신뢰성 있는 얼굴 검출

전봉진[O] 김대진
포항공과대학교 컴퓨터공학과 지능형미디어 연구실
{simple21[O], dkim}@postech.ac.kr

Robust Face Detection

Bong-Jin Jun[O] Daijin Kim
IM Lab, Dept. of Computer Science & Engineering, POSTECH

Face detection is an essential preprocessing step for face recognition[1], surveillance, robot vision interface, and facial expression recognition. It also has many application areas such as picture indexing, tracking, clustering and so on. However face detection has its intrinsic difficulties for the following reasons. First, face is not a rigid object, i.e. every person has different facial shape and different form/location of facial features such as eyes, nose, and mouth. Second, face of the same person looks differently as the facial expression, facial pose, and illumination condition changes. Finally, it is almost impossible to train infinite number of non-face patterns, consequently unexpected false acceptance or false rejection could be occurred.

In this paper, we present a robust real-time face detection algorithm. We improved the conventional face detection algorithms for three different steps. For preprocessing step, we revise the modified census transform to compensate the sensitivity to the change of pixel values. For face detection step, we propose difference of pyramid(DoP) images for fast face detection. Finally, for postprocessing step, we propose face certainty map(FCM) which contains facial information such as facial size, location, rotation, and confidence value to reduce FAR(False Acceptance Rate) with constant detection performance.

Since the main contribution of the paper is the FCM, in this abstract, we concentrate on explaining how to adopt FCM into the face detection algorithm. The detailed explanation of the procedure is given below.

1. For each scanning window centered at (x, y), we compute the confidence value.

$$H_i(T) = \sum_{p \in S_i} h_p(T(p)),$$

where i represents the i-th cascade, p represents the p-th feature location, and S_i is the set of feature locations, respectively.

2. The confidence value cumulated for all n cascade is like following:

$$S(x,y) = \sum_{i=1}^{n} H_i(T), \qquad \text{if } H_i(T) \text{ for all i is above threshold, otherwise 0.}$$

We compute the cumulated confidence value for every pixel position in the image.

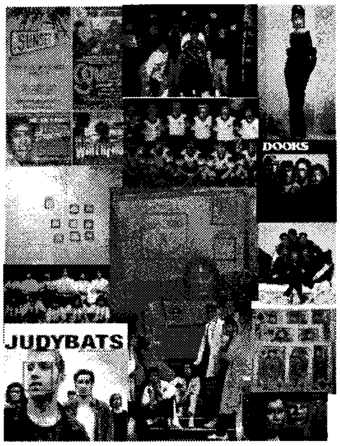
3. We also compute above equation for all pyramid images, then we have $S_p(x,y), p = 1, \cdots, m$, where m is the total number of constructed pyramid images and the pixel locations (x,y) of each down-scaled pyramid image are translated to its corresponding original image locations.

4. The FCM for location (x,y) consists of four items such as $S_{\max}(x,y)$, $W_{\max}(x,y)$, $H_{\max}(x,y)$,

and $C(x,y)$. $S_{\max}(x,y)$ is the maximum confidence value among $S_p(x,y), p=1,\cdots,m$, $W_{\max}(x,y)$ and $H_{\max}(x,y)$ is the width and height of the detected face window which has the maximum confidence value, and $C(x,y)$ is the confidence value cumulated for all m pyramid images $C(x,y)=\sum_{p=1}^{m} S_p(x,y)$.

5. Since we constructed FCM, we can determine the face region using it. First, we look for the values above threshold in $S_{\max}(x,y)$. Then we determine the location (x,y) as the center of face when $C(x,y)$ is above threshold. The non-face region where the maximum confidence value is above threshold is not classified as a face region, since $C(x,y)$ is lower than the threshold. Consequently, we can reduce the FAR using our proposed FCM.

For constructing training face dat, we gathered 17,000 face images from internet. Gathered face images contain multiple human species, variety of illumination conditions, and variation. Each image is aligned by eye location, and we resize images to 22x22 base resolution. In addition, for the robustness to image rotation, we generated another 25,060 face images by rotating gathered face images to −3, 0, 3 degrees. For non-face training data, we collected 5,000 images which include no face image from internet. Then, we extracted image patches from collected internet images by random size and position. After that, we generated 60,000 non-face images by resizing extracted image patches to the same scale of training face images. We used these 60,000 non-face images as the training non-face data for the first stage of cascade. For the next stages of cascade, we used non-face data which are considered as face image by the previous cascade(i.e. we used false positives of previous cascade as training non-face data for training current cascade).



We tested our algorithm on CMU+MIT frontal face test set. Table 1 represents the results of face detection. When we used FCM, the reduction of FAR is ten times better than the cascade adaboost detector with the same detection rate, while the detection time is almost the same. The cascade adaboot detector needs computations for grouping and eliminating overlapped face candidate region. In contrast, proposed detector does not need these computations but needs the computation for FCM. Operating on 320 by 240 pixel images, faces are detected at 23 frames per second on a conventional 3.2 GHz Intel Pentium IV system and 6 frames per second on OMAP5912(ARM9 system).

Table 1. Results of Face Detection

Detector	Number of False Detection
RMCT, adaboost and FCM	3
RMCT and adaboost	93
Viola-Jones	78
Rowley-Baluja-Kanade	167
Bernhard Froba	27