Convex hull과 Robust Hausdorff Distance를 이용한 실시간 얼굴 트래킹

박민식, 박창우, 박민용
연세대학교 전기 전자공학과 지능제어 연구실

A New Face Tracking Algorithm Using Convex-hull and Hausdorff Distance

ICS Lab, Dept. of Electrical and Electronic Engineering, Yonsei University
134, Shinchon-dong, Seodemun-gu, Seoul, Korea
Tel: (02) 2123-2868, Fax: (02) 312-2333, p120100@hanmail.net

Abstract - *This paper describes a system for tracking a face in a input video sequence using facial convex hull based facial segmentation and a robust hausdorff distance. The algorithm adapts YCbCr color model for classifying face region by [1]. Then, we obtain an initial face model with preprocessing and convex hull. For tracking, a Robust Hausdorff distance is computed and the best possible displacement is selected. Finally, the previous face model is updated using the displacement t. It is robust to some noises and outliers. We provide an example to illustrate the proposed tracking algorithm in video sequences obtained from CCD camera.*

Key words : robust hausdorff distance, convex hull, face segmentation, tracking

1. Introduction

Face tracking is an challenging and important problem in computer vision. This subject has been attracting researchers for many years. It has a wide request for many applications such as, interactive human-computer communications, virtual reality interfaces, and automated surveilance systems and video conferencing. The complexity of these issues depends on environment under which the algorithm is applied. In this paper, we applied the tracking algorithm to face tracking in video sequences obtained from CCD camera in laboratory environment.

So far, many different approaches for face segmentation or detection and tracking have been proposed in literature[6]. Eigenfunction and neural network(NNW) are commonly used techniques to locate and detect human faces. Eigenfunctions are specific for a particular pattern. So, it has limitation of the variation on view point. NNWs perform well to detect faces in still images, but difficult to train on scenes with no-faces in the scenes. The approach of skin color is simple and nonparametric forms of density estimation, but if there is not sufficiently accurate model for apparent color, good parametric models for density estimation cannot be obtained. [1] Feature-based approach has the flexibility to be extended to a difficult scale, orientation and viewpoints of faces.[3]

A robust matching method must be able to detect objects that are undergoing translation, rotation, and changes in shape. This excludes basic template matching where the new position is determined by the highest correlation between the model and subsequent frames. The generalized Hough transform, which is used in [4], has been successfully applied to the detection of arbitrarily shaped 2-D binary objects. However, it comes at a high computational cost, especially for a multidimensional Hough accumulator space that includes translation, rotation, and scaling.

Our goal is to track a human face in a video sequence. We find the face location by adapting, in each frame, intersection relationship(ICH) between the convex hull of skin color regions(SCH) and convex hull of hair color regions(HCH) and model matching by robust hausdorff distance measure[2]. In the first frame, we segment a face and track the face using robust hausdorff distance measure.

The performance of the proposed algorithm is shown by experiment. Experimental results show that the proposed algorithm successfully and efficiently tracks the faces in a video sequence.

2. SKin Color and Hair Color region

The segmentation of a face in video sequences is difficult work because of occlusion, cluttered backgrounds, and variation of viewpoint and illumination. The importance of face segmentation is illustrated by its many applications, such as, object-based image encoding, modeling, face recognition and face tracking. Recently, there have been many approaches on face segmentation, but no single method assures robustness. Some proposed approaches involved the combination of shape, motion, and statistical analysis for robustness[6]. In these days, a new approach using color information is proposed. In this paper, we will use the color analysis approach for face segmentation.

2.1 Color segmentation

The first work is classifying pixels of the first frame in video sequence into skin color and non-skin color. In order to do so, we use a skin-color reference map in YCbCr color space by D.chi[1]. The skin-color reference map that is R_{Cb}=[77 127] and R_{Cr}=[133 173] is defined by [1]. With this skin-color reference map, the skin-color segmentation can now begin. Since we are utilizing only the color information, the segmentation requires only the chrominance component of the input image.

The result of this step is described as

$$S(x,y)=\begin{cases}1, & \text{if } [Cb(x,y)\in R_{Cb}]\cap[Cr(x,y)\in R_{Cr}]\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

where R_{Cb} and R_{Cr} are respective skin-color rages of Cb and Cr in YCbCr color space

The output pixel at point (x,y) is classified as skin color and set to one if both the Cr and Cb values at that point fall inside their respective range R_{Cb} and R_{Cr} . Otherwise, the pixel is classified as non-skin color and set to zero.

Next, we must make a hair color reference map for finding hair color region in the first frame. In this paper, the hair color reference map is based on an oriental. The hair color region often show low brightness and chromaticity estimation of low brightness color is not stable. Therefore, the hair color reference map has not sensitive property from luminance condition. Then, we use RGB color space for obtaining hair color region. The hair color reference map that is R_R=[0 30], R_B=[0 30], R_G=[0 30] is defined by experimental result. We find the image points including RGB color space from (0,0,0) to (30,30,30) to detect hair color region.

We can detect the hair color region of the first frame in video sequence from (2)

$$S(x,y)=\begin{cases}1, & \text{if } [R(x,y)\in R_R]\cap[G(x,y)\in R_G\cap[B(x,y)\in R_B]\\ 0, & \text{otherwise}\end{cases}$$

(2)

where is R_R, R_B, R_G are respective hair color ranges of R, G and B in RGB color space

In the next step, we apply the opening operator to remove a noise and find two skin color regions and hair color regions.

2.2 Face Segmentation via convex hull

When we have a large number of points to process, we're interested in the boundaries of the point set. People looking at a diagram of a set of points plotted in the plane, have little trouble distinguishing those in the "inside" of the point set from those lying on the edge[17].

The mathematical name for the natural boundary of a points in the plane is defined to be the smallest convex polygon containing them all. Equivalently, the convex hull is the shortest path surrounding the points.

After preprocessing of the first frame, we assign labeling to each regions in the preprocessed image and make convex hull of hair color region as Hj and that of face color region as Fi. (i=1~n, j=1~m) and intersection region as Iij. Usually, the skin color region and hair color regions have been closely related with intersection of the skin color regions and hair color regions has a very strong possibility that they may be the face and hair.

$$\{\, I_{ij}(x,y)=F_i(x,y)\cap H_j(x,y)\} \qquad \begin{cases}F_i(x,y),\ H_j(x,y)=1 & \text{if } n[I_{ij}(x,y)]>\tau\\ 0 & \textit{otherwise}\end{cases} \tag{3}$$

We can find the facial location in the frame from Iij, which is characterized by (3)

3. Face Tracking Algorithm

Motion estimation is still very challenging task because the apparent motion flow is highly sensitive to noise. Thus, a robust matching algorithm must be able to detect moving faces which are rotating and translating in shape. If a face is segmented and facial location is reliably detected, the face can be tracked over time in an image sequences. In this section, we propose a new automatic tracking algorithm using robust hausdorff distance.

3.1 Robust Hausdorff Distance

Color or gray scale imges are not suitable for template or object matching because they are too sensitive to changes in illumination. Thus, low-level matching algorithms using distance transform and Hausforff distance have been investigated. They are simple and insensitive to the changes of image characteristics.

The Hausdorff distance measure computes the distance between two sets of edge points obtained from the model of previous frame and an convex hull's information of face region in current frame.

The conventional Hausdorff distance between two point sets, model, $M=\{m_1,\cdots,m_{N_M}\}$ and Edge, $C=\{c_1,\cdots,c_{Nc}\}$ of sizes N_M and N_c, is defined as

$$H(M,C) = \max\ (h(M,C),\ h(C,M)) \tag{4}$$

where

$$h(M,C) = \max_{m\in M} \min_{c\in C} \|m-d\| \tag{5}$$

and

$$h(C,M) = \max_{c\in C} \min_{m\in M} \|c-m\| \tag{6}$$

For every model point m, the distance to the nearest convex hull's information c is computed, then the maximum value is assigned to $h(M,C)$. Also, for every convex hull's information c, the maximum distance to the nearest model point m is calculated, then it is assigned to $h(C,M)$. Hausdorff distance is the larger of the two maxima. That is, $h(M,C)=d$ means that every point of the model is within the distance d from some convex hull's information. But There are some problems in equation (6) and (7). If the model or object is outlying as can be seen in Figure. 1., the Hausdorff distance will be very large, even if most of the points are matched well.

Thus, we use a robust Hausdorff distance measure to reduce the problems. It is described as

$$h_R(M,C) = \frac{1}{N_M}\sum_{m\in M}\rho(d_C(m)) \qquad (6)$$

where $d_C(m)$ represents the minimum distance at point m to the convex hull information C, and the cost function ρ is defined by

$$\rho(x) = \begin{cases} |x|, & |x|\le\tau \\ \tau, & |x|>\tau \end{cases} \qquad (7)$$

where τ is a threshold to eliminate outliers. It is useful to reduce the effect of outliers and to compare portion of images. In this paper, we implement above robust Hausdorff distance to track a face in video sequences. The algorithm is as follows.

It is assumed that the Hausdorff distance is smaller than a threshold T. A small shape change in all directions, the displacement $t\in T$ of a face model is obtained by minimizing the following formula :

$$\sigma = \min_{t\in T} H_t(M,E) \qquad (8)$$

That is, $h_R(M+t,\ C)$ and $h_R(C-t,\ M)$ are computed, then the larger of the two is selected as $H_t(M,C)$. When the Hausdorff distance, $H_t(M,C)$ is smallest, t is selected as the best possible displacement.

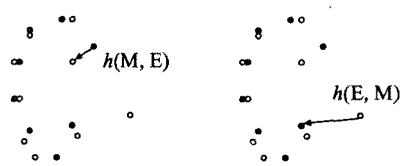

Fig. 1 Hausdorff distance. (a) $h(M,C)$, (b) $h(C,M)$, In this figure, Hsusdorff distance is equal to $h(C,M)$.

3.2 Initerlization of Face Model

First of all, the face position has to be determined. The initial model can be extracted by simple computation of the difference between consecutive frames, but it is too noisy. In section 2, we presented the face segmentation algorithm in the first input image frame using skin-color information and convex hulls information. Then, it is set to the initial model for robust hausdorff distance measure. Using the model and convex hull information of current frame, we calculate the best possible displacement.

3.3 Update Model

Since the tracked object can rotate or change its shape, it is necessary to update the corresponding model every frame. We only invest the skin color region to the new model of frame for finding the best possible displacement t and make the convex hull surrounding the skin color region. Hence, to update these parts we shift the face model of the first frame to the new position of best match, which was located by the robust hausdorff distance. After the best possible displacement t is obtained, the face model is updated. Here, we determine which part of the convex hulls information set must be included in the updated model. The model of the previous frame is shifted to the new position with the displacement t. With t, the part of convex hulls information set that belongs to the updated model is selected by the following equation.

$$M_n = \{c\in C_n \mid h_R(C_n - t, M_{n-1}) \le \sigma\} \qquad (6)$$

The selection of σ is very critical. If σ is too small, we have little tolerance to the shape change. If σ is too large, some nearby noises are determined as new model. In our experiment, σ of 3 pixels is used.

4. Experimental result

We have have applied the proposed tracking algorithm to track a face in image sequences containing a face, which is obtained by CCD camera. The obtained images are 320 x 240 RGB model images of laboratory environment. The host computer administering the hole system uses IBM PC Pentium III 700 MHz.

First, we transform the RGB model images into YCbCr. We segment the candidate region of face from the background using skin color map and hair color map. A face model is obtained by intersection region between convex hull of skin color region and convex hull of hair color region. Finally, the displacement is computed by a robust hausdorff distance and the model is updated. Figure.2 and Figure.3 show the result of skin color region and hair color region in the first frame and the result of initial facial model, respectively. Figure.4 represents. the tracking results.

In this paper, After segmentation the face region in the first frame, the faces of the next frames is tracked by the proposed algorithm used the local information of face, so the algorithm appears the property of robust tracking system in case of existing the another person in the next frame. Figure.5 shows the tracking result of cases including the occlusion among the frames.

Fig.2 The skin color region and hair color region in the first frmae

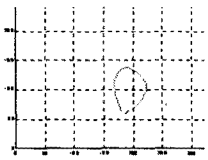

Fig.3 initial face model

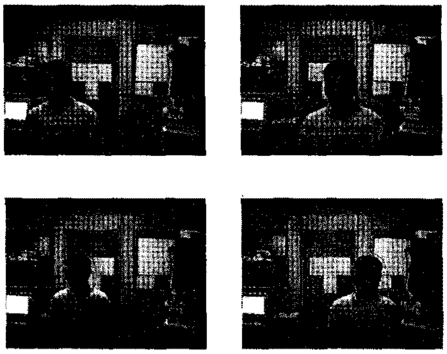

Fig.4 Tracking images

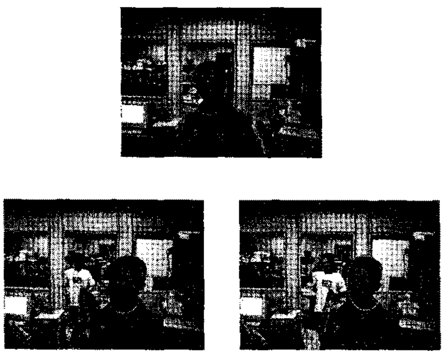

Fig.5 Result including the occlusion among frames

5. Conclusion

In this framework, we have presented an automatic approach for face segmentation and tracking over time. Face segmentation is done by skin color map of YCbCr color model and hair color map of RGB color model in the first frame.

Once faces are segmented, we extract a face model, which can be tracked over time. Tracking the face is performed by computing displacement using a robust hausdorff distance. If a best possible displacement is found, the face model is updated.

However, the proposed method sometimes fails to tracking the real face. Reasons under concern include the following:

1) Hairstyle : Faces with special hair styles, such as skinhead, or wearing a hat, may fail to be tracked.
2) If person wear a clothe of skin color, the clothe may show skin color.
3) If two of more faces are too close, the skin parts of them may be merged together.

Our method may also give some false position under some condition. The most important reason is that we only use the convex hulls for making the face model and ignore all the details about facial features during making the face model. Checking if there are facial features in these face candidates can help deleting all false faces.

Reference

[1] Douglas Chi and King N. Ngan, Face Segmentaion Using Skin-Color map in Videophone Applications , IEEE Trans. PAMI, vol. 9, No.4, JUNE 1999, pp.551-564.

[3] Dong-Gyu Sim, Oh-Kyu Kwon, and Rae-Hong Park, Object Matching Algorithms Using Robust Hausdorff Distance Measures , IEEE Trans. Image Processing, vol. 8, No. 3, March 1999, pp. 425-429.

[3] Jorgen Ahlberg, Using the Active Appearance Algorithm for Face and Facial Feature Tracking , IEEE, 2001, No. 68-72.

[4] Aggarwal, N. Karl, W. C. Line detection in images through regularized Hough transform. Image Processing, 2000. Proceedings. 2000 International Conference, vol.3, 2000 No. 873 -876

Acknowledgment

This work was supported by brain Korea 21 project.