Grid Particle Filter for Human Head Tracking Using 3D Model

Chenguang Liu 1  Jiafeng Liu 2  Jianhua Huang 3  Xianglong Tang 4

1 liu.cg@live.cn 2 jefferyliu@hit.edu.cn

Abstract
A new 3D head-shoulder model based particle filter is presented for finding human head in static images. Edge cues are used as the likelihood function of the proposed particle filter. The positions of head as well as its direction are evaluated simultaneously. At each time step, the proposed algorithm generates a discretized grid state space and takes a subspace of it as the searching space of the particles. Hence the number of particles is well limited and the performance of the particle filter is improved. The experimental results prove that our method is effective and robust.

Keywords: head tracking; face direction; 3D model; grid particle filter

1. Introduction
Monocular sequence based human tracking and analysis is promising for many applications, hence a number of researchers are interested in this field. The human head tracking is important in this researching field since it provides preconditions for the whole human body tracking, human face analysis, face recognition and so on.

Many algorithms for human head tracking use face characters to detect the position of head [1, 2]. Admittedly, the face is always bare and the characters are stable for tracking. But in many important applications such as surveillance, face is not available at any time. While there are also some algorithms take into account the contour of head or head-shoulder structure for head tracking [3, 4, 5, 6]. Jin et al. [3] fulfilled a method based on ellipse matching for human head tracking in complicated background. Yoon et al. [4] proposed a ‘Ω’ model and fulfilled a head tracker. However, both ellipse model and ‘Ω’ model are simple and liable to be affected by noises. Moreover, they do not provide 3D information and can not be used to judge the direction of head.

In this paper we propose a 3D head-shoulder model. This model is similar to the 2D model presented in [4, 5, 6]. But our 3D model is closer to the true structure of the human body and contains more 3D information; therefore the proposed model can better represent the deformation of the projection of the head-shoulder model. Besides, a grid particle filter is presented, which can not only track the position of head but also detect the direction of the face. The tracking result is compared to the literature [5].

2. Particle Filter
The human head tracking problem can be treated as a Bayesian problem. At time step $t$ the state of head is represented by $X_t$, thus the probability framework of the head tracking problem is expressed as
\[ p(X_t | I_t) = c \Phi(I_t | X_t) \int p(X_t | X_{t-1}) p(X_{t-1} | I_{t-1}) dX_{t-1} \] (1)

where \( I_t \) is observation template of the image, \( c \) is a normalization constant, \( p(X_t | I_t) \) is the posterior probability about \( I_t \), \( p(I_t | X_t) \) represents the likelihood measurement, \( p(X_t | X_{t-1}) \) is a dynamic model which predicts the next state of the object, \( p(X_{t-1} | I_{t-1}) \) is the prior probability which provide the prior information about the states of the head at time step \( t-1 \).

Generally, it’s difficult to directly calculate the integral in (1). While particle filter [7] is a method to evaluate the posterior from a set of non-Gaussian and multi-model observation data. The main idea of the particle filter is to represent the posterior density through a set of weighted particles. Therefore, the calculation of the integral is avoided by using particle filtering.

The particle filtering can be treated as a process of selecting ‘the most optimized particles’ from an optimized particle set. A typical particle filter consists of three major steps at time step \( t \): (i) weighted sampling process, during which \( N \) particles are selected on the distribution of the posterior at time step \( t-1 \). The possibility of selecting a particle is decided by its normalized weight; (ii) dynamic process, during which a dynamic model is used to update the \( N \) selected particles of which the states are predicted; (iii) measurement process, during which the observations of the current time step are obtained, and the weight of each of the particles is calculated by likelihood measurement.

2.1. Head-Shoulder 3D Model

We construct a head-shoulder model \( \psi(\nu, \theta_y, \theta_z) \) (Fig. 1) using two connected ellipsoids, of which the parameters are calculated by the proportion of the human body. The model \( \psi \) has three DOFs: \( \nu \) represents the scale, \( \theta_y \) represents the scrolling angle around the coordinate \( Y \), and \( \theta_z \) represents the scrolling angle around the coordinate \( Z \).

The relation between the world coordinate space \( W \) and the image coordinate space \( F \) is shown in Fig. 2, and the plane \( \nu \theta \) in \( W \) and the plane \( \nu \theta \) in \( F \) are parallel with each other.

Now, we project the 3D model \( \psi(\nu, \theta_y, \theta_z) \) to the image plane and then obtain the 2D edge template \( \kappa \{ x, y, \nu, \theta_y, \theta_z \} \), where \( x \) and \( y \) represent the coordinates of template \( \kappa \) in the image plane, \( \nu, \theta_y, \theta_z \) represent the three DOFs of the 3D model \( \psi \) corresponding to \( \kappa \).
The particle in 5 dimensional state space is defined as \( \phi(x, y, v, \theta_y, \theta_z, \omega) \), where \( \omega \) represents the weight of the particle, vector \( \psi(v, \theta_y, \theta_z) \) corresponds to a 3D model \( \psi \), and \((x, y)\) represents the position of the 2D template in the image plane. The coordinate of the head in the 2D template image is calculated beforehand; therefore, when the position of the 2D template is evaluated the position of the head in the image plane can be calculated consequently.

By tuning the value of the three DOFs of model \( \psi(v, \theta_y, \theta_z) \), we can obtain various configurations of the model. The set of different configurations consists of a 3D model database \( S \). Given that the sample rate of \( v, \theta_y \) and \( \theta_z \) are \( \Delta v, \Delta \theta_y \) and \( \Delta \theta_z \), respectively. The model database \( S \) is represented as

\[
S = \{ \psi(v', \theta_y', \theta_z') | v' \in P, \theta_y' \in G, \theta_z' \in Q \} \\
P = \{v | v = v_i + (i-1)\Delta v, i = 1...N_v\} \\
G = \{\theta_y | \theta_y = \theta_{y_0} + (i-1)\Delta \theta_y, i = 1...N_{\theta_y}\} \\
Q = \{\theta_z | \theta_z = \theta_{z_0} + (i-1)\Delta \theta_z, i = 1...N_{\theta_z}\} 
\]

where \( P, G \) and \( Q \) are the sets of the corresponding DOFs, respectively. There are \( N = N_p \times N_g \times N_q \) head-shoulder model in \( S \). The model database \( S \) optimized the searching space of the particle filtering. The state space could be discretized to the grid state space based on \( S \).

### 2.2. Grid Particle Filter

The proposed grid particle filter discretizes the 5 dimensional state space into a grid state space which is used as the searching space of particle \( \phi(x, y, v, \theta_y, \theta_z, \omega) \).

Each node in the discretized state space corresponds to a particle. Thus, the continues state space is discretized to a grid state space and the 'discretizing grade' depends on the sample rate on each dimensions. The bigger the sample rate is, the more efficient the performance is, the smaller the precision of evaluating the configuration is, or vice versa.

The proposed method does not need manually initialization. The grid particle filter contains three major steps at time step \( t \): (1) Weighted sampling process: selecting \( M \) particles with biggest weights from the particle set \( \Phi_{t-1} \) which is generated at time step \( t-1 \); (2) Prediction process: after the tracking process at time step \( t-1 \), the mean value \( \Phi_{t-1} \) is evaluated. As a matter of fact, the difference between two continues frames is very small, therefore, we use \( \Phi_{t-1} \) to predict the particles at time step \( t \). In the 5 dimensional grid state space, a super ellipsoid \( \Omega \) is obtained and its center is \( \Phi_{t-1} \). Then the \( N_p \) particles surrounded by the ellipsoid in the grid state space are predicted new particles. The volume of the ellipsoid can decide the value of \( N_p \) but the volume is decided by the width of each dimension. In order to increase the computing effectiveness, we limit the number of particles by setting reasonable value of the width.
of each dimension. We use (3) to go through the $N_t$ new particles.

$$\eta = \bar{\eta} - w + i \Delta \eta$$
$$\eta \in [\bar{\eta} - w, \bar{\eta} + w]$$  \hspace{1cm} (3)

where $\eta$ represents the value of one dimension on which the width of the super ellipsoid $\Omega$ is $2w + 1$, $\Delta \eta$ represents the interval between each pair of particles in the discretized state space.

Measurement process: each new particle $\phi_t(x, y, \theta, \theta, \omega)\phi_t(x, y, \theta, \theta, \omega)$ uniquely corresponds to a 2D template $\kappa_t(x, y, \theta, \theta, \omega)\kappa_t(x, y, \theta, \theta, \omega)$. An observation template $Z_t$ is the image region covered by $\kappa_t. \kappa_t$. The likelihood between $Z_t$ and the 2D model $\kappa_t$ is calculated as the weight of the particle $\phi_t. \phi_t$. After $N_t$ new particles’ weights have been updated, the particle set $\Phi_t$ is generated at time step $t$.

By finding $M$ particles with the biggest weights and calculating the mean value of them, the position of the head at the current time step is evaluated.

### 2.3. Likelihood measurement

Edge cue is used to construct the likelihood function which is shown as below:

$$\omega_c = \frac{N_c}{N_m} \tag{4}$$

where $N_m$ represents the number of pixels on the edge of the human body in the 2D template image. $N_c = \sum_{i=1}^{N} f(i)$ represents the number of matched pixels comparing the 2D edge template image with the observation image. $f(i)$ represents matching result of the $i$th pixel on the edge of the 2D template. The matching process is defined as:

$$f(i) = \begin{cases} 0 & \text{if } \sum G(x, y) = 0 \\ 1 & \text{if } \sum G(x, y) \neq 0 \end{cases} \tag{5}$$

where $G(x, y)$ represents the gray level of the point $(x, y)$ in a window which is decided by the center point $P_0'$. Given the coordinates of the $i$th pixel on the edge of the 2D template $(x_i, y_i), P_0'$ is the pixel at the same position on the observation edge image. The window’s size is $3 \times 3$ pixels.

### 2.4. Detection of Face Direction

After the grid particle filtering, we obtain the state vector $(\bar{x}, \bar{y}, \bar{\theta}, \bar{\theta}, \bar{\omega})$ of the head at time step $t$. Where vector

$$(\bar{x}, \bar{y}, \bar{\theta}, \bar{\theta}, \bar{\omega})$$

depict the 3D configuration of the head-shoulder structure and the face direction is represented by $\bar{\theta}_y$.

However, when projecting the 3D head-shoulder model to the image plane, the depth ambiguity will occur (Fig. 3). The head-shoulder model database doesn’t have the model of back view $(\theta_y \notin [-90, 90])$. Therefore, when the body is facing back (Fig. 3(b)), the tracker will get a wrong tracking result (Fig. 3(a)).

As a matter of fact, the color difference between the face and the hair. Hence color could be used to distinguish the face
and the back of head. The 2D S-V (Saturation and Value) color histogram $\tilde{H}_{\text{face}}$ and $\tilde{H}_{\text{back}}$ could be generated after the learning process beforehand. After the tracking process is finished, the position of head is calculated too. Then the histogram $H_{\text{head}}$ of a region of the head is calculated based on its position. Bhattacharyya distance [8] is used to describe the distance between $H_{\text{head}}$ and the two histograms $\tilde{H}_{\text{face}}$ and $\tilde{H}_{\text{back}}$. The face direction error is amended according to Eq. (6).

$$\theta' = \begin{cases} 
\theta_y & \text{if } B(H_{\text{head}}, \tilde{H}_{\text{face}}) \leq B(H_{\text{head}}, \tilde{H}_{\text{back}}) \\
180 + \theta_y & \text{if } B(H_{\text{head}}, \tilde{H}_{\text{face}}) > B(H_{\text{head}}, \tilde{H}_{\text{back}})
\end{cases}$$

(6)

where $B(H_{\text{head}}, \tilde{H}_{\text{face}})$ and $B(H_{\text{head}}, \tilde{H}_{\text{back}})$ represent the Bhattacharyya distances between the corresponding histograms.

### 3. Experimental Results

We perform our method on the HumanEva database [9] downloaded from Brown University. Several sequences, including walking, jogging and boxing, are used to test the proposed head tracking method.

When generating the 3D model database, the corresponding parameters are: $\nu \in [0.5, 2.0]$, $\theta_y \in [-90, 90]$ and $\theta_z \in [-20, 20]$; the corresponding sample rate is: $\Delta \nu = 0.1$, $\Delta \theta_y = 45$ and $\Delta \theta_z = 5$, hence totally 720 models are generated. When performing the prediction process of the grid particle filter, the values of the parameters on each of the dimensions are shown as Table 1.

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>$x_t$</th>
<th>$y_t$</th>
<th>$\nu_t$</th>
<th>$\theta_{y,t}$</th>
<th>$\theta_{z,t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Width $W$</td>
<td>$15 \nu_{t-1}$ (Pixel)</td>
<td>$15 \nu_{t-1}$ (Pixel)</td>
<td>0.2</td>
<td>$45^\circ$</td>
<td>$10^\circ$</td>
</tr>
<tr>
<td>Interval $\Delta \eta$</td>
<td>2 (Pixel)</td>
<td>2 (Pixel)</td>
<td>0.1</td>
<td>$45^\circ$</td>
<td>$5^\circ$</td>
</tr>
</tbody>
</table>

Table 1. The values of the dimensions when performing the prediction process at time step $t$.

The ground truth is annotated manually on the test sequences. The average tracking error is 6.97 pixels, and the detected face direction of the test sequences is shown as Fig. 4, Fig.5 and Fig. 6. The experiment is done on an Intel Core2 2.0 GHz PC and the average computation time is 0.41 second per frame. From the results we can see that our method is able to deal with the self occlusion problem. In literature [5] the average tracking error of head is 9.28 pixels and the time is more than 2 seconds per frame on an Intel-Xeon 2.8 GHz PC. Comparing to [5] on human head tracking, our method achieve a better and more effective result.
4. References


