, CONDENSATION has been adopted by Black and Jepson [11]. Unfortunately, the model does not use any prior knowledge on both state transitions and measurement covariances. State predictions are simply previous states plus arbitrary Gaussian noise. Consequently, a very large number of density samples (over thousands) with localised uniform distribution is needed to be initialised and then propagated over time. This is computationally expensive.

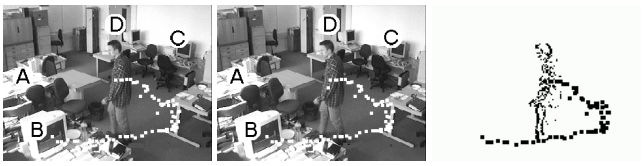

Figure 1. A behaviour of a walking person from station *B* to station *A* in an office with overlaid trajectory. Trajectories are extracted using temporal filtering.

In the following sections, we introduce a method for learning both prior knowledge and a model for recognising structures of behaviour in a state space. We illustrate the method through (i) the recognition of behaviours associated with people walking between different areas of interest in an office environment; (ii) the recognition of *communicative* gestures defined within the context of visually mediated interaction [2] and (iii) the recognition of *symbol* gestures representing alphanumeric characters. Figure 1 illustrates an example of a typical behaviour of a person walking from a station *B* to a station *A* within an office environment. Figure 9 and Figure 11 illustrate examples of gestures defined within the context of visually mediated interaction and gestures representing alphanumeric symbols respectively.

In the rest of this paper, we first introduce the concept of modelling temporal structures by statistical dynamic systems using a first-order Markov process. In Section 3 we show how this approach can be extended to (i) learn prior knowledge on both state distributions and observation covariances and, (ii) perform automatic state selection and segmentation using temporal clustering. In Section 4 we show how to continuously propagate state densities via hidden Markov states both under the constraint of the learned prior and also subject to augmentation by current visual observation. Experiments on the recognition of behaviours and gestures using this model are described in Section 5. We conclude in Section 6.

2 Propagating First-Order Markov Processes

Human activities are temporal. Markov processes can be used to describe statistical dynamic systems with temporal history by a sequence of characteristic states, capturing locations where the system undergoes significant changes, e.g. changes in speed or direction of a movement. The states are mutually connected and the system can undergo a change of state at discrete times. The transitions are based on a set of probabilities associated with each state and its history. State transitions at time (t) depend on the previous state at time ($t-1$) and their previous history. Observation probabilities $p(o_t|s_t)$ are assigned to each state, containing information about the likelihood of an observation o_t occurred in a specific state s_t.

Let the temporal structure of a behaviour or gesture be modelled by a dynamic system described by a Markov process, we now outline an algorithm that aims to both estimate the current model probability $p_t(s_t) \equiv p(s_t|O_t)$ and propagate multiple hypothesis of different models simultaneously over time. First, let us assume that the dynamics of behaviours and gestures can be largely regarded as first-order Markov processes. In other words, a state (s_t) at time (t) will only depend on its previous state at time (t–1) and is independent of its former history $S_{t-1} = (s_1, s_2, \ldots, s_{t-1})$, i.e.

$$p(s_t|s_{t-1}) = p(s_t|S_{t-1}) \tag{1}$$

Second, we assume that a conditional, multi-modal observation probability $p(o_t|s_t)$ is independent of its observation history $O_{t-1} = (o_{t-1}, o_{t-2}, \ldots, o_1)$ and is equal to $p(o_t|s_t, O_{t-1})$, the conditional observation probability given the history. We can then propagate the model (posterior) probability $p(s_t|O_t)$ based on Bayes' rule as follows:

$$p(s_t|O_t) = k_t\, p(o_t|s_t)\, p(s_t|O_{t-1}) \tag{2}$$

where $p(s_t|O_{t-1})$ is the "predicted" prior, $p(o_t|s_t)$ the conditional observation density and k_t a normalisation factor. The prior $p(s_t|O_{t-1})$ for accumulated observation history up to time ($t-1$) can be regarded as a prediction taken from the posterior at the previous time $p(s_{t-1}|O_{t-1})$ and the state transition probability $p(s_t|s_{t-1})$:

$$p(s_t|O_{t-1}) = \int_{s_{t-1}} p(s_t|s_{t-1})\, p(s_{t-1}|O_{t-1}) \tag{3}$$

Such an algorithm can be implemented using factored sampling and is known as CONDENSATION [9]. For prediction based on Equation (3), the posterior $p(s_{t-1}|O_{t-1})$ is approximated by a fixed number of state density samples. The prediction is more accurate as the number of samples increases. Samples are taken from the posterior in proportion to the state probability. This can be done by constructing a cumulative probability distribution and uniformly sampling thereafter. Some states, especially those with high weights, may be selected several times, thus leading to identical samples whilst others with lower probability are likely to be less frequently sampled if at all. New samples are predicted according to their state transition probabilities $p(s_t|s_{t-1})$ and a weight is assigned to all samples in proportion to the value of the observation density $p(o_t|s_t)$. The weighted sample set then serves as a representation for the posterior $p(s_t|O_t)$, and is suitable for sampling.

3 Learning Prior using HMMs and EM

It is clear that without using any prior knowledge on temporal structures, the state densities $p(s_t|s_{t-1})$ can only be poorly "guessed" as the previous estimation plus arbitrary noise. As a result full estimation of the prior $p(s_t|O_{t-1})$ can only be obtained through propagation of a very large

number (thousands) of samples [11]. This is both expensive and sensitive to outliers.

A solution to this problem can be derived from Hidden Markov Models (HMMs). A HMM is defined by a number of discrete states $q \in \{q_1, q_2, \ldots q_N\}$, with probabilistic transitions between states and observation probabilities for each state. A model can be fully described by a set of probabilistic parameters $\lambda = (A, B, \pi)$ where: (i) $A \in \{a_{ij}\}$ is a set of state transition probabilities, describing the probability $p(q_{t+1} = j \mid q_t = i)$ of being in a state q_i at time t and q_j at time $t+1$ where $\sum_{j=1}^{N} a_{ij} = 1$; (ii) $B \in \{b_j(o) = p(o_t|q_t = j)\}$ is the observation density distributions at all states. The observation density $b_j(o)$ can be discrete or continuous, e.g. a Gaussian mixture $b(o) = \sum_{k=1}^{K} c_k \mathcal{G}(o, \mu_k, \Sigma_k)$ with mixture coefficient c_k, mean μ_k and covariance Σ_k for the kth mixture in a given state; (iii) $\pi \in \{\pi_1, \pi_2, \ldots \pi_N\}$ is the initial probabilities of being in state i at time $t = 1$ where $\sum_{i=1}^{N} \pi_i = 1$.

By training HMMs on a set of observed trajectories of activities, prior knowledge on both the state propagation and conditional observation density can be learned by assigning the hidden Markov state transition probabilities $p(q_t = j \mid q_{t-1} = i)$ to the CONDENSATION state propagation densities of

$$p(s_t \mid s_{t-1}) \;=\; p(q_t = j \mid q_{t-1} = i, \lambda) \;=\; a_{ij} \qquad (4)$$

and the Markov observation densities given by the prior on the measurement covariance and mean at each hidden state as the observation conditional density $p(o_t \mid s_t)$

$$p(o_t \mid s_t) \;=\; p(o_t \mid q_t = j, \lambda) \;=\; b_j(o_t) \qquad (5)$$

The observation covariance given by the density function at each hidden Markov state enables the model to cope with measurement covariance and noise. This defines a sample as a vector consisting of the current Markov state q_t for a given model λ.

Learning the prior involves (i) automatic hidden state segmentation through temporal clustering, estimation of (ii) hidden state transition distribution and (iii) conditional observation density distribution at each hidden state. This can be achieved using the Baum-Welch method, an iterative method, which maximises the likelihood $P(O|\lambda)$ for a given model $\lambda = (A, B, \pi)$. Given the number of hidden Markov states to be N, learning the locations of the states (automatic temporal segmentation), their transition probabilities and the conditional observation density distributions associated with each state can be performed as follows:

1. Initialise $\pi = \{1, 0, \ldots 0\}$ and the state transition matrix A as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & 0 & 0 \\ 0 & a_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{N-1N-1} & a_{N-1N} \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

   where $a_{ii} = 1 - \frac{1}{\hat{t}}$ and $a_{i\,i+1} = 1 - a_{ii}$.

   For a first-order HMM, the average time $\hat{t}$ in a state is given by

$$\hat{t} = \sum_{n=1}^{\infty} n\, a_{ii}^{n-1} (1 - a_{ii}) = \frac{1}{1 - a_{ii}} \qquad (6)$$

   and can be estimated as the ratio between the mean trajectory duration $\hat{T}$ of a behaviour in the training set and the number of states N, $\hat{t} = \frac{\hat{T}}{N}$.

2. Use the EM algorithm over a set of M training examples $O = \{O^1, \ldots, O^M\}$ to iteratively perform temporal clustering on the states and estimate model probability distributions A, B and π.

Figure 2 illustrates the iterative process of automatic clustering of the hidden states of a *walking* behaviour going from station *C* to station *B* in an office environment. In this example, four training sequences were used. The number of hidden states is set to 12, with conditional observation density distribution set to 1.

4 Visual Observation Augmented Density Propagation

Recognition based on prior can be made more robust if current observation is also taken into account before prediction. Let us consider the state propagation density $p(s_t|s_{t-1})$ in Equation (3) to be augmented by the current observation, $p(s_t|s_{t-1}, o_t) = p(s_t|s_{t-1}, O_t)$. Assuming observations are independent over time and future observations have no effect on past states $p(s_{t-1}|O_t) = p(s_{t-1}|O_{t-1})$, the prediction process of Equation (3) can then be replaced by

$$\begin{aligned} p(s_t|O_t) &= \int_{s_{t-1}} p(s_t|s_{t-1}, o_t)\; p(s_{t-1}|O_{t-1}) \\ &= \int_{s_{t-1}} k_t p(o_t|s_t) p(s_t|s_{t-1}) p(s_{t-1}|o_{t-1}) \end{aligned} \qquad (7)$$

where $k_t = \frac{1}{p(o_t|s_{t-1})}$ and

$$\begin{aligned} p(s_t|s_{t-1}, o_t) &= \frac{p(o_t, s_t|s_{t-1})}{p(o_t|s_{t-1})} \\ &= \frac{p(o_t|s_t, s_{t-1}) p(s_t|s_{t-1})}{p(o_t|s_{t-1})} \\ &= \frac{p(o_t|s_t) p(s_t|s_{t-1})}{p(o_t|s_{t-1})} \end{aligned} \qquad (8)$$

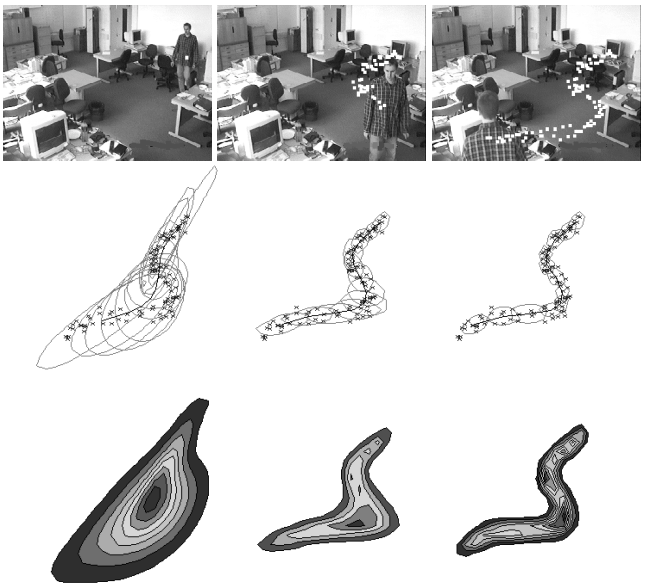

Figure 2. Learning the spatio-temporal structure of a walking behaviour going from station *C* to station *B* using HMM and EM clustering. The process of iteration is shown in the middle and bottom rows from left to right. The images in the middle row show the clustering on positions. The images in the bottom row show the corresponding density distributions over the entire structure based on the clustered hidden states and their distributions in space and time.

Given that the observation and state transitions are constrained by the underlying HMM, the state transition density is then given by

$$p(s_t|s_{t-1}, o_t) = p(q_t = j|q_{t-1} = i, o_t) = \frac{a_{ij}^{\lambda} b_j^{\lambda}(o_t)}{\sum_{n=1}^{N} a_{in}^{\lambda} b_n^{\lambda}(o_t)} \quad (9)$$

The observation augmented prediction unifies the processes of innovation and prediction in CONDENSATION given by Equations (2) and (3). Without augmentation, CONDENSATION performs a "blind prediction" based on observation history alone. Augmented prediction takes the current observation into account and adapts the prior to perform a "guided search" in prediction. This both improves the recognition rate and reduces the number of samples used for propagation.

5 Experiments

To illustrate our approach, we have applied our model to a set of extensive experiments on the recognition of walking behaviours and gestures.

Walking Behaviours: These behaviours were defined within an office environment where four stations were identified as shown in Figure 1. Eight behaviours were selected and a database containing 120 sequences was built. Each behaviour was performed five times by three different subjects and captured at 12 frames per second. Features were extracted using temporal image filtering and stored in a vector $o = (x, y, dx, dy)$ containing the center of mass, relative to the initial starting position and the displacement of the moving person between two consecutive frames. Examples for some of the behaviours and how they overlap can be seen in Figures 3 and 6. Prior knowledge on the state propagation and conditional observation density was learned using four example trajectories from each behaviour. The same examples were also used to compute mean trajectories and variances for each of the behaviours that were used as models for the CONDENSATION algorithm. The remaining 11 trajectories for each behaviour were used for recognition.

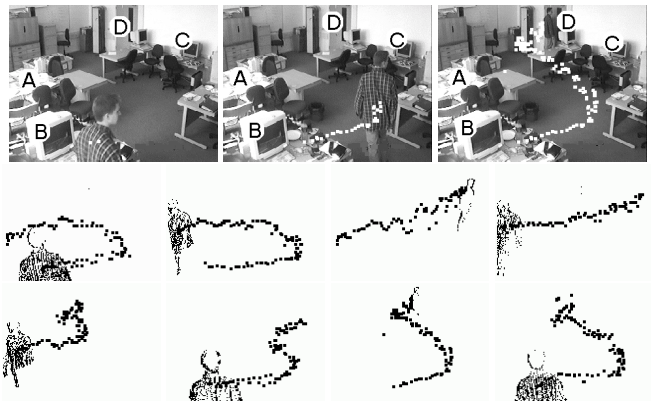

Figure 3. An example trajectory of a person walking from station *B* to station *D* (top). Examples of the eight typical behaviours that the subjects perform in an office. From right to left and top to bottom: (*A to B*), (*B to A*), (*A to C*), (*C to A*), (*D to A*), (*C to B*), (*B to D*), (*D to B*).

Figures 4 and 5 show the probability likelihoods for the behaviours shown in Figure 3. They are obtained by matching the behaviour models to novel trajectories using (i) observation augmented density propagation (top two rows), (ii) non-augmented density propagation based on prior (middle two rows) and (iii) the CONDENSATION algorithm (bottom two rows). For each algorithm we show the model probability estimation for each behaviour during the recognition process and the estimated final probability for the recognised behaviour.

It can be seen that the observation augmented propagation algorithm was able to recognise all behaviours whereas non-augmented propagation algorithm was not able to recognise behaviour *(B to A)*. In addition, the final probabilities estimated for the recognised behaviours by the non-augmented propagation algorithm are lower to that estimated by the observation augmented algorithm. The CONDENSATION algorithm was not able to recognise be-

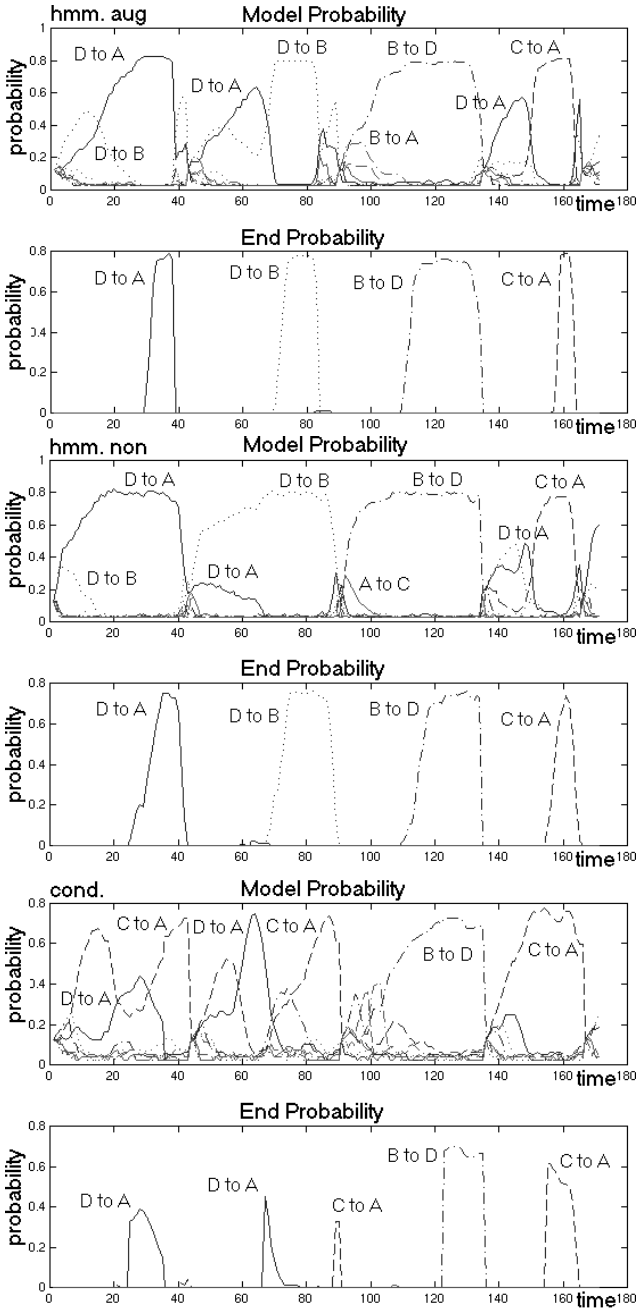

Figure 4. Behaviour likelihoods estimated for the *walking* behaviours (*D to A*), (*D to B*),(*B to D*),(*C to A*), using observation augmented density propagation (top two rows), non-augmented density propagation using prior (middle two rows) and the CONDENSATION algorithm (bottom two rows). The number of samples used for these experiments was 80.

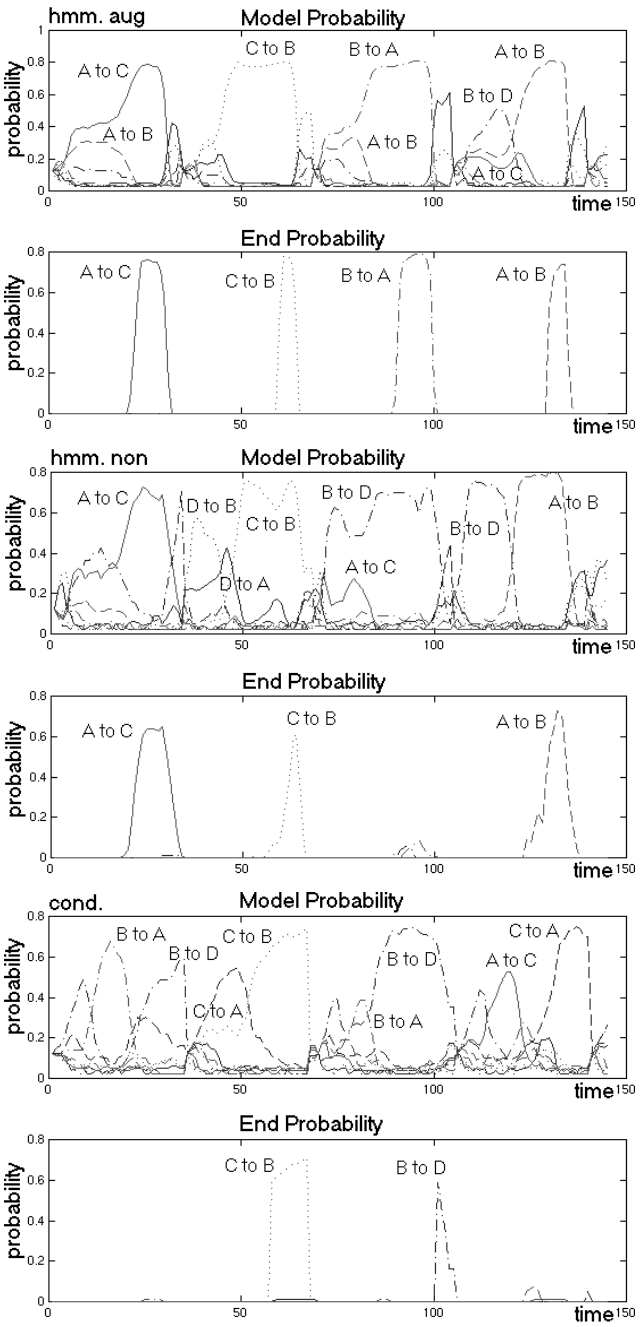

Figure 5. Behaviour likelihoods estimated for the *walking* behaviours (*A to C*),(*C to B*),(*B to A*) and (*A to B*), using observation augmented density propagation (top two rows), non-augmented density propagation using prior (middle two rows) and the CONDENSATION algorithm (bottom two rows). The number of samples used for these experiments was 80.

haviours *(A to C)* and *(A to B)* and misclassified behaviour *(D to B)* as *(D to A)* and *(C to A)* and behaviour *(B to A)* as *(B to D)*. In general, the probability estimated for a recognised behaviour by the CONDENSATION algorithm was much lower to that estimated by both observation-augmented propagation and non-augmented propagation using prior.

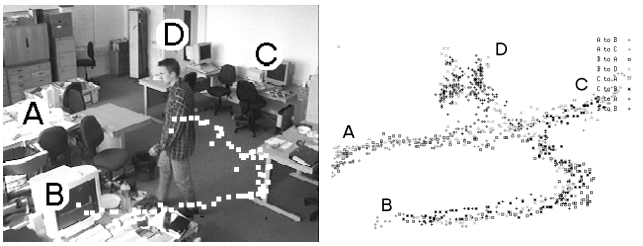

Figure 6. The office environment (left) and overlaid sample trajectories of the *walking* behaviours performed within such an environment (right).

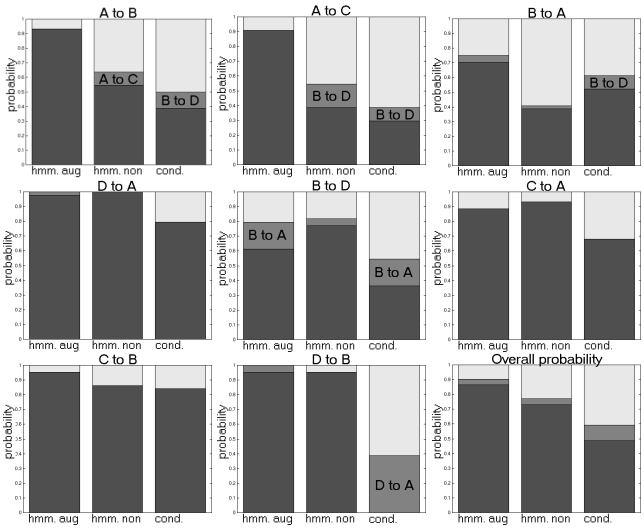

Figure 7. Recognition (black area) and misclassification (dark grey area) rate for the *walking* behaviour and all novel trajectories of the data set using (i) observation augmented density propagation, (ii) non-augmented propagation using on prior, (iii) the CONDENSATION algorithm.

The block diagrams in Figure 7 show the recognition rate (black area) and the misclassification rate (dark grey area) of each of the three algorithms for each behaviour, taking into account all 88 novel trajectories. The observation augmented density propagation algorithm (hmm.aug) recognised most of the trajectories for all behaviours but it misclassified some of the *(B to D)* behaviours as *(B to A)*. The non-augmented density propagation algorithm (hmm.non) misclassified some of the *(A to B)* and *(A to C)* behaviours, whereas the CONDENSATION algorithm misclassified some of the *(A to B)*, *(A to C)*, *(B to A)* and *(B to D)* behaviours and failed to recognise all *(D to B)* behaviours.

Figure 8 shows the recognition and misclassification rate for all walking behaviours with respect to the number of samples propagated. Using only 160 samples the results in Figure 8 (top) illustrate that estimating prior knowledge and incorporating it to our behaviour models increases the overall probability estimation by 60% compared to the probability estimation given by the CONDENSATION algorithm. Further using observation augmented propagation of density functions increases the overall probability estimation by 100% compared to the estimation given by the CONDENSATION algorithm. Compared to the non-augmented propagation algorithm the probability estimation is increased by 25%. It is also significant that the observation augmented propagation algorithm achieves a 64% recognition rate using only 40 samples compared to the 38% recognition rate achieved by the non-augmented algorithm and 29% rate achieved by the CONDENSATION algorithm using the same number of samples. The recognition rate of the observation augmented propagation algorithm can only be matched by the non-augmented algorithm when 640 samples are used. Using 640 sample, observation augmented propagation gives a recognition rate over 70%. Figure 8 (bottom) shows the misclassification rate with respect to the number of samples for the three algorithms. The graphs illustrate that the misclassification rate is much higher for the CONDENSATION algorithm compared to both observation augmented and non-augmented algorithms.

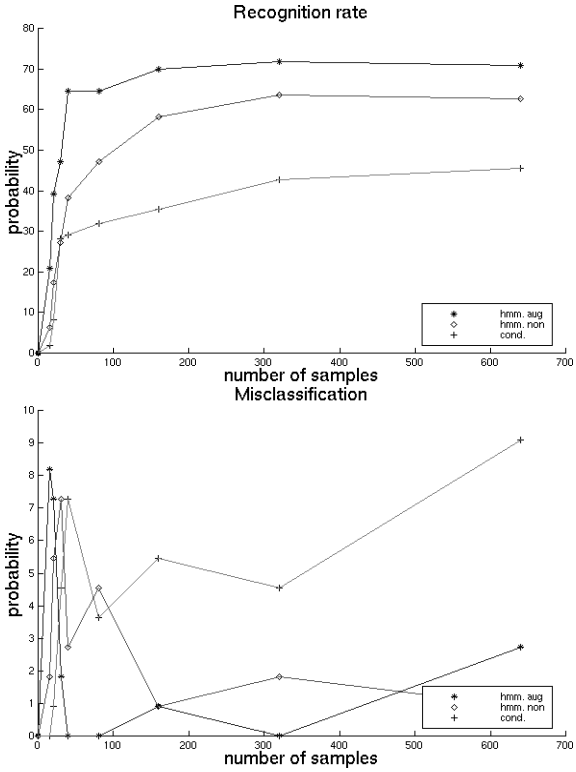

Figure 8. Recognition rate (top) and misclassification rate (bottom) for the *walking* behaviour with respect to the number of samples used for the observation augmented, non-augmented and CONDENSATION algorithm.

Gestures: In addition to the walking behaviours two sets of

Figure 9. Examples for the four *communicative* gestures in visually mediated interaction. From top to bottom: pointing left, pointing right, waving high and waving low.

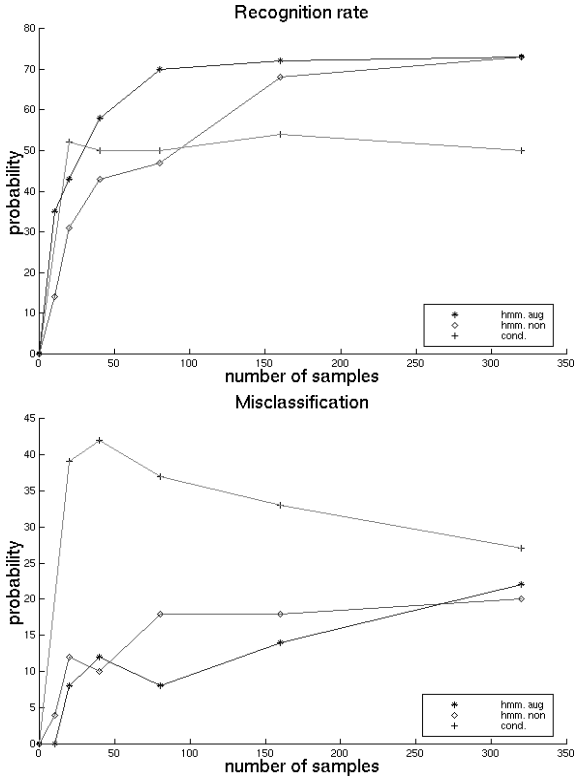

Figure 10. Recognition rate (top) and misclassification rate (bottom) for the *communicative* gestures with respect to the number of samples used for the observation augmented, non-augmented and the CONDENSATION algorithm.

gestures have been used: (i) A set of *communicative* gestures defined within the context of visually mediated interaction and they are: pointing left, pointing right, waving high up and waving low down [2] and (ii) A set of *symbol* gestures similar to that defined in [11] and they are the numerals "2", "3", "5" and letter "l". In the *communicative* gestures the object-centred position and displacement (x, y, dx, dy) of a gesture in time t is determined using moment features estimated from image motion as described in [2]. In the *symbol* gestures, in addition to image motion the skin colour of the hand was also used for extracting the observation vector (x, y, dx, dy). As a result, the *symbol* gestures are less noisy than the *communicative* gestures. A database of image sequences was collected and for the purpose of these experiments we build HMMs using four examples of each *symbol* gesture and six examples of each subject performing *communicative* gestures. Each sequence has on average 40 frames captured at 12 Hz. Examples of the *communicative* gestures and *symbol* gestures can be seen in Figures 9 and 11 respectively.

Figures 10 and 12 show the recognition and misclassification rate for the *communicative* and *symbol* gestures respectively. The results for the *communicative* gestures shown in Figure 10 (top) illustrate that using 160 samples the observation augmented and non-augmented propagation density algorithms increase the overall probability estimation by 40% compared to the probability estimated by the CONDENSATION algorithm. It is important to note that using only 80 samples the observation augmented algorithm achieves a 70% recognition rate. Figure 12 (top) illustrates that using 160 samples to recognise the *symbol* gestures, the observation augmented and non-augmented algorithms increase the overall probability estimation by 25% compared to the probability estimated by the CONDENSATION algorithm. The improved performance of the CONDENSATION algorithm is due to the less noisy nature of the *symbol* gestures.

Figures 10 and 12 (bottom) show the misclassification rate with respect to the number of samples for the three algorithms. The graphs illustrate that for both *communicative* and *symbol* gestures the misclassification rate is much higher for the CONDENSATION algorithm compared to both observation augmented and non-augmented algorithms.

6 Conclusions

We described a general framework to model and recognise temporal structures of human activities. The method is based on first-order Markov model descriptions and continuous propagation of density distributions. Prior knowledge is learned from training examples using Hidden Markov Models. Recognition is made more robust by using the current observation to temporarily adapt the learned prior. From the experiments we have shown, both the observation

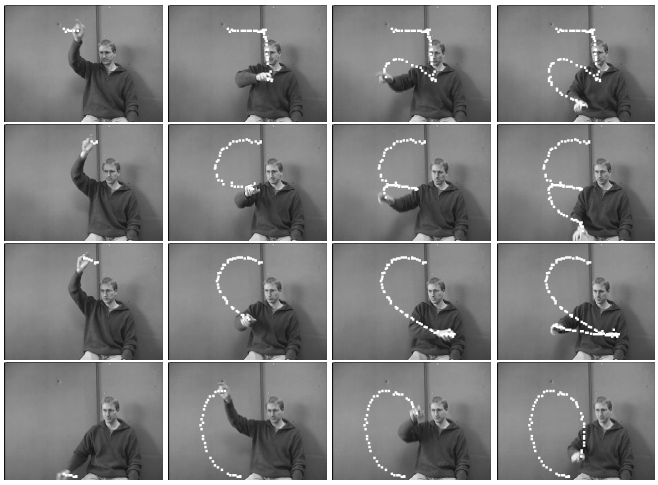

Figure 11. Examples for the four *symbol* gestures. From top to bottom: numerals "5", "3", "2" and letter "l".

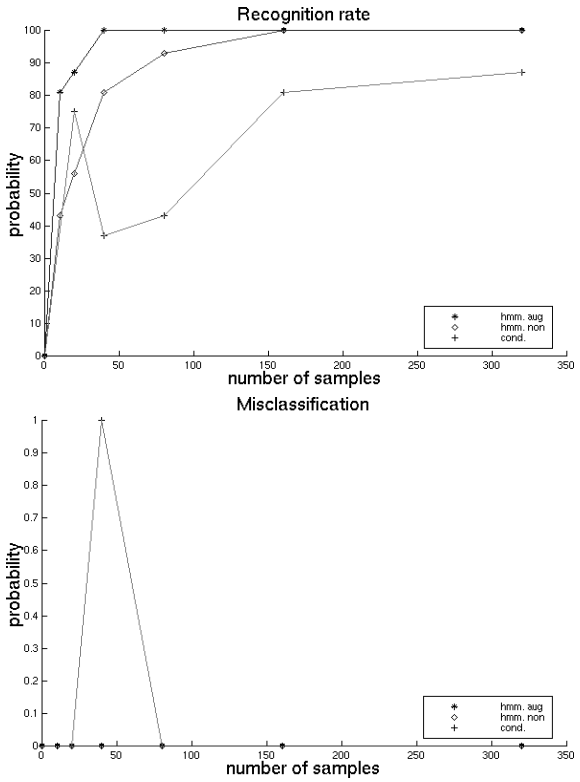

Figure 12. Recognition rate (top) and misclassification rate (bottom) for the *symbol* gestures with respect to the number of samples used for the observation augmented, non augmented and the CONDENSATION algorithm.

augmented and non-augmented algorithms achieve a much higher recognition rate compared to the CONDENSATION algorithm. The recognition rate of the *walking* behaviour is increased by 70% and the recognition rate of the *communicative* and *symbol* gestures is increased by 40% and 25% respectively. In addition we have shown that we can achieve a high recognition rate for the *walking* behaviour and gestures with using only a small number of samples (40). It is significant that such performance improvement is achieved with less computational cost since it only requires a smaller number of samples to be propagated.

References

[1] J. Davis and A. Bobick, "The representation and recognition of action using temporal templates," in *CVPR*, Puerto Rico, June 1997, pp. 928–934.

[2] S. McKenna and S. Gong, "Gesture recognition for visually mediated interaction using probabilistic event trajectories," in *BMVC*, Southampton, UK, 1998, vol. 2, pp. 498–508.

[3] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Signal Processing Series. Prentice Hall, New Jersey, USA, 1993.

[4] R.D. Rimey and C.M. Brown, "Controlling eye movements with hidden markov models," *Int. J. on Computer Vision*, vol. 7, no. 1, pp. 47–66, November 1991.

[5] S. Gong and H. Buxton, "On the expectations of moving objects," in *ECAI*, Vienna, Austria, August 1992, pp. 781–786.

[6] N. Johnson, *Learning object behaviour models*, Ph.D. thesis, School of Computer Studies, University of Leeds, England, September 1998.

[7] T. Starner and A. Pentland, "Real-time american sign language recognition from video using hidden markov models," Tech. Rep. 375, MIT Media Lab, 1995.

[8] A. Bobick and A. Wilson, "A state-based approach to the representation and recognition of gesture," *IEEE PAMI*, vol. 19, no. 12, pp. 1325–1338, December 1997.

[9] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," in *ECCV*, Cambridge, UK, April 1996, pp. 343–357.

[10] M. Isard and A. Blake, "Icondensation: Unifying low-level and high-level tracking in a stochastic framework," in *ECCV*, Freiburg, Germany, June 1998, pp. 893–909.

[11] M.J. Black and A.D. Jepson, "Recognizing temporal trajectories using the condensation algorithm," in *IEEE Conference on Face & Gesture Recognition*, Japan, 1998, pp. 16–21.