FACE TRACKING USING RAO-BLACKWELLIZED PARTICLE FILTER AND POSE-DEPENDENT PROBABILISTIC PCA

Tiesheng Wang\(^{(a)}\), Irene Y.H. Gu\(^{(b)}\), Andrew Backhouse\(^{(b)}\), Pengfei Shi\(^{(a)}\)

\(^{(a)}\)Institute of IPPR, Shanghai Jiao Tong Univ., Shanghai, China
\(^{(b)}\)Dept. of Signals and Systems, Chalmers Univ. of Technology, Sweden
\{tieshengw,pfshi\}@sjtu.edu.cn, \{irenegu,andy@chalmers.se\}

ABSTRACT

This paper deals with tracking of face blobs containing pose changes. We propose a novel tracking method to deal with face pose changes during the tracking. In the method, tracking is formulated as an approximate solution to the MAP estimate of the state vector, consisting of a linear and a nonlinear part. Multi-pose face appearances are described by local linear models, each being related to a single pose and estimated by probabilistic PCA (PPCA). A Markov model with pose indices as its states is used to model the transitions between poses. Shape and locations of face blobs and the associated pose indices are assumed to be nonlinear, and are estimated by a Rao-Blackwellized particle filter (RBPF). This enables a separate estimation of the linear state vector through marginalizing the joint probability. The proposed method has been tested for videos containing frequent face pose changes and large illumination variations, where 5 poses (left, frontal, right, up, down) were modeled. The tracking results are shown to be robust to variable speed of pose changes and with relatively tight boxes.

Index Terms— Object tracking, video surveillance, Rao-Blackwellized particle filters, probabilistic PCA, object pose model, Markov pose model, object appearance model, MAP estimation.

1. INTRODUCTION

Object tracking has become an increasingly important subject. Blob-based tracking refers to tracking by using tight rectangular boxes surrounding the tracked objects. Visual object tracking based on appearance subspace learning [1, 2] is focused on the appearance related to linear transformation and simple distributions, and is not sufficiently robust for handling object pose changes. Other methods consider tracking with multiple-poses. [3] proposed tracking in nonlinear manifolds where global coordinates of local linear models were constructed followed by defining dynamic model according to global coordinates. However, it cannot simultaneously approximate the pose. [4] proposed tracking by using a learned probabilistic appearance manifold without including low dimensional representation of appearance as part of the state vector. In [5] a template matching was switched for face tracking. The weakness of using templates is its sensitivity to illumination changes. Other promising methods on linear model switching deal with different applications, e.g. fault analysis with RBPF [6], and occlusions in visual tracking [5].

Rao-Blackwellization is a technique to reducing variances by marginalizing out the linear part of the state vector [7]. Particle filters have been widely used for dealing with problems that are nonlinear and non-Gaussian. The main advantage of a Rao-Blackwellized particle filter (RBPF) is to divide the state vector into a linear and a nonlinear part. It is more efficient than a standard particle filter due to the probability, being inferred by the particle filter, has a lower dimension. In [6] a RBPF was used to inference on switching linear models, where the nonlinear state is a switching variable. A RBPF [2, 3] was used for eigen-tracking where the nonlinear part of the state vector is the motion parameters. In the blob-based face tracking, appearances corresponding to different poses were represented by their local coordinates. The nonlinear state includes a switching variable and the motion parameters. However, in the RBPF tracking [3], the global coordinates of face appearances must be first constructed before tracking, and the pose cannot be estimated during that time.

Motivated by the above, this paper proposes a novel multi-pose tracking method that combines nonlinear modeling of object shape (described by the surrounding box) and location, and linear modeling of multi-pose object appearances. We use a RBPF for solving our tracking problem however without requiring the construction of global coordinates of appearance manifold. An important assumption on local coordinate transformations is made, that enables the switching between the poses. For modeling the appearance of multi-pose objects, we propose to use a probabilistic PCA (PPCA) in combination with a Markov pose model, where the former is used to model each individual pose and the latter is to model the dynamics of pose. The novelty of such a model is that poses are
embedded as a variable in the state vector.

2. SYSTEM DESCRIPTION

In the proposed system, face tracking is formulated under a nonlinear state space model, where the state vector is split into a linear part and a nonlinear part. The basic idea of splitting the state vector into two subvectors is to enable the use of effective estimation methods to each part of the vector in a marginalized joint probability. The marginalization is obtained by Rao-Blackwellizing the state vector, where one may apply separate models to different parts of the state vector. To enable an effective tracking where a 3D object may appear significantly different in a 2D image as a 2D planar object, enable an effective tracking where a 3D object may appear through an affine transformation, denoted as \( y_t = T(I_t; l_t) \), \( A_t(k) \) and \( C_t(k) \) are the measurement matrix and the system matrix, and the two Gaussian noise items \( v_t(k) \) and \( w_t(k) \) are associated with the dynamics and observations, \( v_t(k) \sim N(0, Q_v(k)) \), \( w_t(k) \sim N(0, Q_w(k)) \). The tracking problem is formulated as finding the approximate of the state vector from observations \( y_{1:t} \) through maximum a posteriori (MAP) estimation of \( p(x_t, l_t, s_t|y_{1:t}) \) in the above switching linear models, where \( x_t \) is marginalized out by the Rao-Blackwellization approach.

4. RAO-BLACKWELLIZED PARTICLE FILTER FOR NONLINEAR STATE VECTOR ESTIMATION

Marginalizing out part of the state vector to estimate remaining part of the state vector by the Rao-Blackwellized particle filter is a frequently used technique [2, 7]. In our system, we use the Rao-Blackwellization method to marginalize out the linear part of the state vector from the nonlinear part. In such a way, the particle filter can be applied to the nonlinear part of a relatively low dimension, and hence is more computationally feasible. Denote the nonlinear part of the state vector as \( n_t = [l_t, s_t]^T \), then the posterior pdf \( p(x_t, n_t|y_{1:t}) \) can be computed by,

\[
p(x_t, n_t|y_{1:t}) \propto p(y_t|x_t, n_t)p(x_t, n_t|y_{1:t-1})
\]

where \( x_t \) is marginalized out by using the integration to (2)

\[
p(n_t|y_{1:t}) \propto \int p(y_t|x_t, n_t)p(x_t, n_t|y_{1:t-1})dx_t
\]

and then estimated by a separate approach. The 2nd term in the right hand side of (3), after using the chain rule and assuming \( p(n_t|x_{t-1}, n_{t-1}) = p(n_t|n_{t-1}) \), can be written as

\[
p(x_t, n_t|y_{1:t-1}) = \int p(x_t, n_t|x_{t-1}, n_{t-1})p(n_t|n_{t-1})dx_t
\]

A set of particles together with their weights \( \{n_{t-1}^{(i)}, w_{t-1}^{(i)}\} \) are used to approximate \( p(n_{t-1}|y_{1:t-1}) \). Then (4) becomes,

\[
\frac{1}{N} \sum_{i=1}^{N} w_{t-1}^{(i)}p(n_t|n_{t-1}^{(i)}) \int p(x_t|n_t, x_{t-1}, n_{t-1}^{(i)})p(x_{t-1}|n_{t-1}^{(i)}, y_{1:t-1})dx_{t-1}
\]

Let \( n_{t-1}^{(j)} = [l_{t-1}^{(j)}, s_{t-1}^{(j)}] \) be the \( j \)th particle drawn according to \( \sum_{i=1}^{N} p(n_t|n_{t-1}^{(i)}) \), \( j = 1, \ldots, N \), using the CONDENSATION version of the particle filter [10]. Assuming
The weight $w_t^{(j)}$ for the $j$-th particle $n_t^{(j)}$ is derived from:

$$
\int p(y_t | x_t, n_t^{(j)}) \int p(x_{t+1} | x_t, n_t^{(j)}) p(x_{t-1} | n_{t-1}, y_{t-1}) dx_{t+1} dx_t
$$

where the second term in (6) is further simplified to $p(x_t | n_t^{(j)}, x_{t-1}, n_{t-1}) = p(x_t | x_{t-1}, s_t^{(j)}, s_{t-1}^{(j)})$. In this way $p(n_t | y_{1:t})$ is approximated by a set of particles and their weights $\{n_t^{(j)}, w_t^{(j)}\}$. The assignment of $w_t^{(j)}$ is described in the next section.

5. PPCA AND MARKOV MODEL FOR TRACKING

For estimating the linear part of the state vector related to the image content within the box, a linear dimension reduction method, PPCA, is used for modeling each single-pose face appearance. PPCA provides a probabilistic explanation of PCA and is related to a factor analysis model by [8]

$$
y = Ax + v
$$

where $A$ is the transformation matrix, $v$ the low dimensional state vector described by the PCA of the boxed image content, and $v$ is the noise, and $y$ is mean centered. PPCA relates to the factor analysis model by assuming $v$ as additive spherical Gaussian noise, $v \sim N(0, \sigma^2 I)$, and $p(x) = N(0, I)$. This leads to the following distributions,

$$
p(y) = N(0, AA^T + \sigma^2 I), \quad p(y|x) = N(Ax, \sigma^2 I), \quad p(x|y) = N(\tilde{A}x, \sigma^2 M^{-1})
$$

where $M = (A^T A + \sigma^2 I)$. Given $y$, the posterior estimate of $x$ is obtained by the expectation of $p(x|y)$, i.e., $M^{-1} A^T y$. The ML estimate of $\sigma^2$ and $A$ is:

$$
\hat{\sigma}^2 = 1/(n-m) \sum_{i=m+1}^n \lambda_i; \quad \tilde{A} = U_m(A_m - \sigma^2 I)^{1/2}
$$

where $n$ is the dimension of the image blob $y$, $m$ is the number of principal components, columns of $U_m$ contain the principal eigenvectors of the sample covariance matrix $S$, and $A_m = \text{diag}(\lambda_1, \ldots, \lambda_m)$ contains the corresponding eigenvalues in the descending order. Based on these, pose $k$ dependent $A_k$, $Q_w(k) = \sigma^2(k) I$, and the mean $m(k)$ are learned.

The dynamic of motion parameter vector is assumed to be $p(l_t | l_{t-1}) \sim N(0, \text{diag}(\sigma_x^2, \sigma_y^2, \sigma_z^2, \sigma_r^2, \sigma_s^2, \sigma_a^2))$, where $\sigma_x^2, \sigma_y^2, \sigma_z^2, \sigma_r^2, \sigma_s^2, \sigma_a^2$ are variances for the translation in $x$ and $y$ directions, the rotation, the scaling, and the aspect ratio of the box, respectively.

To include multiple poses in the model, a Markov model is used. We assume that there are $M$ poses. The probability of a face motion from a previous pose $s_{t-1} = j$ to a current pose $s_t = i$, ($i, j = 1, \ldots, M$), is described by the transition matrix $T$ with

$$
T(j, i) = P(s_t = i | s_{t-1} = j)
$$

of a face motion from a previous pose $s_{t-1} = j$ to a current pose $s_t = i$, ($i, j = 1, \ldots, M$), is described by the transition matrix $T$ with

$$
T(j, i) = P(s_t = i | s_{t-1} = j)
$$

where $T(j, i)$ is the $(j, i)$ entry of $T$.

A Kalman filter is used to update the state vector $x_t$ under a given pose. Since $x_t$ is assumed to change smoothly with time, a simple relation $C(k) = I$ is obtained. The weights $w_k^{(j)}$ of particles in (6) are obtained from Kalman prediction:

$$
w_k^{(j)} = N(y_k^{(j)}; m(k) + A_k \tilde{x}_k^{(j)}, A_k \hat{P}_k^{(j)} A_k^T + Q_w(k))
$$

where $y_k^{(j)} = k$ and $y_k^{(j)} = T(l_k)$, MAP estimates of the pose index $\hat{s}_k = s_k^{(i*)}$ and the motion vector $\hat{l}_k = l_k^{(i*)}$ are obtained by choosing $i^* = \arg \max_{i} w_k^{(j)}$. If $s_k^{(j)} \neq s_k^{(j)}$, then there is a switching of pose-dependent appearance from $s_{t-1}(s_k^{(j)})$ to $s_{t-1}(s_k^{(j)})$, so that the dynamical model is associated with the same pose. In such a case, the input to the Kalman filter is re-initialized by

$$
\tilde{x}_k^{(j)}(k) = M(k)^{-1} A_k^T (y_k^{(j)} - m(k)); \quad \hat{P}_k^{(j-1)} = P_0(k)
$$

where $s_k^{(j)} = k$. Table 1 summarizes the algorithm of the proposed system.

6. EXPERIMENTS AND RESULTS

The proposed method has been tested on several face image sequences used in [4], where faces contain complex motion and illumination changes. Faces for training the PPCA models were collected from image frames containing the corresponding poses, and were trained offline for the time being.
Other videos (not used in the training) from the same person were used for the evaluation. The box size for each face image was normalized to $32 \times 32$ pixels. Small variances $\sigma_x^2 = 5^2$, $\sigma_y^2 = 5^2$, $\sigma_z^2 = 0.02^2$, $\sigma_v^2 = 0.1^2$, $\sigma_w^2 = 0.001^2$ were set to the parameters related to the box location and were determined empirically. In our tests, 5 poses (left, frontal, right, up and down, denoted as 1, 2, 3, 4, 5) were used with the corresponding pose transition probability matrices $T$ in (9),

$$
T_5 = \begin{bmatrix}
.85 & .05 & .0 & .0 \\
.1 & .6 & .1 & .1 \\
.05 & .185 & .0 & .0 \\
.0 & .1 & .85 & .05 \\
.0 & .1 & .05 & .85
\end{bmatrix}.
$$

These values were empirically determined, and were found to be insensitive to a range of values in our tests. Five PPCA models were trained corresponding to the 5 poses. Four principal components were used in each PPCA model. $N=500$ particles were used in the particle filter for estimating $n_t$. Fig.2 shows an example that contains some tracking results from three testing videos. By careful visual inspecting the tracking results through several image sequences, we have observed that the proposed method has well adapted to the pose changes and rightly tracked the faces through the sequences. Comparing with our previous work [9], we observed that the proposed method is more robust in tracking faces with fast pose changes and with a more tight box. The first row of Fig.3 shows the estimated pose changes for two videos. The pose changes are rightly estimated based on the visual inspection. To evaluate the performance, some preliminary tests were performed. We computed the distances between the nose tip (manually marked) and the tracked box center (the second row of Fig.3). Observing the results in Fig.3, we can see that there are frequent pose changes, especially in the video ‘chia’. We also observed the fluctuations in the resulting distance values, indicating the change of positions of nose tip in the box. From the figure, some relatively large distances were observed. We then noticed that these corresponding image frames are mostly related to non-frontal poses, where nose positions are no longer near to the center of the box (also indicated in Fig.2). Therefore, the selected distance measure only partially reflected the tracking performance. This indicated that apart from the visual inspection, more effective evaluation methods should be investigated.

### 7. CONCLUSIONS

The proposed tracking system models the transition of face poses with a Markov model and models face appearance according to each pose with the probabilistic PCA. The state vector of such a system includes the shape and location of face box, the pose index, and low dimensional representation of face appearance. The RBPF is used for inference this mixture state system to obtain the approximate solution to the MAP estimate of the state vector related to face blobs of interest. Our tests have shown that the particle filters are efficient for the low dimensional nonlinear state vector. Also, the switching linear models for multi-pose face appearance is effective. Robust face tracking results have been obtained for several image sequences. Further extensive tests and evaluations will be carried out.

### 8. REFERENCES


