

Robust Head Tracking Based on a Multi-State Particle Filter

Yuan LI¹, Haizhou AI¹, Chang HUANG¹, Shihong LAO²

¹Computer Science and Technology Department, Tsinghua University, Beijing 100084, China

²Sensing and Control Technology Laboratory, Omron Corporation, Kyoto 619-0283, Japan

E-mail: ahz@mail.tsinghua.edu.cn

Abstract

This paper proposes a novel method for robust and automatic realtime head tracking by fusing face and head cues within a multi-state particle filter. Due to large appearance variability of human head, most existing head tracking methods use little object-specific prior knowledge, resulting in limited discriminant power. In contrast, face is a distinct pattern much easier to capture, which motivates us to incorporate a vector-boosted multi-view face detector[6] to lend strong aid to general head observation cues including color and contour edge. To simultaneously and collaboratively perform temporal inference of both the face state and the head state, a Markov-network-based particle filter is constructed using sequential belief propagation Monte Carlo[5]. Our approach is tested on sequences used by previous researchers as well as on new data sets which includes many challenging real-world cases, and shows robustness against various unfavorable conditions.

1. Introduction

Visual object tracking has received much research interest in the computer vision community since it is a ubiquitous fundamental task in many video-based applications. As one of most popular cases of visual object tracking, human head tracking easily finds its way into applications such as visual surveillance, video retrieval, human-computer interaction, auto-focus and stabilizing of imaging devices, etc., many of which require algorithms that make as few assumptions as possible about the tracking environment.

Despite the fact that head tracking is commonly used as experiment scenario to demonstrate general tracking algorithm[5][1], robust automatic head tracking through a wide variety of conditions is still an open problem. Most existing approaches mainly rely on general image cues such as color, corner points, edge, background subtraction, etc.[2][10][4][11], which are often in peril when confronted with challenging situations such as poor illumination and



Figure 1. Head involves much larger appearance variability than face.

strong background noise, due to lack of higher level object-specific knowledge.

On the other hand, tracking methods which exploits more specialized cues are mainly confined to head tracking for face-visible poses or facial feature points tracking[16][3][8]. This is because the head as a whole exhibits much larger appearance variability (especially the head rear, probably caused by hair color and style) and contains no relatively strong and stable features that the face contains, such as the layout and texture of facial organs, which allow for specialized modelling (Figure 1).

In fact, although not always visible, the face pattern as a partial feature of the head can still provide strong aid to head tracking. Consider the human vision system as an analog: to identify an object, we often looks for specific parts with strong characteristics as a hint. See Figure 2 for an example. For the first frame it may be difficult even for human to identify the head rear since it does not present much distinct feature against the dark background. However once we see the face, we easily capture the target throughout the rest frames although the face is occasionally occluded.

Based on this idea, we propose a method that works in a similar manner by fusing statistically learned face observation model and general head observation model within the well-known probabilistic tracking framework – particle filter. For the face part, a vector-boosted multi-view face detector is trained over large data sets and modelled probabilistically to output a face likelihood for each input image patch; while for the head part, we adopt both color histogram and contour edge measurements.

Compared with conventional data fusion approaches with particles, in which different observation models either



Figure 2. Face cue aids head tracking in difficult situations.

share the same state variable[14][15] or work on state variables that form a strict coarse-to-fine hierarchy[11], here the fusion difficulty lies in the coexistence of two distinct states: the head and the face. To overcome this we borrow the idea from non-temporal inference problems of graphical models, namely we extend the underlying Markov chain model of standard particle filter to a temporal Markov network, and use the potential functions associated with links between different states to represent their correlation. The inference over such model is therefore done by sequential belief propagation Monte Carlo first developed in [5]. In [5] the algorithm is used for collaborative multi-scale tracking, where the state variables corresponds to the target state in different scales. Different from their work, our aim is to exploit the spatial or model-induced structure relationship between characteristic aspects. And for head tracking, we show that such method is able to utilize and temporally accumulate information from both face and head to achieve robust tracking.

The rest of this paper is organized as follows. Section 2 presents the multi-state particle filter for fusing observation models with multiple distinct but correlated state variables. This probabilistic tracking framework is enriched in Section 3 and Section 4, which describe the observation models for face and head respectively. We give experimental results in Section 4 and the conclusion in Section 5.

2. Multi-State Particle Filter

Standard particle filter is developed based on Bayesian sequential estimation. Denote the hidden state of target and its observation at time t by \mathbf{x}_t and \mathbf{y}_t respectively, the filtering distribution $p(\mathbf{x}_t|\mathbf{Y}_t)$ stands for the distribution of target state given all observations $\mathbf{Y}_t = (\mathbf{y}_1, \dots, \mathbf{y}_t)$, which can be computed by the well-known predict-update recursion

$$p(\mathbf{x}_t|\mathbf{Y}_{t-1}) = \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{Y}_{t-1})d\mathbf{x}_{t-1}, \quad (1)$$

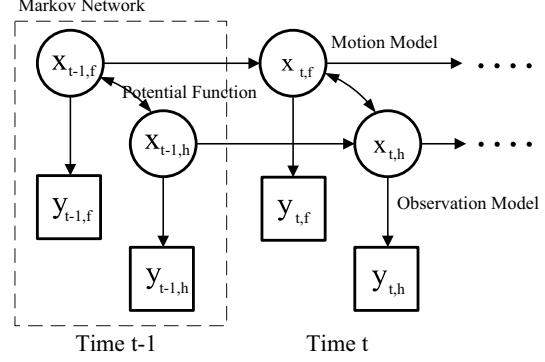


Figure 3. The temporal Markov network of our head tracking algorithm.

$$p(\mathbf{x}_t|\mathbf{Y}_t) \propto p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{Y}_{t-1}). \quad (2)$$

The recursion requires a motion model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and an observation model $p(\mathbf{y}_t|\mathbf{x}_t)$. To handle complicated distributions which lead to analytical intractability, particle filter[7] approximates the two steps by a set of weighted samples $\{\mathbf{x}_t^{(n)}, \pi_t^{(n)}\}_{n=1}^N$.

However, standard particle filter only provides a solution to single hidden state temporal inference problem. To allow for more flexible data fusion involving multiple hidden state variables, we adopt a multi-state particle filter based on sequential belief propagation.

2.1 Sequential Belief Propagation Monte Carlo

When a group of distinct yet correlated state variables $\{\mathbf{x}_i\}$ (for our head tracking method they are the head \mathbf{x}_h and the face \mathbf{x}_f) are evolving over the temporal domain, a temporal Markov network (Figure 3) is formed as a generalized case of Markov chain. Within one time frame t , for each hidden state $\mathbf{x}_{t,i}$, an observation $\mathbf{y}_{t,i}$ is made; and the link from $\mathbf{x}_{t,i}$ to $\mathbf{x}_{t,j}$ is associated with a potential function $\psi_{i,j}(\mathbf{x}_{t,i}, \mathbf{x}_{t,j})$, which in our problem represents the structural relationship between head and face.

For tracking, we are interested in the filtering distribution $p(\mathbf{x}_{t,i}|\mathbf{Y}_t)$ for each $\mathbf{x}_{t,i}$, where \mathbf{Y}_t stands for $\{\mathbf{Y}_{t,j}\}$ meaning that the inference should be made based on observations of all hidden states. We calculate them by sequential belief propagation (SBP)[5].

Similar to the belief propagation algorithm for non-temporal cases, SBP also undergoes a local message passing process. The message passed from state $\mathbf{x}_{t,i}$ to $\mathbf{x}_{t,j}$ is

$$m_{ij}(\mathbf{x}_{t,j}) = \int_{\mathbf{x}_{t,i}} [p(\mathbf{y}_{t,i}|\mathbf{x}_{t,i})\psi_{i,j}(\mathbf{x}_{t,i}, \mathbf{x}_{t,j}) \int_{\mathbf{x}_{t-1,i}} p(\mathbf{x}_{t,i}|\mathbf{x}_{t-1,i})p(\mathbf{x}_{t-1,i}|\mathbf{Y}_{t-1})d\mathbf{x}_{t-1,i} \prod_{k \in \mathcal{N}(\mathbf{x}_{t,i}) \setminus j} m_{ki}(\mathbf{x}_{t,i})] d\mathbf{x}_{t,i}, \quad (3)$$

where $\mathcal{N}(\mathbf{x}_{t,i})$ denotes all state variables with a link to $\mathbf{x}_{t,i}$. The messages are passed iteratively until convergence, and the filtering distribution is given by

$$p(\mathbf{x}_{t,i}|\mathbf{Y}_t) \propto p(\mathbf{y}_{t,i}|\mathbf{x}_{t,i}) \prod_{k \in \mathcal{N}(\mathbf{x}_{t,i})} m_{ki}(\mathbf{x}_{t,i}) \int_{\mathbf{x}_{t-1,i}} p(\mathbf{x}_{t,i}|\mathbf{x}_{t-1,i}) p(\mathbf{x}_{t-1,i}|\mathbf{Y}_{t-1}) d\mathbf{x}_{t-1,i}. \quad (4)$$

Since closed-form solutions to the two distributions either do not exist or are computationally expensive to obtain, a Monte Carlo version of sequential belief propagation (SBPMC) is developed in [5]. Same as in the standard particle filter, the filtering distribution is represented by weighted samples, i.e.,

$$p(\mathbf{x}_{t,i}|\mathbf{Y}_t) \sim \{\mathbf{x}_{t,i}^{(n)}, \pi_{t,i}^{(n)}\}_{n=1}^N. \quad (5)$$

Further, each message at time t is also approximated by

$$m_{ji}(\mathbf{x}_{t,i}) \sim \{\mathbf{x}_{t,i}^{(n)}, \omega_{t,i}^{(n)}\}_{n=1}^N. \quad (6)$$

Such approximation enables SBPMC to perform in a completely non-parametric manner similar to particle filter. Instead of elaborating on the general SBPMC algorithm, we will discuss its application in our head tracking approach.

2.2 Multi-State Particle Filter for Head Tracking

By representing the head by an ellipse and the face by a square region, the state variables are defined as

$$\mathbf{x}_h = \langle x_h, y_h, a, b \rangle; \quad (7)$$

$$\mathbf{x}_f = \langle x_f, y_f, s, \theta \rangle. \quad (8)$$

For the head state, $\langle x_h, y_h \rangle$ denotes the center coordinates of the ellipse, a and b denote the semimajor and semiminor axes. For the face state, $\langle x_f, y_f \rangle$ and s are the center coordinates and side length of the face square, and θ is the pose (with five discrete values: frontal, left and right half profile, left and right full profile). Observations of the head and the face are $\mathbf{y}_{t,h}$ and $\mathbf{y}_{t,f}$ respectively, and our goal is to obtain the distributions $p(\mathbf{x}_{t,h}|\mathbf{Y}_t)$ and $p(\mathbf{x}_{t,f}|\mathbf{Y}_t)$, where \mathbf{Y}_t stands for $\{\mathbf{Y}_{t,f}, \mathbf{Y}_{t,h}\}$.

SBPMC is applied to this specific model to derive the multi-state particle filter for head tracking, which is shown in Table 1.

The remaining problems are: what are the potential functions ($\psi_{h,f}(\mathbf{x}_h, \mathbf{x}_f)$, $\psi_{f,h}(\mathbf{x}_f, \mathbf{x}_h)$), the motion models ($p(\mathbf{x}_{t,f}|\mathbf{x}_{t-1,f})$, $p(\mathbf{x}_{t,h}|\mathbf{x}_{t-1,h})$) and the observation models ($p(\mathbf{y}_{t,f}|\mathbf{x}_{t,f})$, $p(\mathbf{y}_{t,h}|\mathbf{x}_{t,h})$)?

The potential functions are expected to reflect the spatial correlation between face and head, which varies under different poses. Figure 4 shows the center position of

Table 1. Multi-state particle filter for head tracking

With the particle set of the head state $\{\mathbf{x}_{t-1,h}^{(n)}, \pi_{t-1,h}^{(n)}\}_{n=1}^N$, the face state $\{\mathbf{x}_{t-1,f}^{(n)}, \pi_{t-1,f}^{(n)}\}_{n=1}^N$, and a boolean label v indicating whether face is visible at time $t-1$, proceed as follows at time t :

- Re-sample:
 - For the head, simulate $\alpha_n \sim \{\pi_{t-1,h}^{(n)}\}_{n=1}^N$, and replace $\{\mathbf{x}_{t-1,h}^{(n)}, \pi_{t-1,h}^{(n)}\}_{n=1}^N$ with $\{\mathbf{x}_{t-1,h}^{(\alpha_n)}, 1/N\}_{n=1}^N$;
 - For the face, if v , simulate $\alpha_n \sim \{\pi_{t-1,f}^{(n)}\}_{n=1}^N$, and replace $\{\mathbf{x}_{t-1,f}^{(n)}, \pi_{t-1,f}^{(n)}\}_{n=1}^N$ with $\{\mathbf{x}_{t-1,f}^{(\alpha_n)}, 1/N\}_{n=1}^N$; else re-initialize face samples according to the output head state at the previous time step.
- Prediction: simulate $\mathbf{x}_{t,h}^{(n)} \sim p(\mathbf{x}_{t,h}|\mathbf{x}_{t-1,h}^{(n)})$; simulate $\mathbf{x}_{t,f}^{(n)} \sim p(\mathbf{x}_{t,f}|\mathbf{x}_{t-1,f}^{(n)})$.
- Update:
 - $\pi_{t,f}^{(n)} \leftarrow p(\mathbf{y}_{t,f}|\mathbf{x}_{t,f}^{(n)})$, $\pi_{t,h}^{(n)} \leftarrow p(\mathbf{y}_{t,h}|\mathbf{x}_{t,h}^{(n)})$;
 - if $\sum_{n=1}^N \pi_{t,f}^{(n)} > \Gamma$ (Γ is a constant threshold),
 - Face visible, $v \leftarrow true$;
 - For $n = 1..N$, calculate messages between face and head states:
 - $\omega_{t,f}^{(n)} \leftarrow \sum_{m=1}^N \pi_{t,h}^{(m)} \cdot \psi_{h,f}(\mathbf{x}_{t,h}^{(m)}, \mathbf{x}_{t,f}^{(n)})$,
 - $\omega_{t,h}^{(n)} \leftarrow \sum_{m=1}^N \pi_{t,f}^{(m)} \cdot \psi_{f,h}(\mathbf{x}_{t,f}^{(m)}, \mathbf{x}_{t,h}^{(n)})$;
 - For $n = 1..N$, pass messages:
 - $\pi_{t,f}^{(n)} \leftarrow \pi_{t,f}^{(n)} \cdot \omega_{t,f}^{(n)}$, $\pi_{t,h}^{(n)} \leftarrow \pi_{t,h}^{(n)} \cdot \omega_{t,h}^{(n)}$;
 - Re-sample and iterate the update step till convergence.
 - Else, face invisible, $v \leftarrow false$.
- Normalize sample weight so that $\sum_{n=1}^N \pi_{t,h}^{(n)} = 1$ and $\sum_{n=1}^N \pi_{t,f}^{(n)} = 1$.
- Inference:
 - Estimate the head state by MMSE:
 - $\hat{\mathbf{x}}_{t,h} \leftarrow \sum_{n=1}^N \mathbf{x}_{t,h}^{(n)} \cdot \pi_{t,h}^{(n)}$;
 - If v , also estimate the face state:
 - $\hat{\mathbf{x}}_{t,f} \leftarrow \sum_{n=1}^N \mathbf{x}_{t,f}^{(n)} \cdot \pi_{t,f}^{(n)}$.

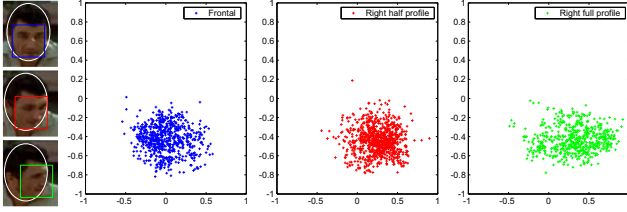


Figure 4. Position of face relative to head under different poses (left: frontal, middle: right half profile, right: right full profile).

face relative to head (head is centered at the origin and its bounding box is scaled to $[-1, 1] \times [-1, 1]$). θ of the face state enables us to model potential functions separately for each pose. Gaussian distributions are used with parameters learned from data with manual labelled head contour and detected face.

For the motion models we avoid sophisticated models since zero-order Gaussian diffusion seems to suffice for our tracking algorithm. In the next two sections we introduce the observation models of the tracker.

3 Face-based Observation Model

Concerning integrating face detection into tracking, the nature of the head tracking problem makes two requirements: 1) because of the wide-range pose change of the head, a detector must cover multi-view face to be truly effective; 2) the speed of detection should be fast enough to be part of a real-time system.

Although much research work has been done on face detection, it is until recent years that face detection algorithms which meet the above requirements are developed[18][6]. To our knowledge, works involving the integration of detection technique into tracking are numbered. [16][9] are two representative attempts related to our work. [16] gracefully integrates the detector of [13] in face tracking, but only reaches a speed of 26 seconds per frame (352x288 pixel); [9] trains a cascade of classifiers[18] for the proposal distribution of particle filter rather than an observation model.

We construct a probabilistic observation model based on a vector-boosted tree structure detector using Haar-like features[6], which covers a range of $\pm 90^\circ$ out-of-plane rotation (yaw) and $\pm 45^\circ$ in-plane rotation (tilt). The detector itself is a coarse-to-fine structure: as the tree branches, the face space is divided into smaller subspaces according to pose. Each node of the tree is a strong classifier boosted from numerous weak classifiers, and it decides whether an input image patch should be rejected or passed to one or more of its son nodes. In one detection run, an image patch starts from the root node to travel along tree paths by width-first search, and is accepted as face if and only if it passes

one or more leaf-nodes (each leaf-node represents a different pose).

While face detection only focuses on whether an input image patch is a face or not, we are now interested in how likely it is a face, i.e., $p(\mathbf{y}_f|\mathbf{x}_f)$, which is rarely discussed in existing works. Here we propose a probabilistic approximation of $p(\mathbf{y}_f|\mathbf{x}_f)$ from the output of a series of boosted strong classifiers by diving further into the training process.

For any given state \mathbf{x}_f , let I_x be the corresponding image patch. The face likelihood can be formulated as

$$\begin{aligned} p(\mathbf{y}_f|\mathbf{x}_f) &= p(\text{face}|I_x) \\ &= \frac{r \cdot p(I_x|\text{face})}{r \cdot p(I_x|\text{face}) + p(I_x|\overline{\text{face}})}, \end{aligned} \quad (9)$$

where $r = p(\text{face})/p(\overline{\text{face}})$ is the a priori ratio.

Any strong classifier (tree node) v that I_x passes projects I_x to some feature space and outputs a confidence $f_v(I_x)$. Hence we approximate $p(I_x|\text{face})$ and $p(I_x|\overline{\text{face}})$ in the projected space by $p(f_v(I_x)|\text{face})$ and $p(f_v(I_x)|\overline{\text{face}})$, which can be adequately estimated by two single gaussian distributions learned from training samples on which v is trained.

On the other hand, any strong classifier v also has a different a priori ratio r since the face and non-face subspaces have been continually refined by its ancestor classifiers. If an initial r_0 for the root node is assumed, and the face sample accepting rate α and non-face sample rejecting rate β are given during training, the a priori ratio r_v for any node v can be computed by

$$r_v = r_0 \prod_{i=0}^{l-1} \frac{\alpha_{v_i}}{1 - \beta_{v_i}}, \quad (10)$$

where $\{v_0, \dots, v_{l-1}\}$ is the path from root node v_0 to v .

Therefore,

$$p(\mathbf{y}_f|\mathbf{x}_f) \approx \frac{r_v \cdot p(f_v(I_x)|\text{face})}{r_v \cdot p(f_v(I_x)|\text{face}) + p(f_v(I_x)|\overline{\text{face}})}, \quad (11)$$

where v is the node with the maximum layer number among all nodes that I_x have passed.

4 Head-based Observation Model

To accomplish head tracking with full range out-of-plane rotation, we further introduce two relatively general image cues concerning the head as a whole: one is the color histogram[11] inside the head region, and the other is intensity edge along head contour[7]. Since these two cues and their variations are widely studied and adopted for visual tracking, we only give a brief summary.

For the color cue, we use histogram in the three channels of the RGB color space which have N bins: $\mathbf{h} =$

(h_1, h_2, \dots, h_N) . Each candidate histogram is compared to a reference histogram by the distance metric defined based on Bhattacharyya coefficient:

$$D(\mathbf{h}_x, \mathbf{h}_{\text{ref}}) = \left(1 - \sum_{i=1}^N \sqrt{h_{x,i} h_{\text{ref},i}}\right)^{1/2}. \quad (12)$$

And the color likelihood is given by

$$p(\mathbf{y}_h^{\text{color}} | \mathbf{x}_h) \propto \exp(-D^2(\mathbf{h}_x, \mathbf{h}_{\text{ref}}) / 2\sigma_{\text{color}}^2). \quad (13)$$

The contour edge likelihood is based on the assumptions that the clutter along any contour normal is Poisson process with spatial density λ and that the measurement error of any true edge point obeys a gaussian distribution $N(0, \sigma_\varepsilon)$. A certain number of contour normals are selected uniformly as measurement lines, along each such line l , edge points are searched for. The edge likelihood at l is given by

$$p(l | \mathbf{x}_h) \propto q + \frac{1-q}{\sqrt{2\pi}\sigma_\varepsilon\lambda} \sum_m \exp\left(-\frac{d_m^2}{2\sigma_\varepsilon^2}\right), \quad (14)$$

where q is the probability that the true edge point is not detected, and d_m is the distance from the m -th detected edge points to the contour. Assuming that observations of measurement lines are independent, the overall likelihood of the contour is

$$p(\mathbf{y}_h^{\text{edge}} | \mathbf{x}_h) = \prod_l p(l | \mathbf{x}_h). \quad (15)$$

Combining the color and the edge observation we obtain

$$p(\mathbf{y}_h | \mathbf{x}_h) = p(\mathbf{y}_h^{\text{color}} | \mathbf{x}_h) p(\mathbf{y}_h^{\text{edge}} | \mathbf{x}_h). \quad (16)$$

5. Experiments

In implementation, we simply use independent filters for multi-target tracking (since our focus is not multi-target interaction) by means similar to [17]. Because of the presence of detector, new target can be automatically initialized by performing full-frame detection every several frames. The tracker runs at about 15 frames per second on a laptop with a Pentium M740(1.73MHz) CPU and 512M RAM (video frame size 320×240).

We find that quantitative comparison with other methods is difficult due to lack of standard test set and performance statistics in previous works. Nevertheless, Figure 5 shows a comparison of results on a video sequence by frame-based detection, color-and-edge-based tracking and our method. All three methods have the same initialization, and the strong background intensity gradient indicates a difficult case for the edge cue (a). The detector fails from time to time since the face is blurry and poorly illuminated (b); tracker using only general cues like color and edge is also quickly distracted by clutter (c). However our method is



(a) Initialization on the first frame and the Sobel gradient map



(b) Frame-based detection by method in [6]



(c) Tracking by standard particle filter using color histogram and gradient



(d) Tracking by our method

Figure 5. Comparison of results by different methods.

able to track both the head contour and the face successfully by temporally combining and accumulating information from face and head cues.

Our tracker is tested on a wide range of data including the popular test cases first used by [2], desktop online capture, commercial movie clips and family video clips shot by hand-held video cameras. The latter two categories are more challenging because they contain various complicated scenes and irregular camera motion. Our method has shown robustness against target position, size and pose change as well as unfavorable conditions such as occlusion, poor illumination and cluttered background. See Figure 6 for a few examples.

6. Conclusion

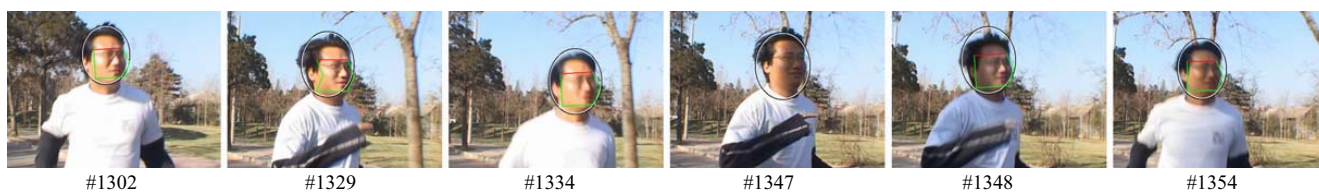
In this paper we describe a novel head tracking method by fusing face and head cues within a particle filter which supports multiple correlated hidden state variables. In addition to conventional color and contour cues, latest advances in face detection is integrated by modelling the detector into a probabilistic observation likelihood to significantly increase the tracker's discriminant power. The resulting real-time tracker, which has shown robustness under various circumstances in experiments, can automatically track not only head but also face as well as infer the face pose. The



(a) Tracking in low resolution video with occlusion (video clip from [2]).



(b) Tracking an actress and an actor with a face mask in a cluttered scene.



(c) Tracking a running man with motion blur, illumination change and camera vibration.

Figure 6. Tracking results.

performance may be further improved by adopting stronger head observations. Finally, our idea of a more free-form data fusion manner in the sense of simultaneously and collaboratively exploiting different characteristic parts of the tracking target can also be extended to other object tracking problems.

7. Acknowledgements

This work is supported mainly by a grant from OMRON Corporation. It is also supported in part by National Science Foundation of China under grant No.60332010.

References

- [1] S. Avidan. Ensemble tracking. In *CVPR*, 2005.
- [2] S. Birchfield. Elliptical head tracking using intensity gradients and color histograms. In *CVPR*, 1998.
- [3] M. L. Cascia, S. Sclaroff, and V. Athitsos. Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models. *PAMI*, 22(4):322–336, 2000.
- [4] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *CVPR*, 2000.
- [5] G. Hua and W. Y. Multi-scale visual tracking by sequential belief propagation. In *CVPR*, 2004.
- [6] C. Huang, H. Ai, Y. Li, and S. Lao. Vector boosting for rotation invariant multi-view face detection. In *ICCV*, 2005.
- [7] M. Isard and A. Blake. Condensation – conditional density propagation for visual tracking. *IJCV*, 28(1):5–28, 1998.
- [8] K. Lee and D. Kriegman. Online learning of probabilistic appearance manifolds for video-based recognition and tracking. In *CVPR*, 2005.
- [9] K. Okuma, A. Taleghani, D. Freitas, J. J. Little, and D. G. Lowe. A boosted particle filter: Multitarget detection and tracking. In *ECCV*, 2004.
- [10] P. Perez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In *ECCV*, 2002.
- [11] P. Perez, J. Vermaak, and A. Blake. Data fusion for visual tracking with particles. *Proceedings of IEEE (issue on State Estimation)*, 2004.
- [12] H. A. Rowley. *Neural Network-based Human Face Detection*. PhD thesis, Carnegie Mellon University, 1999.
- [13] H. Schneiderman and T. Kanade. A statistical to 3d object detection applied to faces and cars. In *CVPR*, 2000.
- [14] C. Shen, A. Hengel, and A. Dick. Probabilistic multiple cue integration for particle filter based tracking. In *Digital Image Computing: Techniques and Applications*, 2003.
- [15] M. Spengler and B. Schiele. Toward robust multi-cue integration for visual tracking. In *Intl. Workshop Computer Vision Systems*, 2001.
- [16] R. C. Verma, C. Schmid, and K. Mikolajczyk. Face detection and tracking in a video by propagating detection probabilities. *PAMI*, 25(10):1215–1228, 2003.
- [17] J. Vermaak, A. Doucet, and P. Perez. Maintaining multi-modality through mixture tracking. In *ICCV*, 2003.
- [18] P. Viola and M. Jones. Robust real-time object detection. In *IEEE Workshop on Statistical and Theories of Computer Vision*, 2001.