COMBINING FEATURES AND DECISIONS FOR FACE DETECTION

Jie Wang, K.N.Plataniotis, A.N.Venetsanopoulos

Multimedia Laboratory
The Edward S.Rogers Sr.Department of Electrical and Computer Engineering
University of Toronto, Toronto,M5S 3G4, ONTARIO, CANADA
jwang, kostas, anv@dsp.utoronto.ca

ABSTRACT

In this paper, we propose a novel face detection algorithm which detects faces in color images using a combination of feature and decision fusion mechanisms. In addition to commonly used skin color information, two additional features, namely average face template matching score and horizontal edge template matching score are utilized. A mean shift algorithm operating on the combined feature space is used to determine face candidate areas. Face candidate and its flipped pattern are then inputed to a multiple layer perceptron based classifier. Two outputs along with the correlation value between candidate and its flipped pattern are then combined to give the final decision. Experimentation on two different test databases indicates that the proposed method performs well under a variety of scale, expression and environmental conditions, outperforming commonly used approaches.

1. INTRODUCTION

Face detection is a fundamental step in many applications such as face recognition, video surveillance and human computer interface. The performance of face detection process, therefore, critically affects the overall systems characteristics.

Face detection is a challenging problem since face patterns vary significantly under different illuminations, poses, expressions and scales. In literature, various face detection algorithms have been proposed and recent surveys can be found in[1][2]. Face detection methods can be roughly classified into two categories, feature-based approach and image-based approach[2]. In feature-based approach, low level features such as texture, color or motion are extracted and encoded according to human knowledge. A face pattern is detected if the encoded features can match a defined model or satisfy a set of rules. Image-based approaches treat the face detection problem as a two class (face and non-face) pattern recognition problem. A classifier is built on the extracted feature space to classify the input pattern is a face or not. Therefore methods used for recognition such as principle component analysis (PCA) and linear discriminant analysis (LDA) can be utilized.

Most of the face detection system are trying to find the feature or classifier which can give the best performance. However, it is found that lost information provided by those not best features or classifiers may be complementary and useful[3]. This motivated the introduction of multiple learner system. Other than choose a best learner, set of learners are combined to get a better performance. In this paper, we propose a novel framework to detect faces in color images by combining both features and decisions. In addition to skin color which is widely used in face detection in color images, we proposed another two features, average face template matching score and horizontal edge template matching score, to compensate the unreliability of skin color model since the appearance of skin color is often affected by foreground and background lightings[1]. Thus three features which provide the complementary information from the viewpoints of color, edge and texture are combined by a product rule and drive mean shift algorithm to detect face candidates. In order to verify the face candidates, a multilayer perceptron classifier is built following Sung's[4] view-based framework. Based on the knowledge that a real face is usually symmetric, both the face candidate and its flipped pattern are inputted to the classifier. Final decision is obtained by combining the two outputs and the correlation value between candidate and its flipped pattern through a sum rule, product rule and stacked generalization.

The rest of the paper is organized as follows: Section 2 introduces the proposed framework. In section 3, face candidate localization module is introduced. The verification module utilized in this work is described in section 4. In section 5, simulation results are included in order to demonstrate the effectiveness of the proposed method. Conclusions are drawn in Section 6.

2. SYSTEM OVERVIEW

The proposed here algorithm can be divided into two parts, face candidate localization and verification (Fig.1). The input image is transformed to HSV color space, in which H,S components are used for skin color filtering and V component is used for template matching. Skin color information, as well as average face matching score and horizontal edge matching score are combined to drive a non-parametric searching procedure, Mean Shift[5]. The modes which are found by mean shift algorithm are the locations of face candidates. Verification of face candidates is implemented based on Sung's view based approach[4]. Both the face candidate and its flipped pattern are verified. The outputs of the verifier together with the correlation value between the face candidate and its flipped pattern are combined to make the final decision.

3. FACE CANDIDATE LOCALIZATION

For most face detection system in color images, skin region detection is usually the first step and further processing is operated on skin regions.Thus the whole performance relies much on skin color model. However, even with the most complex model, skin color filtering can not always be reliable due to the environmental factors[1]. Therefore, we adopt other two features, horizontal

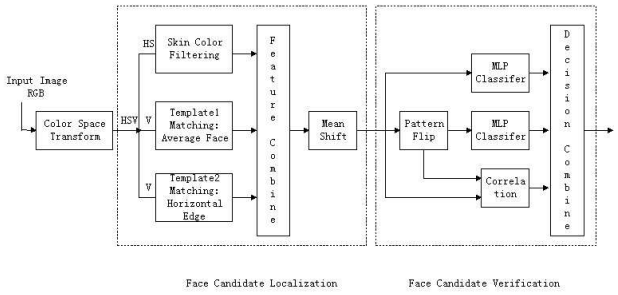

Fig. 1. System framework

edge template and average face template matching scores, combined with skin color in parallel to detect face candidates. Edge representation of a face provides the schematic information of face pattern. Horizontal edge template is motivated from the observation that faces in most images which are frontal or near frontal have plenty of horizontal edges in eyes and mouth region. In addition, influence of hair which usually has many vertical edges can be avoided. Horizontal edge template proposed here is a binary template which captures the information of eye and mouth structure. In addition to edge, which describes the structure of local facial features, average face template is utilized to provide global texture information of face pattern.

3.1. Features

Feature 1: Skin Color The skin color model utilized here is the HSV variant suggested in [6]. The RGB color values are transformed to HSV color space and a two dimensional ([h,s]) Gaussian model is obtained from the actual skin pixels. Thus skin likelihood of an input pixel can be calculated with the mean and covariances obtained from the training samples.

Since pixel level likelihood cannot indicate the pattern is a face pattern or not, we use the ratio of skin pixel number to the pattern size as the pattern level probability estimated by skin color while the skin pixels are those with the likelihood value over a threshold.

Feature 2: Average Face Matching Score Average face template (Fig.2 (a)) is generated from 299 gray scale face images in BioID database[7].

Pattern matching is processed between V component of the image and the template using normalized cross correlation. Directly calculate matching score is extremely time consuming since 2D cross correlation function needs to be computed for every point in the image. Therefore we applied the fast cross-correlation algorithm proposed by Lewis[8] which is to use Fourier Transform of the image and the template to perform convolution.

Feature 3 - Horizontal Edge Matching Score This template is also generated from 299 gray scale face images from BioID database[7]. The edge is detected for each image by using Sobel operator and averaged over all 299 images. The final template (Fig.2 (b)) is binary by setting a edge strength threshold on the averaged edge map followed by some morphological operations.

For the correlation between binary image $f(x, y)$ and template $t(x, y)$, we proposed a new measure which could be efficiently

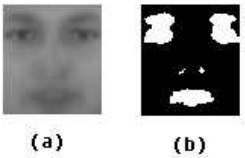

Fig. 2. (a)Average face template (b)Horizontal edge template

calculated using Fourier transform. The correlation is defined as:

$$c(u,v) = \frac{\sum\limits_{x,y} f(x,y)t(x-u,y-v)}{N_w} \cdot (1 - \frac{\sum\limits_{x,y} f(x,y)t'(x-u,y-v)}{N_b}) \quad (1)$$

,where $t'(x, y)$ is the negative template with changing 0 to 1 and 1 to 0, N_w is the number of white pixels in template foreground region and N_b is the number of black pixels in template background region. The intuitive physical meaning of the measure is that the better the image window pattern matches the template, more white pixels ($f(x, y) = 1$) are within the region of template foreground and fewer white pixels are within the region of template background. Therefore, the above measure can be efficiently computed by $c(u, v) = \frac{\mathcal{F}^{-1}\{\mathcal{F}(f)\mathcal{F}^*(t)\}}{N_w} \cdot (1 - \frac{\mathcal{F}^{-1}\{\mathcal{F}(f)\mathcal{F}^*(t')\}}{N_b})$

3.2. Feature Combination

Each of skin color and two templates will provide an pattern level probability map from the viewpoint of color, edge and texture, $P_c(x, y), P_e(x, y), P_t(x, y)$. These three maps are combined by using a product rule $P(x, y) = P_c(x, y) \times P_e(x, y) \times P_t(x, y)$. With a product rule, only three features all report a high value, the pattern may be determined as a face candidate. This helps to reduce a lot of background regions which may be detected as face by one of the features.

In order to capture the faces of different size, all above processing is done across different scales(Fig.3). Based on trained template size (73×65), 9 different scales have been tried, which can detect the faces from size (29×26) to (146×130).

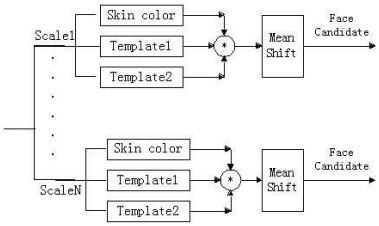

Fig. 3. Framework of combining skin color and other two templates matching

3.3. Detect Face Candidate

The output of the above module at each scale i is a pattern level probability distribution map $P(x, y)$ whose pixel value indicates how possible the window pattern of size i centered at that pixel (x, y) is like a face. Then face candidate centers which are corresponding to the mode of the conditional distribution $f(x|face)$,

$f(y|face)$ are detected by a well-defined algorithm – Mean Shift: $m(x) - x = \frac{\sum_{i=1}^{n} x_i g(||\frac{x-x_i}{h}||^2) w(x_i)}{\sum_{i=1}^{n} g(||\frac{x-x_i}{h}||^2) w(x_i)} - x$, where $w(x)$ is the weight and in our case is $P(x, y)$, $g(x) = -k'(x)$ and $k(x)$ is the profile of Kernel function used for estimating density function[5].

In this work, we adopt Continuously Adaptive Mean Shift (CAMSHIFT) algorithm proposed by Bradski [9] on the combined pattern level probability distribution map $P(x, y)$. The algorithm is operated with different initial locations. Other than exhaustively try every location in the image, we start with the points whose pattern level probability value is greater than a threshold.

4. VERIFICATION

Since the combination of three features will still include some non face patterns and mean shift algorithm, as many iterative searching methods, will stick into the local maximums, verification is necessary for further reducing false detections. Here we use Sung and Poggio's view-based method[4]. In addition to verify the face candidate, its flipped pattern is also verified. The flipped pattern is the horizontally reversal version of the original image, i.e., $F(x, y) = I(-x, y)$ where $I(x, y)$ is the original image , $F(x, y)$ is the flipped image and the origin point is in the middle of the image. Since the face pattern is usually symmetric, the flipped pattern of a real face can also be classified as a face pattern with high probability. At the same time, based on the observation that a face pattern is very similar to its flipped pattern, their correlation value can be viewed as a third estimator which is combined with the outputs of Sung's approach to give a more robust decision.

In [4], Sung and Poggio developed an image-based approach to detect faces. In their system, large number of face and non face samples are modelled as six Gaussians clusters respectively by using a modified K-means clustering algorithm. Feature vector is composed of the Mahalanobis distances to each of the cluster in a lower eigenspace obtained by a standard Principle Component Analysis. Then a multilayer perceptron classifier is built on the feature space to classify an input pattern is a face or not. In our work, we adopt Sung's scheme and built a MLP based classifier to verify the face candidate and its flipped pattern. In addition to the above classifier outputs, cross correlation value between the face candidate and its flipped pattern is calculated as a third estimator based on the knowledge that human face is symmetric.

Three scores which are classifier outputs with face candidate input S_MLPorg, with flipped pattern input $S\_MLPflip$ and the correlation value S_corr are then combined to give the final decision. Here, we explored three combination scheme, sum rule, product rule and stacked generalization.

Sum rule and product rule are very simple combination forms just taking the weighted sum and product from these three scores, i.e., $S\_sum = \alpha_1 S\_MLPorg + \alpha_2 S\_MLPflip + \alpha_3 S\_corr$, $S\_product = S\_MLPorg^{\beta_1} \times S\_MLPflip^{\beta_2} \times S\_corr^{\beta_3}$. However, before the combination, these three scores should be translated into a common domain to represent same concept with same scale. The output value of MLP classifier ranges from -1 to 1 with the positive value indicating a face pattern and the negative value indicating a non-face pattern. The higher the value, the pattern more likely to be a face. Therefore, S_MLPorg and $S\_MLPflip$ can be viewed as the posterior probability after normalizing it to [0-1]. However, the correlation value which is also ranges from -1 to 1 can not be viewed as the posterior. In order to estimate posterior distribution of S_corr, we calculate the corresponding correlation value of 2000 face pattern samples and 2000 non-face pattern samples. The real correlation value is quantified into 20 bins each of which has a interval of 0.1. The the obtained histograms of face patterns and non-face patterns are viewed as the conditional probability distribution of correlation score given the face pattern and non-face pattern $P(bin(i)|\omega_k), i = 1, ..., 20, k = 1, 2, \omega_1 : face, \omega_2 : non - face$. Therefore, posterior probability of the given correlation value to be a face can be obtained by Bayesian rule:

$$P(\omega_1|bin(i)) = \frac{P(bin(i)|\omega_1)P(\omega_1)}{P(bin(i)|\omega_1)P(\omega_1) + P(bin(i)|\omega_2)P(\omega_2)} \quad (2)$$

Assume that the priori probability $P(\omega_1) = P(\omega_2) = 0.5$, equation 2 changes to $P(\omega_1|bin(i)) = \frac{P(bin(i)|\omega_1)}{P(bin(i)|\omega_1)+P(bin(i)|\omega_2)}$

Since correlation value is a simple but weaker feature compared to MLP classifier output, we set a smaller weight for correlation value though combination. By experiments, we empirically choose the weights of two MLP outputs and correlation value as 0.45,0.45,0.1 respectively for both sum rule and product rule.

For stacked generalization, these three scores comprise a new feature space on which a second level classifier is built to make the final decision. In this work, we use K-nearest neighbor as a second level classifier with K=100 on the train set of 2000 face patterns and 2000 of non-face patterns .

5. EXPERIMENT AND RESULTS

In order to demonstrate the effectiveness of the proposed framework, more than one thousand images from different sources have been tested. The test databases used by Hsu in [10] is also adopted for a comparison.

5.1. Training Database

Two templates used in this work are created from BioID database[7]. This is a database consisting of 1521 gray level images, each of which contains a frontal view face. In addition, the locations of two eyes of each image are also provided. With this information, we rescaled and cropped the images by setting the distance of two eyes to 40 pixels and resulting in face patterns of size 73×65. Altogether 299 images are selected arbitrary across all 23 persons for generating templates. Average template is obtained by taking the average pixel value over all 299 face patterns. Horizontal edge template is obtained by taking the average of the horizontal edge map of each face pattern and followed by a binarization step.

The database used to train verification classifiers contains 6599 gray scale face patterns and 6617 nonface patterns. Some images are from MIT's face training database while others are manually collected from the Internet. The pattern size is 19×19. In all these samples, 4599 face patterns and 4617 non-face patterns are used to train the MLP classifier while other 2000 face patterns and 2000 nonface patterns are for second level K-Nearest Neighbor classifier and estimating likelihood distribution of correlation value.

5.2. Test Databases and Results

For face detection in color images, there is no standard database for testing. The available databases which are commonly used by face detection and recognition community like FERET database, CMU database, MIT database are gray scale image databases. Thus two

	Champion			News Photo		
	DR	FDR	IMP	FDR	DR	IMP
MLP	96.48	17.43	-	84.92	10.98	-
Sum	96.4	15.58	1.77	85.24	10.32	0.98
Product	96.32	12.86	4.41	84.42	9.67	0.81
K-NN	95.6	7.43	9.12	82.13	4.59	3.6

Table 1. Detection Result

test databases to be used for color image detection are selected by ourselves. The first one is the Champion database. This freely available database (http:// www.libfind.unl.edu / alumni / events/ breakfast_for_champions.htm) contains 1251 compressed images with 1251 faces. The faces in these images are frontal and near frontal. This database is also used by Hsu[10], however, only 227 images of the database were tested. The detection rate and the number of false detections obtained in[10] are 91.63 percent and 14 respectively. The second database is a "news photo database" of 386 images with 610 faces. These images are collected from yahoo news web site.

The detection rate (DR) and false detection rate (FDR) of only MLP classifier used with face candidate input (1st level) and combination with other two scores are shown in Table 1. The false detection rate is defined as the ratio of number of falsely detected patterns to number of available face patterns. In addition the improvement (IMP) with respect to the decease of total error rate (miss detection rate + false detection rate) by using decision combination scheme is also shown in Table 1.

In order to reduce the false detection without decreasing the detection rate, the patterns which are decided by MLP classifier as a face pattern with high confidence, i.e. with high S_MLPorg will be determined as face directly without further classification by K-Nearest Neighbor classifier or rule based approach.

Champion database is comparatively simple with faces of fontal and near frontal and we got a detection rate over 95% even without any of the combination. However, the false detection rate is quite high with 17.43%. All three combination schemes reduce the false detection rate without a big loss on detection rate. News photo database is much more challenging, since the faces contained under the complex background with all kinds of poses and expressions. From the experiments, we find stacked generalization with K-Nearest Neighbor classifier has the greatest performance improvement with respect to decreasing the total error. Sum and product rule can achieve a higher detection rate, however, it still have a high false alarm. Some detection results with K-NN combination scheme are shown in Fig.4.

6. CONCLUSION

In this paper, we have proposed a face detection algorithm in color images. The algorithm combine both features and decisions to provide a robust performance. The color image is firstly transformed to HSV color space, where HS components are used to detect skin-tone pixels by using a Gaussian skin color model. Three features, skin color, average face template matching score and horizontal edge template matching score are combined to drive CAMSHIFT algorithm to detect face candidates. Finally, face candidates and their flipped patterns are verified by multiple perceptron based classifier, and the decisions with the correlation value between face candidate and its flipped pattern which is viewed as the third estimator are combined using three combination schemes, sum rule, product rule and stacked generalization. Extensive experiments on different images show that the system performs well under various conditions, and shall be therefore preferred to other systems when it comes to using face detection in practical applications.

Fig. 4. Detection Results Using K-NN: New Photo Database(1st row),Champion Database (2nd row)

7. REFERENCES

[1] M.Yang, D.J.Kriegman, and N.Ahuja, "Detecting Faces in Images: A Survey ", *IEEE Transactions on PAMI*, vol.24,no.1,pp.34–58, January 2002.

[2] Erik Hjelmas, "Face detection: A survey", *Computer Vision and Image Understanding 83*, pp. 236–274, 2001.

[3] J. Ghosh, "Multiclassifier systems: Back to the future", in *Proceedings of Third International Workshop on Multiple Classifier Systems*,2002,pp.1–15.

[4] K.-K. Sung and T. Poggio, "Example- based learning for view-based human face detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.20, pp.39–51,1998.

[5] Y.Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.17, pp.790–799,1995.

[6] S.Tsuruoka, A.Kinoshita, T.Wakabayashi, Y.Miyake, and M.Ishida, "Extraction of hand region and specification of finger tips from color image", in *Proc. of International Conference on Virtual Systems and Multimedia*,1997,pp.206–211.

[7] R. Frischholz. Jesorsky, K. Kirchberg, "Robust face detection using the hausdorff distance", in *In J. Bigun and F. Smeraldi, editors, Audio and Video based Person Authentication*, 2001,pp. 90–95.

[8] Lewis J.P, "Fast normalized cross-correlation," *Industrial Light and Magic*.

[9] G.R.Bradski, "Computer vision face tracking for use in perceptual user interface", *Intel TEchnology Journal*,1998.

[10] R.Hsu, M.A.Mottleb, and A.K.Jain, "Face Detection in Color Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696–706, May 2002.