Robust Object Tracking Using an Adaptive Color Model

Gi-jeong Jang and In-So Kweon

Department of Electrical Engineering & Computer Science
Korea Advanced Institute of Science and Technology
373-1 Kuseong-dong, Yuseong-gu, Taejon, Korea
gjjang@covra1.kaist.ac.kr, iskweon@kaist.ac.kr

Abstract

In this paper we present a new robust face tracking method based on the Condensation algorithm [1, 2] that uses a sampling based density representation. A two-dimensional color model is used to approximate the face color. We modified the Condensation algorithm to provide color adaptability to the abrupt change of illumination and to the tracking of differently colored people. According to the face size and location uncertainty, the searching range is automatically determined and it makes the algorithm extremely robust and efficient.

The tracker operates at real-time and actively controls a camera pan-tilt in order to locate a person’s face in the center of the image. Experimental results show the algorithm’s robustness to the agile motion of face and to the dramatic change of illumination in the presence of complex background.

1. Introduction

Robust real-time face tracking has many applications in HCI (Human Computer Interaction). Face tracking is used not only for directing the computer’s attention to a user but also for segmentation of face as a prerequisite stage for recognition and understanding. This enhances the performance of such a system as face identification, facial expression and emotion recognition, and lip reading. Nowadays many applications are in direct commercial use such as video conferencing, automatic surveillance system.

There have been a large variety of face tracking algorithms.

For a simplest form, there is a technique of background differencing [3, 4], where pixelwise difference has been considered. The difference map is usually binarized using a predefined threshold to classify foreground region and then find face region. This is only effective for a static camera. As a similar method, interframe difference [5] was introduced for a moving camera. However it has some limitations. When the face is static or the region is untextured uniform, the difference map become unclear.

Model-based approaches [3, 6, 7] use the face model space, which is either the real 3D world or the 2D projection (image) space. During tracking, each input video image is projected onto the surface texture map of the face model. Model parameters such as scaling, translation, rotation, deformation, and illumination condition are updated via image registration in the texture map space. These methods are more reliable and accurate than the difference-based methods, but suffer from high computational cost. If the model is simplified as, for example, 2D planar, we can save the computational cost but the system will lose the robustness to the facial rotation.

In color-based approaches [8, 9, 10, 11], the face color is assumed to be unimodal in color space. Therefore color model can be easily expressed as statistical parameters (mean, variance) or color histogram. These methods are relatively robust to facial orientation and partial occlusion, also simple and fast, but not robust to the change of lighting condition. There have been attempts to cope with this problem: Using a recursive filter to modify the model color itself or using additional parameters describing the illumination condition. But these are not satisfactory when the illumination changes abruptly.

In this paper, we present a real-time algorithm to track a person’s face using Condensation in which a two-dimensional color model is used to approximate the face color. We modified the Condensation algorithm to provide color adaptability to the abrupt change of illumination and to tracking of differently colored people. According to the face size and location uncertainty the searching range is automatically determined through prediction and update process. With the sample density on the image, we represent the propagation of probability densities over time. The predefined positional constraints between face and body make the system more robust. Thanks to the real-time camera pan-tilt control, the moving camera provides a large searching range and this actually slow down the relative speed of the face in the acquired image. So we can obtain more accurate and robust estimation of...
the position and size of the tracked face in the presence of complex background.

2. Face Tracking Algorithm

2.1 Review of Condensation (Conditional Density Propagation) algorithm

In the proposed face tracking system, we estimate the state of the face at the current time step \( k \). Here, the state vector \( \mathbf{x}_k \) represents only sample position.

\[
\mathbf{x}_k = (x_{k-1}, y_{k-1}, x_{k-1}'m - 1, y_{k-1}'m - 1, \ldots, x_k, y_k)^T
\]

where, \( m \) = order of the motion model-1.

Given the information about the initial state and all measurements \( Z^k = \{ z_i, i=1..k \} \) up to the current time, we can construct the posterior density \( p(\mathbf{x}_k \mid Z^k) \) of the current state. Here, after initializing the candidate face region we choose the mean of the posterior density as estimator, because during the tracking period the density can be assumed as unimodal.

To track the face we recursively calculate the density \( p(\mathbf{x}_k \mid Z^k) \) at each time step. This is done alternating two phases: prediction and update.

In the prediction phase we use a motion model from which conditional density \( p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_{k-1}) \) is calculated under the assumption \( \mathbf{x}_k \) is dependent only on the previous state \( \mathbf{x}_{k-1} \) and the input \( \mathbf{u}_{k-1} \). So the motion model can be expressed as

\[
\mathbf{x}_k = A\mathbf{x}_{k-1} + B\mathbf{u}_{k-1} \quad (1)
\]

where \( A \) is a matrix representing dynamic characteristic, \( B \) denotes a scaling factor, \( \mathbf{w}_k \) means Gaussian noise vector.

The current position of the face can be predicted in the form of PDF \( p(\mathbf{x}_k \mid Z^{k-1}) \):

\[
p(\mathbf{x}_k \mid Z^{k-1}) = \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_{k-1})p(\mathbf{x}_{k-1} \mid Z^{k-1})d\mathbf{x}_{k-1} \quad (2)
\]

In the update phase, the posterior PDF \( p(\mathbf{x}_k \mid Z^k) \) is obtained. We use a measurement model that is given in terms of likelihood \( p(z_k \mid \mathbf{x}_k) \). Here, the model is expressed as a histogram for face color. From the color of the sampled point, we calculate the measurement by

\[
p(\mathbf{x}_k \mid Z^k) = c_z p(z_k \mid \mathbf{x}_k) p(\mathbf{x}_k \mid Z^{k-1}) \quad (3)
\]

where \( c_z \) is normalization factor.

After the update phase, the process is repeated recursively. At the initial time step, without prior information we assume that the initial state is uniform.

2.2 The implementation of the face tracking system

In this section, we present the implementation details of the proposed face tracking system.

Condensation is a sampling-based tracking method, which represents the posterior probability density \( p(\mathbf{x}_k \mid Z^k) \) with the local density of a set of \( N \) random samples or \( S_k = \{ s'_i, i=1..N \} \) as described in the previous section.

The goal is then to recursively compute at each time step \( k \) the set of samples \( S_k \) that is drawn from \( p(\mathbf{x}_k \mid Z^k) \).

![Figure 1: Structure of face tracking system](image)

In the prediction phase, with the set of particles \( S_{k-1} \), apply the motion model to each particle \( s'_{k-1} \) by sampling from the density \( p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_{k-1}) \). As a result a new set \( S'_k \) is obtained that approximates the predictive density \( p(\mathbf{x}_k \mid Z^{k-1}) \).

The update phase includes “measure” and “select” block as shown in Figure 1. In this phase, the measurement \( z_k \) is found first, according to the newly adapted color model. This is done by the color histogram look-up. Then the samples in \( S'_k \) are weighted by \( \pi'_k = p(z_k \mid s'_k) \). We then obtain \( S_k \) by resampling from the weighted set. The number of resampling is determined in proportion to the weight scale.

Sample propagation scheme is depicted in Figure 2, where the circle stands for the face colored region. Samples (point) are spread around the expected position to get the state (face position), then they are measured according to how much they are close to the face color model. Figure 3 shows an example on real image.

At the initial time step, we uniformly spread the samples on the whole image. Then, before the posterior PDF \( p(\mathbf{x}_k \mid Z^k) \) is close to unimodal, in other words, before the samples gather into a small region, the stage remains localization period. In this period, the motion model is not applied and only the general face color model is used, because the prior knowledge is not confident. But
in the proposed method, the localization period is relatively short and the transition from localization to tracking is automatic and seamless.

After the face position is determined, we control the pan-tilt unit to place the face in the center of the image. Pan-tilt control is processed with an independent thread.

Figure 2: Sample propagation scheme

Figure 3: Sample propagation on real image

3. Adaptive Color Model

3.1 Color model

It was found that the distribution of skin colors of different people can be grouped to a small region in the chromatic color space [5]. Although skin colors of different people vary over a wide range, they differ much less in color than in brightness. So we choose the chromaticity space as color model:

\[ r = R / (R + G + B) \]
\[ g = G / (R + G + B) \]  

(4)

Figure 4: Histograms for face color

We make a two dimensional histogram to represent face color as shown in Figure 4. It shows the skin color model for general face (a) and for one face (b). We can observe that the two histograms are both near Gaussian distributions but (a) spreads more wide. Larger variations of face color and illumination will result in a larger variance of the histogram.

Therefore when the general face color model is used, tracked face color cannot be represented well and larger variance of the model will increase the probability to make the tracker diverge to similarly colored background. After the tracker is fixed to the expected face, the adaptation process of the color model should be necessarily considered.

3.2 The structure of adaptive color model

The adaptive color model uses the basic structure of Condensation. The difference is that the tracking object is color histogram in the chromatic color space.

The procedure for color model adaptation is as follows:

1. Randomly assign the selected samples in the previous frame to the predicted samples in the present frame.

\[ s^{(n)}_{k} < - - - s^{(j)}_{k-1} \]  

(6)

2. Measure the color distance between the two samples in the previous step in the chromatic color space. Then using this distance, find the value of the Gaussian function of one face.

\[ \pi_{1, k}^{(n)} = N(\| s^{(n)}_{k} - s^{(j)}_{k-1} \|; 0, \sigma_{one}) \]  

(7)

where, \( N(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ - \frac{(x-\mu)^2}{2\sigma^2} \right\} \)  

(8)

3. With the color of the predicted sample, find the value of the Gaussian function of the general face.

\[ \pi_{2, k}^{(n)} = N(\| s^{(n)}_{k} - m \|; 0, \sigma_{general}) \]  

(9)

4. Set the measurement value multiplying the two values from the step 2, 3. Then select new samples according to the given measurement to construct a new color model histogram.

\[ \pi_{1, k}^{(n)} = \pi_{1, k}^{(n)} \ast \pi_{2, k}^{(n)} \]  

(10)
As seen in Figure 5, when the illumination changes slowly or even abruptly, the proposed algorithm moves the face color region in the model to the nearest from the previous position in color space. To accomplish this, we assigned high measurement value to the samples that are close to the color of previous samples. To prevent the face color model from moving to far from the general face color, $\pi_2$ was multiplied to $\pi_1$.

Figure 6 shows the mean of face color drifts on the chromatic color space when the illumination changes abruptly. The abrupt color changes in r and g components depict the successful tracking and color adaptation.

4. Estimation of Face Size

Face segmentation using color can be accomplished by a simple pixel grouping process such as region growing. The proposed method estimates the face size just in the course of resampling (selection) of predicted points by bounding the region of the samples.

\[
\begin{bmatrix}
0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & B' \\
\end{bmatrix}
\]

where, 

\[
\begin{bmatrix}
\sigma & 0 \\
0 & \sigma \\
\end{bmatrix}
\]

The order of this matrix will be determined by the order of the system.

In Condensation algorithm, when $\sigma$ is large, samples spread wide on the image. It widen the searching range but enlarge the possibility of distracting the tracker to the background clutter. On the other hand, when $\sigma$ is small, samples assemble to a narrow part of tracked area. This reduces the searching range, and tracker may fail to catch up the dynamically moving face. Therefore we need to adjust $\sigma$ according to the face size on the image.

When the portion of the face region on the image is large, $\sigma$ becomes greater to spread the samples more wide. On the other hand, when the portion of the face region on the image is small, $\sigma$ is made smaller to gather the samples near the face region.

When the number of samples is predetermined, if most of the samples are contained in the face region, the sum of measurements of all samples will increase, in the opposite case, will decrease. So when the sum of measurements goes below the predetermined threshold, we diminish $\sigma$, otherwise, increase $\sigma$. Thanks to the fact that the face color model is adjusted to the tracked face color, we can use the same threshold.

In Figure 7 (a) shows the case where $\sigma$ is too large. Too many samples are out of the face region. Samples are confined to a small region in (b) for a small $\sigma$. Adaptive $\sigma$ by the proposed method produce an accurate estimation of face size as shown in (c) and (d).

5. Learning dynamics of the motion

In (1), $A$ is the matrix representing the dynamics of the object motion. Learning dynamics means determining the elements of matrix $A$ using the sequence of tracked position at each frame of the learning data. $A$ can be calculated using MLE (Maximum likelihood Estimation) [12].

Generally the dynamics of face is very irregular, so modeling such a dynamics with a single matrix is almost impossible and the output of MLE is meaningless. But in the proposed system, with the help of fan-tilt control, the trajectory of the face oscillates around the center of the
image as shown in Figure 8. With the oscillating characteristics, the dynamics can be well described with only $A$. The estimated matrix $A$ can be used to predict the face position more accurately.

With learned dynamics we could get the satisfactory enhancement of tracking performance as Figure 9. In (a) the predicted position of samples are approximately in the face, but in (b) the samples are much shifted right.

![Figure 9: Predicted position of samples (a) with / (b) without dynamic model](image)

### 6. Geometrical Constraints

The robustness can be enhanced using some constraints [13]. Here, we use the geometrical relation between the face and the body. The color model of the body is predetermined and two Condensation modules are processed for each part. Because we can assume the position of the face is above the body, the prediction of the face position can be more accurate. Even when one of the two parts is fully occluded, we can predict the occluded part with reasonable accuracy as shown in Figure 10. When the occluded part begins to reappear, the tracker can resume the tracking process rapidly.

When we fail to track both of the two parts, the system is set to reinitialize, spreading the samples uniformly over the whole image.

![Figure 10: Prediction of occluded region using geometrical constraints](image)

### 7. Experimental Results

We have implemented the proposed face tracking algorithm on a Pentium III 500MHz PC with 320*240 image size. And it has been tested in different environments with various backgrounds and lighting conditions even in the change of illumination both for gradual and abrupt case.

The proposed face tracking algorithm requires a relatively small amount of computation. Without the time consumed to grab images, the net time for the track algorithm is 1.5 sec to process 100 images. In this case, 1000 samples were used to run the Condensation algorithm, which was sufficient for the testing environments.

Figure 11 shows an example of tracking results with the change of illumination. In the fourth image the illumination changes abruptly, but tracking is successful without any disturbance. Figure 12 shows an example of tracking both face and body in very dynamic environments. After the full occlusion, the tracker recovers fast.

![Figure 11: Tracking result with the change of illumination](image)

![Figure 12: Tracking result with full occlusion](image)
8. Conclusion

In this paper, we present a real-time face tracking algorithm which is based on a robust and efficient statistical method named Condensation. One major unique feature of the proposed algorithm is that it adapts its face color model to the tracked face at each frame. We set another tracker to track the face color model itself in the chromatic color space. This enables the system robust in the presence of abrupt change of illumination and the change of tracked face color. The segmenting face region is very simple. With the Condensation algorithm, the examined region is automatically selected near the face. So we just only have to bound the region of selected samples to segment the face. Using the constraints of relative position between the face and the body, we could enhance the prediction accuracy.

Work is currently underway to develop a multi-resolutional version of the proposed method by combining various features such as color, motion, face pattern, contour or edgy for more robust tracking.

References