Joint Bayes Filter: A Hybrid Tracker for Non-rigid Hand Motion Recognition

Huang Fei and Ian Reid

Department of Engineering Science,
University of Oxford, Parks Road, Oxford, OX1 3DP,UK

Abstract. In sign-language or gesture recognition, articulated hand motion tracking is usually a prerequisite to behaviour understanding. However the difficulties such as non-rigidity of the hand, complex background scenes, and occlusion etc make tracking a challenging task. In this paper we present a hybrid HMM/Particle filter tracker for simultaneously tracking and recognition of non-rigid hand motion. By utilising separate image cues, we decompose complex motion into two independent (non-rigid/rigid) components. A generative model is used to explore the intrinsic patterns of the hand articulation. Non-linear dynamics of the articulation such as fast appearance deformation can therefore be tracked without resorting to a complex kinematic model. The rigid motion component is approximated as the motion of a planar region, where a standard particle filter method suffice. The novel contribution of the paper is that we unify the independent treatments of non-rigid motion and rigid motion into a robust Bayesian framework. The efficacy of this method is demonstrated by performing successful tracking in the presence of significant occlusion clutter.

1 Introduction

Many computer vision applications such as surveillance, sports, and human-computer interfacing require robust estimation and understanding object motion. Tracking of rigid and non-rigid objects such as vehicles [2] and humans [1] has been under extensive investigation in recent years. In this paper, we address the existing problems of single-view tracking and recognition of articulated hand motion in complete occlusion scenes. In these situations, simultaneous estimation and recognition of articulated hand poses can be challenging but is crucial for gesture recognition. By utilising separate image cues, we decompose complex motion into two independent (non-rigid/rigid) components. Rigid motion is approximated as the motion of a planar region and approached using a *Particle filter* while non-rigid dynamical motion is analysed by a *Hidden Markov Model (HMM) filter*. Due to its generative learning ability, hand articulation is correctly estimated even under significant occlusions.

A considerable body of work exists in hand tracking and gesture recognition. All tracking methods have some associated tracker representations, either kinematic model-based [8], [11], [5] or appearance model-based [12]. Kinematic

T. Pajdla and J. Matas (Eds.): ECCV 2004, LNCS 3023, pp. 497–508, 2004.

model-based methods construct a geometrical model before tracking. Although it can provide more information about hand configurations than 2D appearance, tracking is usually made possible with a careful initialisation and tedious model fitting/searching process, and could fail to maintain tracking where there is fast appearance deformation and/or severe occlusions; in gesture recognition, significant change point in appearance deformation is usually important for semantics interpretation. PCA appearance models [12] have the advantage of the ability to generate a new appearance using a small training set, but linear correlations impose a limit to its applications. Complex scenes and occlusion clutter pose serious distractions to all these representations. In recent years, exemplars [16] have become a popular choice for tracker representations because object models can be extracted from the original data, and the non-linear nature of articulations can be conveniently represented by a sequence of examples which exhibits first-order Markov dependence.

In addition to representation, motion estimation algorithm also contributes significantly to a tracker's performance. Besides the successful application of Particle filter[4], Dynamic Bayesian Network (DBN) [10] in visual tracking, the Hidden Markov Model (HMM) considered in this paper is another statistical time-series modelling tool [15]. Since it was firstly introduced for sign-language analysis and gesture recognition [3], there have been some research work in human motion classification using HMMs. Bregler proposed a probabilistic compositional framework [22] to classify dynamic human representation (coherence blobs), Brand [19] makes a further study in this direction by identifying kinematic structure evolution as a controlling dynamic process in 3D human motion synthesis given a set of motion capture data. Despite these efforts, the powerfulness of *dynamic* Markov model as a tracker to analyse the non-rigid motion has not been recognised until recent years [16]. In the *Metric Mixture* tracker proposed by Toyama and Blake [16], exemplars are assumed aligned in the center of probilistic mixtures. The object tracking problem then transforms to the filtering and recognition of the representative examples in the spatial-temporal domain.

In [16], two dynamic processes (global motion and shape changes) share the same joint observation density provided by the chamfer distance. This leads to an attractive homogeneity of description and implementation. However it necessitates the use of a large particle set, which must *simulataneously* represent hypotheses of both shape and position. In many applications these processes can (or even should) be decoupled, potentially leading to a more economical use of particles, and therefore greater effciency and reliability.

We propose such a decoupling for the analysis of complex hand motion. In this application, the hand motion is separated into two components assumed independent: cyclic shape variations and hand region motion. Unlike [16], each has its own dynamic process and observation density. The former is modelled by a probabilistic discrete-state system (using a Hidden Markov Model) while the latter is achieved via a particle-based region tracker. Each maintains a full pdf of its respective component of the motion, and they interact via an importance sampling function. We term this the **Joint Bayes filter (JBF)**.

Fig. 1. A video sequence of articulated hand motion.

In JBF, *monte-carlo* approximation provides an optimal localization of the hand region regardless of its articulation and clutter distraction, which reduces the distraction to the non-rigid shape inference. On the other hand, the HMM filter's output model the dynamics of the particle set and provides the importance weights of the particles, thus improves the accuarcy and efficiency of the region tracking. The overall benefit of our approach is clear: Tracking and recognition of hand articulations can be performed simultaneously; due to the independence of two observation processes, JBF tracker can withstand significant occlusion distractions and perform well in cluttered scenes.

2 The Problem

Figure 1 shows consecutive example frames from a video sequence of articulated hand motion. Changing appearances between successive frames can be significant, and thus rule out the standard assumption of constant intensity underlying optical flow based tracker [12]. Although an edge-based tracker [18] [11] [21] may perform well in normal situations, the strong occlusion clutter introduced later in the paper will damage a tracker without prior dynamics [12], or with only weak prior dynamics [18]. In this situation, even dense sampling will not help tracking articulations because of the lack of a proper *dynamical model of hand articulations*. These difficulties call for compact hand representations and stronger dynamical model based on the representations.

To deal with fast appearance deformation, strong occlusion clutter and complex scene, combination of both shape and colour as tracker representation [18], [23] becomes a natural choice. Although a global shape region provides a more reliable measure of the object identity than sparse boundary edges, in the presence of heavy occlusion clutter, a strong dynamic model of hand shape variation is still needed to infer what is happening behind the scene. We believe that such shape dynamics learned from regions are more reliable than their edge-based counterparts. Colour is another useful cue, because it can not only provide task-specific object representation (for example, skin colour can segment the hand from the shadows and form a silhouette sequence), but also provide a good measure of the moving region when we need to approximate 'rigid region' motion. In this paper, we exploit the fact that colour-histogram of the region of interest is invariant to geometrical distortions provided that it is correctly tracked. This rigid colour appearance has been studied in [6], [14]. A multiple hypothesis based particle filter together with colour representations [14] has been demonstrated as a good basis for region-based tracking.

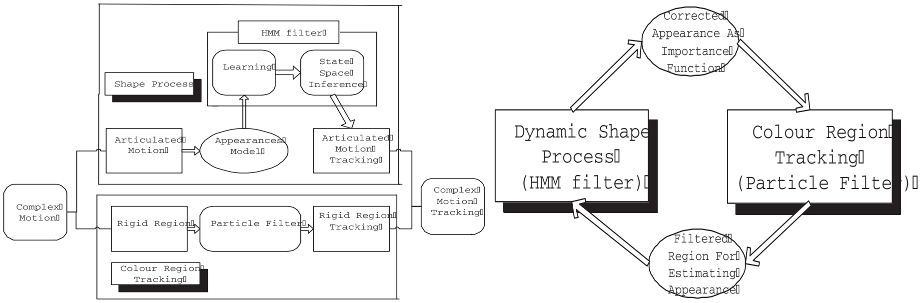

Fig. 2. (a)The flow chart of new tracking system; (b)The relationship between the two independent components. In each video frame, the Particle filter first locate the hand, and then the HMM filter infers about its' appearance. The appearance is used to update the Particle filter for the next video frame.

Although a colour and shape based tracker was proposed in [18] [23], our approach differs significantly from theirs. No dynamic models of hand shape variations are learned in [18] [23], yet this dynamics together with relatively robust features are crucial for tracking non-rigid motion under severe occlusions. A compact global silhouette representation as image moments [19] can avoid the tedious need to find the hand shape from multiple observations at local regions, and thus save computational resources for the colour-histogram computation and non-rigid shape tracking. In our Joint Bayes Filter (JBF) method, a colour-based particle filter provides a robust estimation of non-rigid object translation and localizes the most likely hand location for the HMM filter. In turn, the shape output from the HMM filter provides the importance weighting for the particle set before the resampling stage and the particle set updating in the prediction stage. This combination distinguishes our method from others. For illustrative purposes, we introduce the overall tracking system in Figure 2. The relationship between the two independent Bayesian filters, the HMM filter and the Particle filter, is also summarized.

3 Discrete Shape Tracking

In this section, we discuss the first component (the discrete-state shape tracking) of our system. Like a speech signal, we assume the articulation of hand motion as a time sequential process and can provide time-scale invariance in recognition and tracking. In most situations non-rigid motion periodically causes appearance changes. The underlying motion patterns of the articulations are often intractable, while the appearance changes often observe statistical constraints. Though a silhouette of the hand is one of the weak cues and barely preserve 3D structure of the hand, it could provide a reasonable amount of information about the articulated shape changes without resorting to a complex hand model. Image moments ($X = \{m_0, \ldots, m_n\}$ where m_i is the i^{th} moment of the shape)

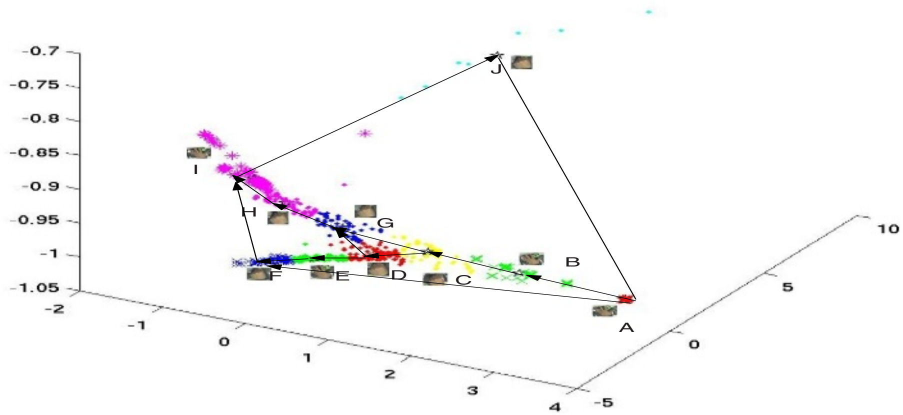

Fig. 3. Articulated hand motion embedded into a 3D metric space using Local Linear Embedding (LLE) [9] algorithm.

of the silhouettes are computed to estimate the shape class. Although a single silhouette and its image moments cannot reveal the truth about the underlying motion patterns, when we accumulate this weak evidence given sufficient amount of training data, a useful dynamic appearance manifold embedded in the training data can be discovered (Figure 3).

In line with recent advances in manifold learning [9], we embed our image moments sequence in a metric space. Figure 3 shows the distributions of articulated hand appearances. In the vector-quantized 3d metric space, not only are the spatial-temporal continuity of the non-rigid hand motion (i.e. the hand shape changes) well preserved, but also smooth trajectories (which approximate the linear Markov chains, confirming the Markov assumption underlying HMM) in the feature space are identifiable. This preliminary examination validates in a certain degree that the image features (silhouette moments in this case) is sufficient to account for non-rigid motion.

3.1 Learning and Tracking

Having assumed that non-rigid motion causes a dynamic appearance manifold, and verified that a sequence of image moments can actually replicate such dynamics (Figure 3), we concentrate on the essential learning and inference aspects for our tracking objective. Similar as the statistical learning in HMMs, dynamic motion data are aligned to discrete states and motion dynamics are approximated from the states. During tracking, in order to estimate what is going to be next most likely appearance correctly, the best underlying state sequence has to be decoded from current observation and a prior dynamic appearance model.

A classical VQ algorithm [13] is used in the learning stage to group the hand appearance data into clusters. Although complex gaussian mixtures are usually used in approximating motion data, in this paper simple L_2 distance measure is

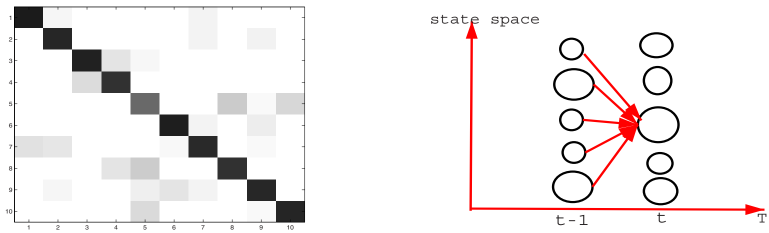

Fig. 4. (a)Dynamic model provides strong motion prior; (b)The size of the circles represent the distribution of shape indentities from the observation process, together with the dynamic model (weighted along the trajectories), determine the MAP estimation results in each frame.

used without strong parametric assumption. Thus we not only obtain the tracker representations, but also 'embed' articulated human motion into discrete states. Having obtained the codebook, the essential aspect of the HMM tracker: *spatial-temporal filtering using shape dynamics*, is straightforward. HMM provides such strong motion dynamics $P(X_t|X_{t-1})$ (Figure 4),

$$P(X_t|X_{t-1}) = \frac{\sum_{t=1}^{T-1} \xi_{ij}(t)}{\sum_{t=1}^{T-1} \gamma_i(t)} \tag{1}$$

where $\xi_{ij}(t)$ denotes the probability of being in state i at time t and at state j at time $t+1$ given the model and the observation, $\gamma_i(t)$ be defined as the probability of being in state i at time t, given the entire observation sequence and the model. This can be related to $\xi_{ij}(t)$ by summing as $\gamma_i(t) = \sum_{j=1}^{N} \xi_{ij}(t)$, N is the number of latent states.

We argue that this predictive model $P(X_t|X_{t-1})$ is typically stronger than those of *Kalman filter* or *Particle filter* tracker (Figure 7). In those trackers, the dynamic motion is usually built through a rather *ad-hoc* prediction (usually an AR model) of local feature measurements. In *HMM*, the dynamical model $P(X_t|X_{t-1})$ is built through the statistics of latent states evolution. After the alignment of the human appearances into discrete samples, latent states have been associated with meaningful example appearances (either deterministic or probabilistic). Estimating the expected number of transitions from state s_i to state s_j and the expected number of transitions from state s_i in the training set determines the likelihood that one appearance evolves to another appearance sequentially, thus approximate the prior probability distribution for non-rigid motion.

On the other hand, object tracking can be viewed as feature matching or object registration in the time domain, it shares similarities with object recognition. In discrete appearance tracking, we are interested in more than simply recognizing static patterns at a particular time t, dynamic appearance patterns are more important for tracking. We define X_t as the discrete shape tracker state (the associated exemplars) at time t, and Z_t represents image observations (image moments of the silhouettes in this case) at time t. We emphasize that the trajectories between different states A,B, C etc observe Markov assumptions. The visual tracking problem can be interpretated as making inference about the 'random walk' on the appearance manifold, however such 'randomness' must observe Markov independence property. For example, a first order Markov model observe $P(X_t|X_{1:t-1}, Z_{1:t-1}) = P(X_t|X_{t-1})$, its geometrical interpretation is, for instance, when object appearances evolve from $A \rightarrow B \rightarrow C \rightarrow G \rightarrow H \rightarrow I$ on the manifold. At state I, the example appearance is more closely related to state H than the previous appearances. This assumption is plausible because state H is closest to state I, agreeing with the general energy (distance) minimal principle as used in *Kass Snakes* tracker [21].

The *Viterbi* algorithm [20] is adapted for decoding the best underlying state sequence as well as tracking non-rigid hand motion. In order to find the single best state sequence, $Q = (q_1, q_2, ..., q_t)$ (also known as the motion trajectory such as $A \rightleftharpoons B \rightleftharpoons C \rightleftharpoons G \rightleftharpoons H$), for the given observation $O = (o_1, o_2, ..., o_t)$ (the measurements such as $\hat{A}, \hat{B}, \hat{C}$... etc), we first define the quantity

$$\delta_t(i) = \max_{q_1, q_2, \cdots, q_{t-1}} P[q_1 q_2 ... q_{t-1}, q_t = i, o_1 o_2 ... o_t|\lambda] \tag{2}$$

where λ refers to a particular HMMs for hand motion analysis. $\delta_t(i)$ is the best score (highest probability) along a single path, at time t, which accounts for the first t observations and ends in state i. Since during visual tracking, making the right predictions at each time instants is the major objective, we come to the *Bayesian* tracking formula.

$$P(X_t|Z_{1:t}) = \max_{1 \leq i \leq N} [\max_{1 \leq i \leq N} [\delta_{t-1}(X) \cdot P(X|X_{t-1})] \cdot P(Z_t|X_t)] \tag{3}$$

4 Colour-Region Tracking

Now we briefly introduce the second component (colour-region tracking) of our system. Tracking non-rigid hand motion cannot be successful without a robust global motion estimator, and this is an equally important issue when occlusion occurs. A particle filter is now the standard tool to handle such multimodal nature distractions. Colour-histogram is a relative robust and efficient region descriptor invariant to non-rigid object translation, rotation and occlusion. In this paper, only when the hand region is correctly localized, can colour segmentation provide an accurate silhouette input to the HMM filter.

Traditional colour-based particle filter tracker has some drawbacks. First it lacks a sophisticated mechanism for updating the region's scale changes. This difficulty can cause troubles for deterministic methods [6]. In [6] and [14], no clear solutions for updating the scales are given. A recent work attacks this problem by utilizing scale-space concepts [7]. In fact, the adaptive scale corresponds to the non-rigid shape changes. In our JBF framework, we explicitly model the dynamics of the particle set as a first-order AR process, updated by output from the HMM filter. A second problem with the traditional particle filter is that *factored sampling* often generate many lower-weighted samples which have little contribution to the posterior density estimation. Accuracy and efficiency are sacrificed accordingly. However, the HMM filter in the JBF tracker provides an additional sensor which can reweigh the particles and form an 'important' region for the particle filter.

5 Joint Bayes Filter

The essential aspect of Joint Bayes Filter is that at every video frame, each of the process (either shape changes or global hand motion) maintains an independent motion prior for the shape/position of the hand. The likelihood of each process is evaluated separately (nearest neighbour classification in the shape feature space, histogramming similarity measure in the colour space), but the posterior of each process is "resampled" to maintain accuracy and efficiency (the HMM filter provides importance sampling for the Particle filter, Particle filter determines the most likely hand region [1]), therefore, complementary Bayesian beliefs are reinforced and propagated through the Markov chain.

For the sake of clarity, we summarize the JBF algorithm in detail. In the HMM filter, let X_t represent the shape tracker state (associated with exemplars), and Z_t denote the image observations (image moments of the silhouette in this case) at time t. $d(X_t, Z_t)$ refers to the distance measure in feature space. The state vector of the Particle filter is defined as $x_t = (x, y, s_x, s_y)$, where x, y, s_x, s_y refer to the rectangle location $L(x, y)$ in the image plane and scales along x,y coordinates. $R(x_t)$ is the candidate region thus defined, M is the number of particles used. $b_t(u) \in \{1, \ldots, N\}$ is the bin index associated with the colour vector $y_t(u)$ at pixel location u in frame t. Assume we have a reference colour histogram: $q^\star = q^\star(n)_{n=1,\ldots,N}$ obtained at initial frame. $q_t(x_t)$ denotes the current observation of the colour histogram.

$$q_t(x_t) = C \sum_{L \in R(x_t)} \omega(|u - L|)\delta[b_t(L) - n]. \tag{4}$$

where C is a normalization constant ensuring $\sum_{n=1}^{N} q_t(X_t) = 1$, ω is a weighting function. $D([q^\star, q_t(x_t)])$ represents the Bhattacharyya distance.

The $g_t(X_t)$ used is similar to the one proposed in *ICondensation* [18], $g_t(X_t) \sim exp(-\lambda(C(S_t) + \Delta x_t))$ where $C(S_t)$ denotes the centroid of the shape, and Δx_t is the offset between the centroid of the shape and the colour region. In the following, $A_H(x_t)$ denotes the most likely hand region, which is a rectangle area. $A_S(X_t)$ refers to the shape tracker output from the HMM filter.

[1] Although it seems plausible to directly update the posterior of the HMM filter from the posterior of the Particle filter (i.e. estimate the mean shape in the feature space from the sample shape output from the whole distribution of colour regions), we believe such treatment has no significant improvement and would like to investigate this topic further in the future.

Joint Bayes Filter Algorithm

1. **Initialization.**
   Particle Filter: Select the hand region, obtain the reference colour-histogram $q^\star$. For $i = 1, 2, \ldots, M$, select the initial particle set $x_0^{(i)}$.
   HMM Filter: Obtain $A_H(x_0)$ from the tracker initialization. Perform colour segmentation in $A_H(x_0)$ to obtain the silhouette.
2. **Prediction.**
   Particle Filter: For $i = 1, 2, \ldots, M$, draw new sample set $\tilde{x_t}^{(i)} \sim p(x_t|x_{t-1}^{(i)})$, here the dynamics process is a first order AR model. Calculate the colour-histogram distribution $q_t(\tilde{x}_t)$. Evaluate the importance weights $\tilde{\omega_t}^{(i)} = \frac{p(x_t|x_{t-1}^{(i)})}{g_t(X_t)} p(z_t|\tilde{x_t}^{(i)})$, where $p(z_t|\tilde{x_t}^{(i)}) \sim \exp(-\lambda D^2[q^\star, q_t(x_t)])$, and normalize the importance weights.
   HMM Filter: Generate the new prior $P(X_t|Z_{1:t-1})$ by propagating $P(X_{t-1}|Z_{t-1})$ through the markov chain.
3. **Update.**
   Particle Filter: Resample with replacement N particles $(x_t^{(i)}; i = 1, 2, \ldots, N)$ from the set $(\tilde{x}_t^{(i)}; i = 1, \ldots, N)$ according to the importance weights. Output the $A_H(x_t)$ from the particle filter.
   HMM filter: Obtain the $A_H(x_t)$ from the Particle filter, perform colour segmentation, get the observation density $P(Z_t|X_t) \sim \exp(-\lambda d(X_t, Z_t))$. Combine with the prior $P(X_{t-1}|Z_{t-1})$ to estimate $P(X_t|Z_{1:t})$ which is the most likely appearance at time t.

6 Experiment

Several experiments are designed to examine the performance of the JBF tracker.

1. **Tracking dynamic appearances using JBF**. We obtain a long video sequence of cyclic hand motion. 70% of the data is used for training the dynamic appearance model $P(X_t|X_{t-1})$ and selecting the exemplar set, the rest for tracking. 300 particles are used to approximate the distribution of the candidate regions, each is associated with the colour-histogram density in YUV colour-space (8 bins for each channel). Near real-time performance has been achieved for the overall tracking system. The result is shown in Figure 5. Small non-rigid appearance deformations and varying changing speed between successive frames are well captured. In fact, the gait of the articulated hand motion is encoded in the strong appearance dynamics which is built in the learning stage. We also notice that even using the weak cue of image moments alone, tracking non-rigid hand poses in the JBF framework can achieve rather good performance.
2. **Coping with occlusion**. In Figure 6, we demonstrate that the region tracker provides relatively robust object detection/localization, and therefore reduces the distractions to the HMM filter. Of most interest is whether the performance of the HMM filter tracker will degenerate under several frames

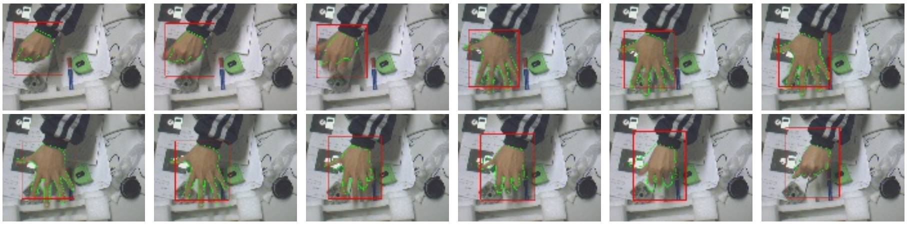

Fig. 5. Tracking results of the JBF tracker, the Particle filter determines the most likely hand region (the red rectangle), the HMM filter produce the most likely hand shapes (the green contours of the exemple shapes are drawn to show results).

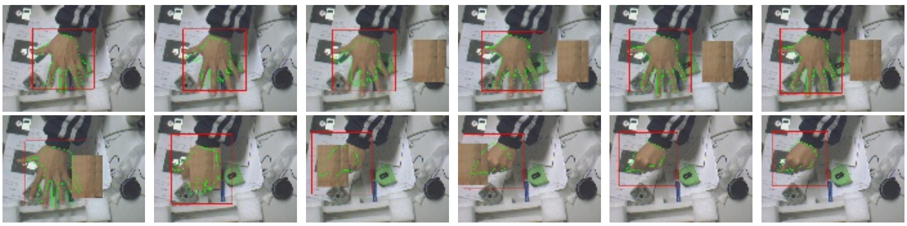

Fig. 6. The Particle filter in JBF reduces the distractions to the HMM filter.

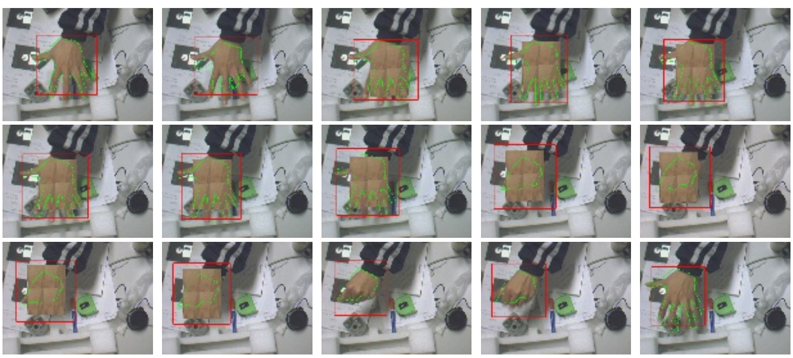

Fig. 7. The HMM filter in JBF withstands several frames of occlusion clutter.

of significant occlusion. In experiment (Figure 7), we clutter the hand regions with skin colour patch for several frames, and observe that the tracker is not only able to recover sufficient information of the hand shape, but also can correctly predict the significant change point in hand articulation. This suggests that our HMM component in JBF framework is relatively robust to significant occlusion clutter.

Here we summarize the mechanism of the HMM filter to handle occlusion clutter as demonstrated in (Figure 7) : $P(X_t|X_{t-1})$ represents the strong appearance dynamics of hand motion learned, $P(Z_t|X_t)$ represents the observation density.

In discrete appearance tracking, the most likely shape appearance estimation is given by $P(X_t|Z_t) \sim [P(X_{t-1}|Z_{t-1}) \cdot P(X_t|X_{t-1})] \cdot P(Z_t|X_t)$.

Suppose up to frame t, there are no occlusion or unreliable observations. From frame $t+1$, a significant occlusion is introduced into the video sequence. Then the observation density $P(Z_{t+1}|X_{t+1})$ contributes little to the shape appearance tracking with the first two components corresponding to the dynamic prior being most influential. A strong dynamic appearance model $P(X_t|X_{t-1})$ obtained during the learning stage, and a correct initial estimate $P(X_0|Z_0)$ in the tracking stage, are two important factors which enable the HMM filter tracker to give an optimal estimate even under harsh conditions.

7 Conclusions

This paper presents an unifying framework for non-rigid motion tracking and understanding. The contributions of our approach are as follows:

1. Explicitly separate the articulated hand motion into two independent observation processes: non-rigid motion and rigid region motion. Different dynamic models in JBF (dynamic appearance model in the HMM filter modelling the shape changes, auto-regressive process in the Particle filter updating the particle set) are complementary for articulated hand motion tracking.
2. Demonstrate the probabilistic inference mechanism of the HMM filter in visual tracking is Bayesian (MAP). In contrast to the multiple hypothesis in particle filter, we show that state-based inference is also robust to occlusion clutter and unreliable measurements. Both methods are fully Bayesian and therefore this combination (JBF filter) gives robust tracking results in real-world applications.
3. In contrast to the previous work, we associate shape descriptors with the HMM filter, colour-histogram appearance model with the Particle filter, independent target representations are closely related to the motion estimation objective(non-rigid/rigid motion), in a hand tracking and recognition application.

References

1. Aggarwal. J, Cai.Q: Human Motion Analysis: A Review. Computer Vision and Image Understanding (1999)
2. Koller. D, Weber. J and Malik. J: Robust Multiple Car Tracking with Occlusion Reasoning. Proc. European Conference on Computer Vision (1994)
3. Starner.T, Pentland.A: Visual Recognition of American Sign Language Using Hidden Markov Model. Proc. International Workshop on Automatic Face and Gesture Recognition (1995)
4. Isard. M, Blake. A: Active Contour. Springer-Verlag (1998)
5. Morris.D and Rehg.J: Singularity Analysis for Articulated Object Tracking. Proc. Computer Vision and Pattern Recognition (1998)

6. Comaniciu.D, Ramesh.V and Meer.P: Real-Time Tracking of Non-Rigid Objects using Mean Shift. Proc. Computer Vision and Pattern Recognition (2000)
7. Collins.R: Mean-shift Blob Tracking through Scale Space. Proc. Computer Vision and Pattern Recognition (2003)
8. Rehg.J and Kanade.T: Model-Based Tracking of Self-Occluding Articulated Objects. Proc.International Conference on Computer Vision (1995)
9. Roweis.S and Saul.L: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science (2000)
10. Palovic.V, Rehg.J, Cham.T, Murphy.K: A Dynamic Bayesian Network Approach to Figure Tracking Using Learned Dynamic Models. Proc. International Conference on Computer Vision (1999)
11. Stenger.B, Thayananthan.A, Torr.P, and Cipolla.R: Filtering Using a Tree-Based Estimator. Proc. International Conference on Computer Vision (2003)
12. Black.M ,Jepson.A: Eigen tracking: Robust matching and tracking of an articulatedn objects using a view based representation. Proc. European Conference on Computer Vision (1996)
13. Linde.A , Gray.R: An algorithm for vector quantization design. IEEE.Trans. on Communications (1980)
14. Pérez.P, Hue.C, Vermaak.J and Gangnet.M: Color-based probabilistic tracking. Proc. European Conference on Computer Vision (2002)
15. Rabiner.R: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proc.IEEE (1989)
16. Toyama.K and Blake.A: Probabilistic Tracking with Exemplars in a Metric Space. Proc. International Conference on Computer Vision (2001)
17. Sidenbladh.H, Black.M: Learning Image Statistics for Bayesian Tracking Proc. International Conference on Computer Vision (2001)
18. Isard.M and Blake.A: ICondensation: Unifying low-level and high-level tracking in a stochastic framework Proc. European Conference on Computer Vision (1998)
19. Brand.M: Shadow Puppetry. Proc.International Conference on Computer Vision (1999)
20. Viterbi.A: Error bounds for convolutional codes and an asymptically optimum decoding algorithm. IEEE. Transanction on Information Theory (1967)
21. Kass.M, Witkin.A, Terzopoulos.D: Snakes: Active contour models. International Journal of Computer Vision (1987)
22. Bregler.C:1997: Learning and Recognizing Human Dynamics in Video Sequences. Proc.Computer Vision and Pattern Recognition (1997)
23. Birchfield.S: Elliptical Head Tracking using Intensity Gradients and Colour Histogram. Proc. Computer Vision and Pattern Recognition (1998)
24. Gavrila.D, Philomin.V: Real-time Object Detection for ‘smart’ Vehicles International Conference on Computer Vision (1999)