Probabilistic Tracking and Recognition of Non-Rigid Hand Motion

Huang Fei and Ian Reid
Department of Engineering Science
University of Oxford, Parks Road, OX1 3PJ, UK
[fei,ian]@robots.ox.ac.uk

Abstract

Successful tracking of articulated hand motion is the first step in many computer vision applications such as gesture recognition. However the nonrigidity of the hand, complex background scenes, and occlusion make tracking a challenging task. We divide and conquer tracking by decomposing complex motion into non-rigid motion and rigid motion. A learning-based algorithm for analyzing non-rigid motion is presented. In this method, appearance-based models are learned from image data, and underlying motion patterns are explored using a generative model. Non-linear dynamics of the articulation such as fast appearance deformation can thus be analyzed without resorting to a complex kinematic model. We approximate the rigid motion as planar motion which can be approached by a filtering method. We unify our treatments of nonrigid motion and rigid motion into a single, robust Bayesian framework and demonstrate the efficacy of this method by performing successful tracking in the presence of significant occlusion clutter.

1. Introduction

Visual tracking and motion analysis play an important role in many areas such as surveillance, sports, and human-computer interfacing. 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 problems of tracking and recognition of articulated hand motion in complex scenes, and scenes with significant occlusions in a single view. In these situations, simultaneous estimation and recognition of articulated hand poses can be a challenging problem but is crucial for real-world applications of gesture recognition. With this aim in mind, we explicitly decompose the articulated hand motion into rigid motion and non-rigid dynamical motion. Rigid motion is approximated as motion of a planar region and approached using a *Particle filter* while non-rigid dynamical motion is analyzed by a *Hidden Markov Model (HMM) filter*. Although all existing methods have some difficulties in tracking non-rigid motion, our unified method demonstrates its strength by successfully recovering continuously evolving hand poses in complex scenes. Due to its ability to link the observations and underlying motion patterns in a generative fashion, hand articulation is correctly estimated even under significant occlusions.

A considerable body of work exists in hand tracking and gesture recognition. Depending on the complexity of the task, tracking can be done at different levels of accuracy. A common approach is to assume that the gesture information is encoded solely by trajectory which can be obtained via simple blob tracking. In contrast, we aim to preserve shape information as well. All tracking methods have some associated representations, either kinematic model-based [10], [8], [5] or appearance model-based [12], [15]. Kinematic model-based methods construct a geometrical model before tracking. Although it can provide more information about hand configurations than 2D appearance, during tracking it is usually associated with a tedious model fitting process, and could fail to maintain tracking where there is fast appearance deformation; in gesture recognition this is important for semantic 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 clutters pose serious distractions to these representations. Therefore, in this paper, we adopt an exemplar representation similar to that proposed by Toyama and Blake [17]. Exemplars have advantages over other representations because a useful object model and noise can be learned directly from raw data, and the non-linear nature of articulations can be conveniently represented by a sequence of exemplars which exhibits first-order Markov dependence.

Although the Hidden Markov Model [16] was proposed for sign-language analysis [3] and gesture recognition nearly a decade ago, it is not until recent years that some intrinsic aspects of the Hidden Markov Model for tracking non-rigid motion have been fully discussed [17]. We draw inspiration from Toyama and Blake [17], but our approach differs from theirs in the following aspects:

1. In [17], shape exemplars are built from edge-maps. Edge cues are sparse features [18], so tracking normally requires dense sampling in a particle filter

Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03)
0-7695-2010-3/03 $ 17.00 © 2003 IEEE

method. Usually the applications are limited to a relatively uncluttered environment such as in [17], since cluttered background and occlusions will introduce more errors for observations than multiple hypotheses can handle. However, as demonstrated later, exemplars can also be constructed from regions, which are robust to distractions.

2. Two independent dynamic processes (global motion and shape changes) share the same joint observation density provided by the chamfer distance in [17]. Thus simultaneously tracking both shape changes and global motion is only made possible via *monte-carlo* approximation using a large particle set, which must both represent shape and position. In contrast, we separate the observation processes, particle filter is used only to track the region locations, therefore less particles are used.

In this paper we propose to reexamine the following relevant issues of estimation of non-rigid hand motion using a learning-based approach: (1) **The articulated hand description and associated motion analysis methods.** (2) **Robustness to complex scenes, significant occlusions, and self-articulation distortions.** Rather than modelling a joint observation density as in [17], we explicitly separate the complex hand motion into two components assumed independent: cyclic shape variations and hand region motion, each has its own dynamic process and observation density. Then we unify two standard solutions, the HMM filter and the Particle filter, into a novel **Joint Bayes filter (JBF)**. The overall benefit of our approach is obvious: Tracking and recognition of 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. The organization of this paper is as follows: in section 2, we will introduce the problem, and a brief outline of our method will be analyzed in a research context. In section 3, the *Learning-Based Tracking* component is examined in details. In the next section, the rigid motion component is analyzed by a particle filtering approach. Afterwards some experimental results will be given. Finally, we present a summary of the work, and an outline of future research directions.

2. Joint Bayes Filter Method

Figure 1 shows us two example frames from a video sequence used for experiments. Changing appearances between successive frames can be significant, and thus rule out the standard assumption of constant intensity underlying optical flow [12]. Although an edge-based tracker [19] may perform well in normal situations, the strong occlusion clutter introduced later in the paper will damage a tracker

Figure 1: Two successive frames from a video sequence of hand motion.

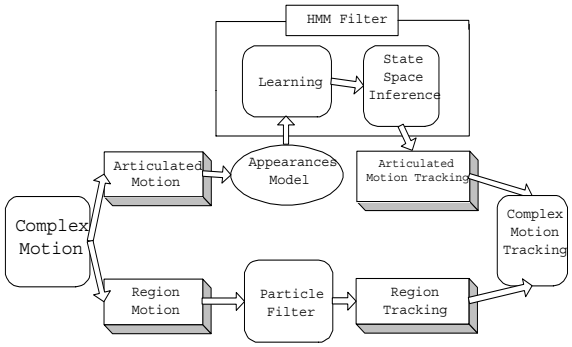


Figure 2: The tracking system diagram.

without prior dynamics [12], or with only weak prior dynamics [19]. 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 better hand representations and proper dynamical model based on the representations.

To deal with *fast appearance deformation, strong occlusion clutter and complex scene*, we propose the use of shape and colour for hand representations. Although a global shape region provides a more reliable measure of the object than sparse boundary edges in tracking, in the presence of heavy occlusion clutter, we still need a strong dynamic model of hand shape variation 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 form a good basis for region-based tracking.

Although a colour and shape based tracker was proposed

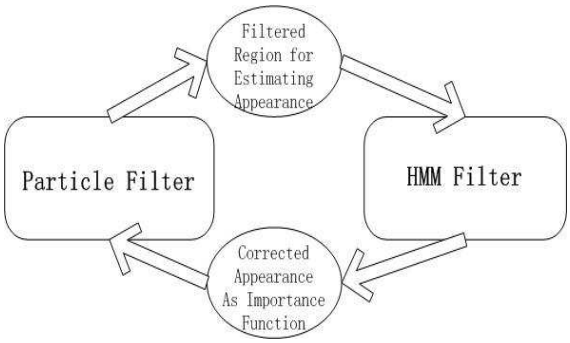


Figure 3: In each video frame, the Particle filter first locate the hand, and then the HMM filter infers about its' appearance. The appearance is used to update the Particle filter for the next video frame.

in [19], our approach differs significantly from theirs. No dynamic models of hand shape variations are learned in [19], 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 [20] ($X = (m_0, \ldots, m_n)$ where m_i is the i^{th} moment of the shape) 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 summarized in Figure 3.

3. Learning-Based Tracking

In this section, we briefly analyze the process of Learning-Based Tracking (i.e. the HMM filter) in our Joint Bayes Filter. Like a speech signal, we assume the articulation of hand motion is 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 intractable, but the appearance changes often observe statistical constraints. We can sometimes replace the problem of tracking non-rigid motion with tracking appearance changes. Though a silhouette of the hand is one of the weak cues and cannot preserve 3D structure of the hand, it could provide a reasonable amount of information about the articulated shape changes in the image plane without resorting to a complex hand model. We compute image moments of the silhouettes as in [20]. 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. In line with recent advances in manifold learning [9], we embedded our image moments sequence in a metric space for illustrative purposes. Figure 4 shows the distributions of articulated hand appearances in the manifold. Here, tracing any obvious trajectory will complete a cycle of articulated hand motion. The sparseness of the clouds not only presents the evidence of possible hidden states lying under the motion sequence, but also encodes the belief of possible state transitions in the articulated motion. Certain intermediate stages cannot be bypassed while the articulations in real hand motion cause appearance deformations. This property presents an important clue for our tracker to cope with significant occlusion as discussed in Section 5.

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 4), we can concentrate on the essential learning and inference components for our tracking objective. Similar to the statistical learning in HMMs, we have to acquire example appearances as tracker representations and construct the dynamical appearance model from the examples. 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.

Besides the need for handling small non-rigid motion, it is neither convenient nor necessary for us to keep all the information of the object during tracking. Therefore a classical VQ algorithm [13] is used in the learning stage. A sample set of typical appearances is thus obtained (Figure 5).

1. *Initialization:* Let the length of codebook be N; Randomly genererate initial codebook: $Y_N = (Y_1^0, Y_2^0, \ldots, Y_N^0)$; Set the stopping criterion ϵ; Obtain training sequences: $T_S = (X_n; n = 1, 2, \ldots, M)$;

2. *Division:* Use Y_N^n as the centroid of the classes. Divide the training sequences into N classes using L_2 distance measure, $S_j^n = (X | d(X, Y_j^n) < d(X, Y_i^n))$, $i \neq j$, $Y_i, Y_j \in Y_N^{(n)}$,$X \in T_S$; Calculate the average loss and

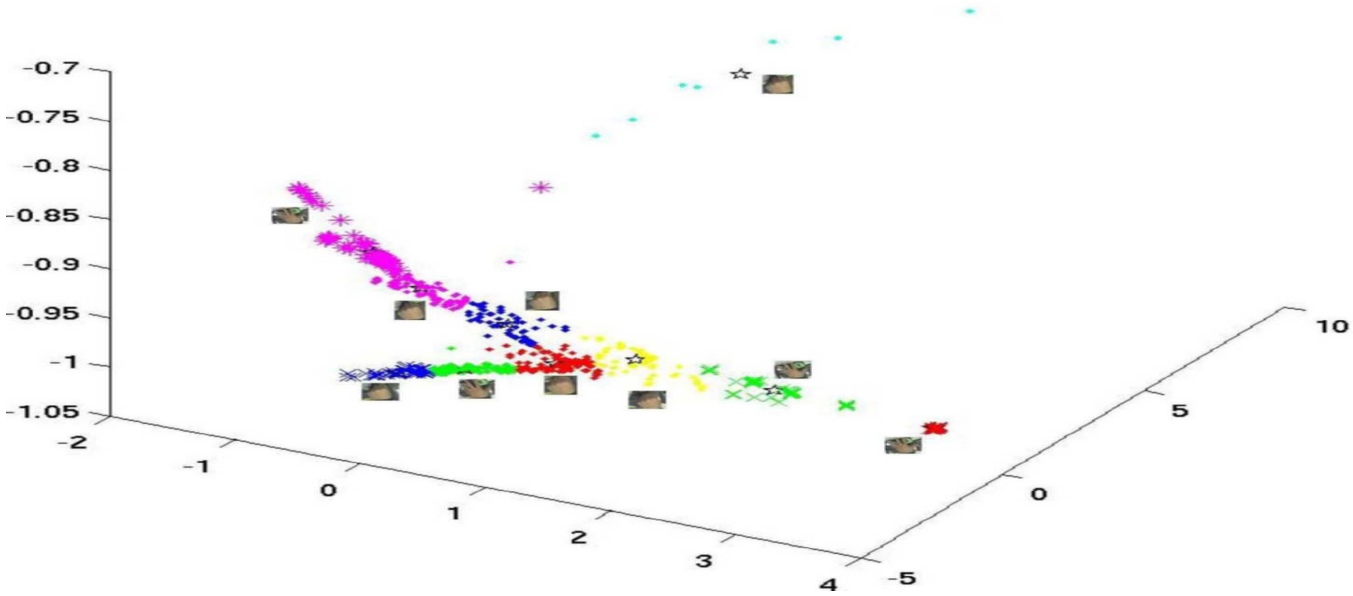


Figure 4: Manifold of the hand articulations embedded in a metric space.

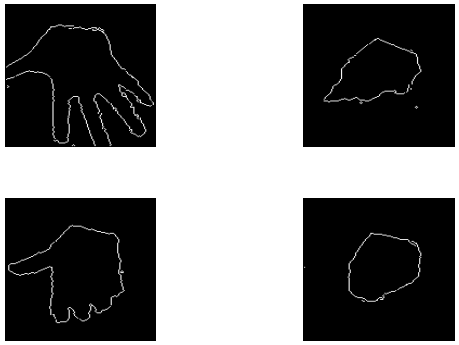


Figure 5: Sample examplars of hand apperances.

relative loss: $D^n = \frac{1}{M}\sum_{r=1}^{M} \min d(X_n, Y_n)$; $\hat{D}^n = |\frac{D^{n-1}-D^n}{D^n}|$ if $\hat{D}^n \leq \epsilon$ stop computation, else continue.

3. *Generating new examplar:* Y_i=$\frac{1}{|S_i|}\sum_{X \in S_i} X$, from the N new centroids, construct a new codebook, goto (2) until $\hat{D}^{n+1} \leq \epsilon$.

The *non-parametric* learning algorithm used is very similar to the K-NN algorithm in the first step of Local Linear Embedding (LLE) [9] which we use to visualize the underlying motion patterns in Figure 4. Given temporally aligned visual data, optimal estimation of the codebook size N can be easily achieved. Having obtained the codebook, it is straightforward to learn the temporal transitions (i.e.the shape dynamics) using standard HMM techniques (in particular the *Baum-Welch* algorithm [16]), here first-order Markov denpendence is assumed..

Discrete appearance tracking, in contrast to the problem of learning in our learning-based tracking component, has as its ultimate goal the estimation of the most likely exemplar stored in the codebook given past appearance history and current observations. As in speech recognition and gesture recognition, we make the assumption that during articulated appearance tracking, making the best decision at each video frame has a global optimal tracking result. In line with the latest developments in probabilistic tracking [4], we demonstrate that learning-based tracking in JBF can be fully incorporated into a Bayesian framework. Here, X_t represents the 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. Furthermore we assume that Z_t is conditionally independent of Z_{t-1} given X_t. $d(X_t, Z_t)$ refers to the distance measure in the feature space. In the algorithm described below, $A_S(X_t)$ denotes the shape tracker output from the HMM filter. $A_H(x_t)$ is the most likely hand region estimated by the Particle filter. The *Learning-Based Tracking* algorithm is summarized below:

1. *Belief Propagation:* Generate a new prior $P(X_t|Z_t)$ by propagating $P(X_{t-1}|Z_{t-1})$ through HMM or general case Markov Chain.

2. *Hypothesis Generation:* Acquire the most likely region $A_H(x_t)$ from the region tracker discussed in Section 4, within the region, perform colour segmentation and measure the moments of the silhouette, get the

(a) Frame 1

(b) Frame 30

(c) Frame 50

(d) Frame 70

Figure 6: The masks represent the region being tracked, left-hand for deterministic colour tracking results, right-hand for probabilistic colour tracking results.

likelihood function $P(Z_t|X_t)$ where

$$P(Z_t|X_t) \sim \exp(-\lambda d(X_t, Z_t)). \quad (1)$$

3. *Decision Making:* Compute *maximum a posteriori* (MAP) Probability $P(X_t|Z_t)$ via Bayes' rule. Output the most likely shape $A_S(X_t)$.

4. Colour-Region Tracking

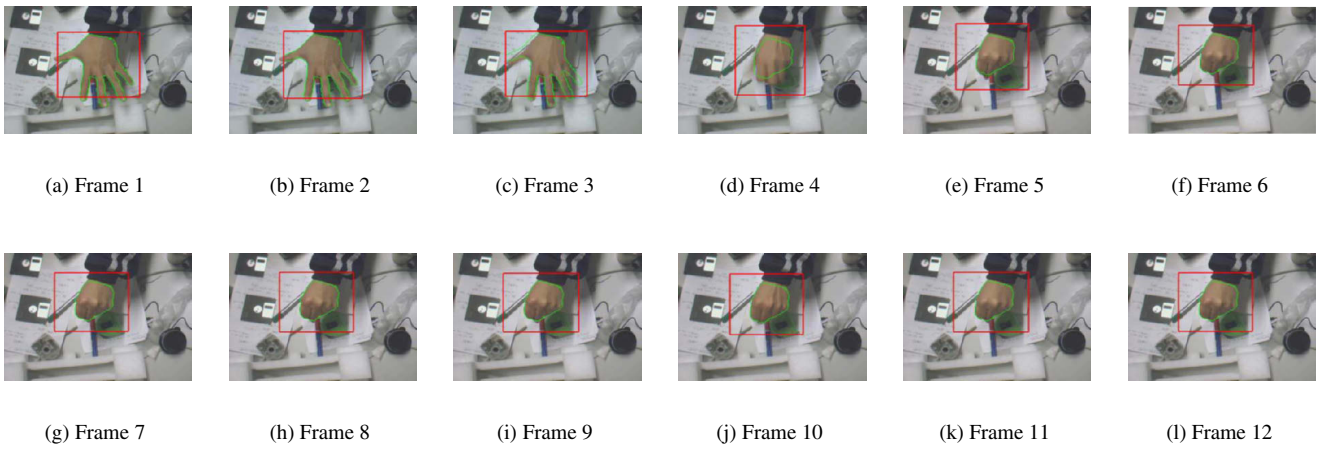
Tracking non-rigid hand motion cannot be successful without a robust global motion estimator, and this is an equally important issue when occlusion occurs. A particle filter is now the standard tool to handle such multimodal nature distractions. Colour-histogram is a relative robust and efficient region descriptor invariant to non-rigid object translation, rotation and occlusion. Colour-histogram based particle filter [14] provides a robust region estimation of the articulated hand. In our JBF framework, only when the hand region is correctly localized, can colour segmentation provide an accurate silhouette input to the HMM filter.

Traditional colour particle filter tracker has some drawbacks. First it lacks a sophisticated mechanism for updating the region's scale changes. This difficulty rules out the possibility of using a correlation based method [11] and cause troubles for deterministic methods [6]. In [6] and [14], no clear solutions for updating the scale 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 $A_S(X_t)$ 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. Acurracy and efficiency are sacrificed. However, the HMM filter in the JBF tracker provides an additional sensor which can reweight the particles and form an 'important' region for the particle filter.

In short, in the JBF framework, *monte-carlo* approximation provides an optimal localization of the hand region regardless of its articulation, background clutter, and even significant occlusion. This advantage reduces the distraction of occlusion and background clutter 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 reweighting of the particles, thus improves the accuracy and efficiency of the region tracking. In each video frame, complementary optimal Bayes beliefs are thus fused and propagated through the Markov chain.

We summarize the Particle filter part of Joint Bayes Filter algorithm as follows: 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. M is the number of particles used. $b_t(u) \in \{1, \ldots, N\}$ means 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 the initial frame. $q_t(x_t)$ denotes the current observation of the colour histogram of the region, $q_t(x_t) = C \sum \omega(|u - L|)\delta[b_t(L) - n]$, 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 coefficients [6].

Similar to the weighting scheme in [19], the importance function $g_t(X_t)$ is obtained from the HMM filter. $g_t(X_t) \sim \exp(-\lambda(C(S_t) + \Delta x_t))$ where $C(S_t)$ denotes the centroid



(a) Frame 1 (b) Frame 2 (c) Frame 3 (d) Frame 4 (e) Frame 5 (f) Frame 6

(g) Frame 7 (h) Frame 8 (i) Frame 9 (j) Frame 10 (k) Frame 11 (l) Frame 12

Figure 7: 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 appearances (the green contours).

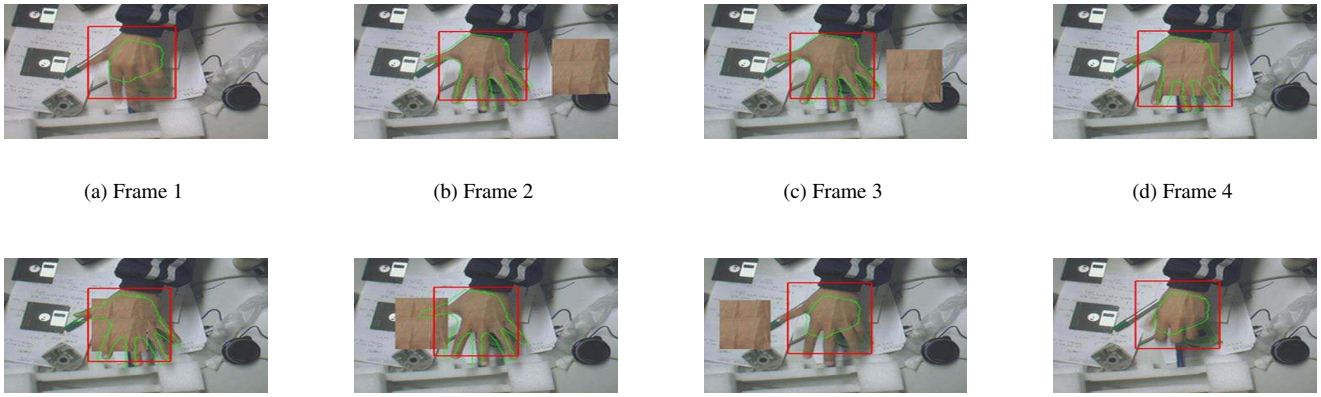
of the shape. Δx_t is the offset between the centroid of the shape and the colour-region, thus the particles which are faraway from the target have lower weights.

1. *Initialisation:* Select the hand region, obtain the reference colour-histogram $q^\star$. For $i = 1, 2, \ldots, M$, randomly generate the initial particle set $x_0^{(i)}$.
2. *Importance sampling step:* (a) For $i = 1, 2, \ldots, M$, draw new sample set from $\tilde{x_t}^{(i)} \sim p(x_t|x_{t-1}^{(i)})$, here the dynamic process is a first order AR model. (b) For $i = 1, 2, \ldots, M$, calculate the colour-histogram distribution $q_t(\tilde{x_t}^{(i)})$. 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 observation density $p(z_t|\tilde{x_t}^{(i)}) \sim \exp(-\lambda D^2[q^\star, q_t(\tilde{x}_t^{(i)})])$. (c) Normalize the importance weights.
3. *Selection step:* Resample with replacement M particles $(x_t^{(i)}; i = 1, 2, \ldots, M)$ from the set $(\tilde{x}_t^{(i)}; i = 1, \ldots, M)$ according to the importance weights. The mean of 10% of the best particles is estimated as $A_H(x_t)$.

5. Experiments and Results

We design several experiments to examine the performance of the JBF tracker.

1. **Colour-region tracking**. The aim of this experiment is to evaluate the particle filter's improvement on colour region tracking. We implement the algorithm given in [14] and perform a test on a face sequence, in which we initialize the tracker on the face region selected at the initial frame. Deterministic colour-histogram distance minimization is easily distracted by the background clutter while multiple-hypothesis particle filter helps stabilize the tracker. See Figure 6 for illustration.
2. **Tracking dynamic appearances using JBF**. We obtain a long video sequence of cyclic hand motion. 60% 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. 200 particles are used to approximate 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 7. Small non-rigid appearance deformations and varying changing speed between successive frames are well captured by the HMM filter. 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.
3. **Coping with occlusions**. In Figure 8, we demonstrate that the region tracker reduces the distractions to the HMM filter. Of most interest is whether the performance of the HMM filter tracker will degenerate under several frames of significant occlusions. Experiment (Figure 9) shows that skin colour occlusions do not prevent the tracker from recovering the articulated hand poses. This suggests that our learning-based tracking component in JBF framework is robust to significant occlusions and unreliable observations.



(a) Frame 1 (b) Frame 2 (c) Frame 3 (d) Frame 4

(e) Frame 5 (f) Frame 6 (g) Frame 7 (h) Frame 8

Figure 8: The Particle filter in JBF reduces the distractions to the HMM filter.

Here we summarize the mechanism of the HMM filter to handle occlusion clutter as demonstrated in (Figure 9) : $P(X_t|X_{t-1})$ represents the strong appearance dynamics of hand motion learned, $P(Z_t|X_t)$ represents the observation density (see Equation 1. 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 occlusions 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.

6. Conclusions and Future Work

This paper is an extension to recent efforts to combine the HMM filter and the Particle filter for the purpose of visual tracking. We place emphasis on non-rigid hand motion analysis in the presence of scene clutter and occlusion, situations which are not covered by previous work. The successful tracking and recognition results of this work have potential applications in gesture recognition and augmented reality. The experimental results presented in the paper can be found at http://www.robots.ox.ac.uk/˜fei.

We make the following contributions in our paper:

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.

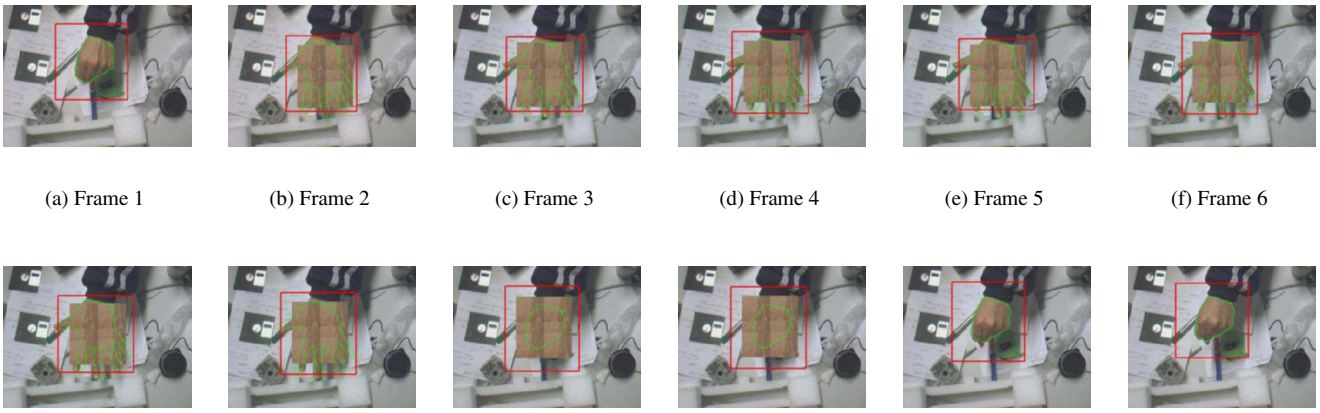
2. Demonstrate the probabilistic inference mechanism of the HMM filter in visual tracking. In contrast to the multiple hypothesis in particle filter, we show that state-based inference is also robust to occlusion clutter and unreliable measurements. Both methods are fully Bayesian and therefore this 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, in a hand tracking and recognition application.

Due to the complexity of hand articulations, there are some issues that remain to be explored in the future. A few are summarized as follows:

1. Hand appearances invariant to the viewpoint. Simple image moments are robust enough for estimation appearances in hand articulation even under significant distractions. However the changing of viewpoint will degenerate the performance significantly. Multiple-view representation seems to be a potential research direction.

2. We propose joint filters in estimation of both motion and appearance. The combination of state-based infer-



(a) Frame 1 (b) Frame 2 (c) Frame 3 (d) Frame 4 (e) Frame 5 (f) Frame 6

(g) Frame 7 (h) Frame 8 (i) Frame 9 (j) Frame 10 (k) Frame 11 (l) Frame 12

Figure 9: The HMM filter in JBF withstands several frames of occlusion clutters.

ence method and sampling method is still a research issue. We plan to investigate further.

References

[1] J. K. Aggarwal and Q. Cai "Human Motion Analysis: A Review" *Computer Vision and Image Understanding*,1999

[2] D. Koller, J. Weber and J. Malik "Robust Multiple Car Tracking with Occlusion Reasoning" *Proc.Third European Conference on Computer Vision*,1994

[3] T.Starner and A.Pentland "Visual Recognition of American Sign Language Using Hidden Markov Models" *Proc.International Workshop on Automatic Face and Gesture Recognition*,1995

[4] M.Isard and A.Blake "Active Contour" *Springer-Verlag*,1998

[5] D.D.Morris and J.M.Rehg "Singularity Analysis for Articulated Object Tracking" *Proc.Computer Vision and Pattern Recognition*,1998

[6] D. Comaniciu, V. Ramesh and P. Meer "Real-Time Tracking of Non-Rigid Objects using Mean Shift" *Proc.Computer Vision and Pattern Recognition*,2000

[7] R.T.Collins "Mean-shift Blob Tracking through Scale Space" *Proc.Computer Vision and Pattern Recognition*, 2003

[8] J.M.Rehg and T.Kanade "Model-Based Tracking of Self-Occluding Articulated Objects" *Proc.International Conference on Computer Vision*,1995

[9] S.Roweis and L.Saul "Nonlinear dimensionality reduction by locally linear embedding" *Science*,2000

[10] B.Stenger, P.R.S.Mendonca and R.Cipolla "Model-Based 3D Tracking of an Articulated Hand" *Proc.Computer Vision and Pattern Recognition*,2001

[11] G.Hager and K.Toyama "The XVision system: A general-purpose substrate for portable real-time vision applications" *Computer Vision and Image Understanding*,1998

[12] M. J. Black and A. Jepson "Eigen tracking: Robust matching and tracking of an articulatedn objects using a view based representation" *Proc.4th European Conference on Computer Vision*,1996

[13] A.Linde and R.Gray "An algorithm for vector quantization design" *IEEE.Trans.on Communications*,1980

[14] P. Prez, C. Hue, J. Vermaak and M. Gangnet "Color-based probabilistic tracking" *Proc.European Conference on Computer Vision*,2002

[15] R.Bowden and M.Sarhadi "A non-linear Model of Shape and Motion for Tracking Finger Spelt American Sign Language" *Image and Vision Computing*,2002

[16] R.Rabiner "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition" *Proc.IEEE*,1989

[17] K.Toyama and A.Blake "Probabilistic Tracking with Exemplars in a Metric Space" *Proc. Int. Conf. on Computer Vision*, 2001

[18] H.Sidenbladh and M.Black "Learning Image Statistics for Bayesian Tracking" *Proc.Int.Conf.on Computer Vision*,2001

[19] M.Isard and A.Blake "Icondensation: Unifying low-level and high-level tracking in a stochastic framework" *Proc.5th European Conference on Computer Vision*,1998

[20] M.Brand "Shadow Puppetry" *Proc.Int.Conf.on Computer Vision*,1999