Enhancement of Template-Based Face Detection by Belief Propagation in Ordered Component Search

Hossein Nejati[1], Elaheh Bayat[2], Amin J. Torabi[3], and Rasool Maghareh[1]

[1] School of Computing, National University of Singapore (NUS), Singapore 117417
{nejati,maghareh}@nus.edu.sg
[2] Faculty of Information Technology, Multimedia University (MMU), Malaysia
[3] School of EEE, Nanyang Technological University (NTU), Singapore

Abstract. Face detection problem is a high-dimensional problem. Here, we proposed an enhancement to face detection approaches, particularly, template-based approaches, in which we avoid the exhaustive search while we obtain face scale, rotation angle, and facial component locations. Our biologically inspired algorithm quickly deduces face position, scale, and orientation, as well as location of facial feature, by an ordered search and belief propagation. Our approach has several advantages over image-based and feature-based face detection methods. In addition, our approach can be used to detect both real faces and non-real faces including symbolic faces, cartoons, puppets, and caricatures. We used our approach to enhance Viola-Jones face detector in terms of running time and accuracy. We presented results of our tests on Multi-PIE dataset which show decrement of the running time from 145.11 seconds to 1.48 seconds and increment the detection accuracy from 67.4% to 93.8%.

Keywords: Face Detection, Template-based, Belief Propagation.

1 Introduction

Face detection is a prerequisite, and also a bottle neck for face recognition and face analysis applications. Aim of the face detection system is to identify and locate all faces regardless of their positions, scale, orientation, lighting conditions, expressions, etc. We can divide the face detection methods into two main categories: image-based methods and feature-based methods.

The image-based methods (e.g. [9], [12]) are usually consisted of two main parts: the core, and the sweep. The core is a statistically trained classifier(s) which classifies a fixed size window (of pixels) into face and non-face classes. The sweep part, analyses the core in a fixed size window over the entire image. To find all faces, the sweep should be performed using windows with all scales and rotations which is an exhaustive search with a high time order.

In feature-based approaches, one or more apparent properties of the face, such as skin color or face geometry are exploited to find face candidates. Depending on which features to use, these methods may or may not perform window sweeping across the image. For example, methods which use face skin color perform a color classification

Z. Qian et al. (Eds.): Recent Advances in CSIE 2011, LNEE 124, pp. 693–698.
springerlink.com

which reveals face candidate positions without window sweeping (e.g. [1]). In contrast, methods which use facial component or geometric features should perform window sweeping to find these features (e.g. [8], [2]). Although using features such as face skin color has the advantage of avoiding the exhaustive search, these approaches have some disadvantages such as ignoring gray scale images, not providing rotation, pose, or facial feature location information, and sensitivity to skin color differences and illumination variations.

Salient point based approaches (e.g. [10], [5], [13]) also avoid exhaustive search. In these approaches, salient points are detected and based on color, texture, or relative position of these points, face candidates are located. However, these works are restricted to large scale, frontal views of faces, with relatively good illumination conditions. This is mainly because the salient points are not discriminative to faces and a large number of non-face related salient points are detected in the image, even after the points are verified (e.g. with gist [10]).

In contrast with face detection algorithms, it seems that humans can detect faces with different sizes and rotations, significantly fast with facial features pin pointing [11]. This is done by avoiding exhaustive search and applying a belief propagation technique on more important facial components, to detect the face and other facial components. It means that after detecting a shape, the probability of that particular shape to be a facial component is decided. Based on this probability, location, scale, and rotation angle of other facial components are predicted.

Inspired by how the brain works, we introduce a computational low cost approach to detect faces with different rotation angles, scales, and poses in color and gray scale images. Our approach is a combination of image-based and feature-based approaches that first performs an "ordered search" with belief propagation to find face candidates without exhaustive search and then verifies these candidates with a template-based algorithm (which can be almost any template matching classifier).

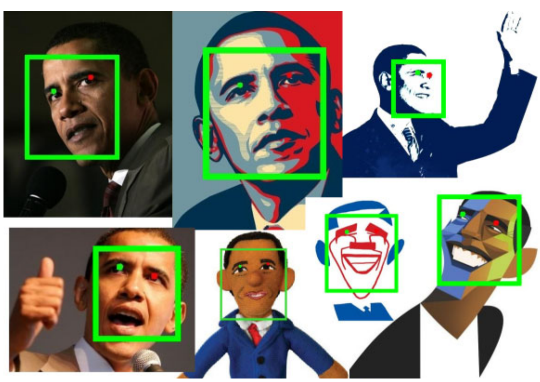

Fig. 1. Examples of the wide range of faces, detected by our approach

In order to ensure a fast ordered search, we use a method similar to the human visual system strategy. We first detect all components in the image and search for a component representing the right eye. Based on this component, we estimate the location of search for the left eye. Then based on the eyes we estimate face scale and rotation, and the estimated location of the mouth and the nose. Estimations of locations, scale and orientation of facial components are based on several works on geometric properties of the face [4], [1].

In our approach we avoid exhaustive search, only require training for a set of fixed size, up-right faces (advantages over image-based approaches), handle a wide range of skin colors, illumination, scale, and to some extent poses (advantages over feature-based approaches). In addition, due to ordered search, our method can be applied to detect non-realistic faces such as puppets, line-drawings, signs and symbols, and cartoons (see Fig. 1).

2 The Face Detection Algorithm

Our algorithm consists of 3 steps: 1) pre-processing, 2) face candidate detection, and 3) verification. In the first step, image normalization and segmentation is performed. Important facial components (eyes, mouth and nose) are relatively darker than the skin, regardless of the illumination condition [9]. Therefore, gradual image thresholding based on pixel intensities can provide our required segmentation. We iteratively threshold the input image into a black and white image, with the theresold level set from 1 to 255, and select connected components as segments. Because different image segments have different intensity levels, we need to store and label all segments across iterations. We also apply Contrast-Limited Adaptive Histogram Equalization (CLAHE) [14] in order to cope with illumination variations. CLAHE improves results in poorly lit images (e.g. side lighting).

In the second step, face candidate detection, we apply two sets of rules to detect image segments that are likely to represent a face in the image. The first set of rules includes the three Golden Ratio properties of the face (see section 2), deducted from the specific geometric properties of the face. The second set of rules includes brightness differences between face segments, adopted from intensity properties of the face (e.g. brightness of cheek bones and fore-head in comparison to eyes and mouth). The first set of rules finds candidate segments which have face-like geometric order. The second set of rules rejects candidate segments which do not have the intensity properties of a face.

We use a specific order of searching for facial components to avoid the exhaustive search. We first search for the right eye, then the left eye, then the mouth, and finally the nose. In each step of the search, components are checked against both sets of rules. Although this may not be the best possible order for searching, this order is confirmed by biological studies on the importance level of the different facial component for the human visual system [12].

Searching in an order, unlike searching for all facial components at once, enables a step by step applying rules and deducing information about location, scale, and rotation of the facial components. For example, location of the left eye is deduced by the orientation and size of the right eye; and scale and orientation of the face is deduced by length and orientation of the line connecting the eyes. This is similar to the belief propagation strategy used by the human visual system to detect objects, and in particular, faces [7]. Scale and orientation of each component is calculated based on their respective segmented region in the Pre-processing section.

We interpret both sets of rules into a set of binary probability functions for the facial components. To introduce ordered search in these functions, we relate each function to the previous one, so that one cannot be calculated unless its previous function is calculated. This connection between probability functions leads to a fast rejection whenever a violation of rules occurs. Equations 1 to 3 illustrate these functions for the right and left eyes, the mouth, and the nose:

$$P(s_r = E_r) = P(A(s_r) \in A_{limit}). \tag{1}$$

$$P(s_l = E_l \mid s_r = E_r) = P(loc(s_l) \in E_{l\_est}(s_r)) \times P(A(s_l) = \sigma \times A(s_r) \pm \varepsilon) \tag{2}$$

$$P(s_r, s_l, s_m, s_n \in face\_candid) = P(s_r = E_r) \times P(s_l = E_l \mid s_r = E_r) \times P(s_m = M \mid s_r = E_r, s_l = E_l) \times P(s_n = N \mid s_r = E_r, s_l = E_l, s_m = M) \tag{3}$$

where s_i is an arbitrary image segment; E_r, and E_l are segments for the right eye, the left eye, respectively; *P()* is a binary probability function; function *A(s)* returns area of segment *s*; A_{limit} is the maximum acceptable segment area based on image dimensions; function *loc(s)* returns Cartesian coordinates for the centroid of segment *s*; function $E_{l\_est}()$ returns estimated locations of the left eye; and finally, *face_candid* is the face candidates set. Note that we estimate the location of a facial component based on the geometric properties of the previously found components in the ordered search. The *a priori* probabilities and regulation constants are learned by simple feed-forward back propagation neural network, trained on 249 images from the Multi-Pie dataset (high resolution images).

The geometric properties of the face are interpreted into location estimation functions (e.g. E_{l_est}) and the intensity properties of the face are interpreted into area constraints. Enforcing the intensity properties of the face significantly reduces false face candidates. The ordered search for segments dramatically reduces the search space for face candidate segments. As the ordered search proceeds from one step to the next, the search space is even more reduced as more constraints are required to be satisfied. Finally, a set of four image segments, are accepted as a face candidate if and only if each segment is accepted based on its respective properties, and its relation to the previous segments. When a face candidate is detected, that image portion is cropped, scaled to a fixed size, and rotated to the up-right angle for the next step, face verification.

The third step of our approach is the verification of face candidates by fixed scale face template matching. We use the well-known Viola-Jones face detector [12] to perform the verification. Note that, as the scale and rotation angle of the face is deduced in the ordered search, we only need to train the Viola-Jones face detector on upright faces, with a fixed scale and frontal pose. Each face candidate is cropped, scaled to that specific fixed scale, and rotated to the upright angle. Therefore, the Viola-Jones face detector is reduced to simple fixed size template matcher.

It should be noted that although we use Viola-Jones face detection for final verification, almost any similar approach can be used for this purpose. This is because at this step, face location, scale and rotation angle are already known. Therefore, given any fixed size face detection algorithm, face candidates with various scales and rotation angles can be transformed and tested without window sweeping.

3 Experimental Results and Discussion

We implemented and tested our approach using Matlab on a PC with Windows XP operating system running on a 2.33GHz Core(TM)2 Duo Intel(R) chip and 4GB of main memory. We used the Matlab implementation of the Viola-Jones face detector. We used 128 gray-scale images in 18×27 pixels to train the Viola-Jones face detector. These images include 69 up-right faces and 59 non-faces. In the training time, the classifier is trained to obtain error rate equal to 0.001, using same Haar-like features used in [12]. The testing images were selected from the 250 subjects of the Multi-PIE face dataset [6]. We used color images from each subject in 320×240 pixels under 3 different rotation angles (-45, 0, and 45), 2 different poses (frontal, and 45 left angles), and 3 different lighting conditions (low, bright, and left lighting). Finding face candidates in each image took 0.17 second on average. Face verification in each image took 0.79 second on average (including verification of several face candidates). The entire testing of each image took 1.48 sec. on average.

We tested the same set of test images, except for rotated images, on the same Matlab implementation of the Viola-Jones face detector, used in our algorithm. In order to enable this implementation of the Viola-Jones detector to find different scales of faces, we resize the original input image to 1.0, 0.8, 0.6, 0.4, and 0.2 of the original size (320×240 pixels). All faces detected with confidence level more than 0.8 are accepted. Average time required for processing each image was 145.11 seconds (including all sizes). Average times required for resized images with ratios 1.0, 0.8, 0.6, 0.4, and 0.2 were 77.77, 41.4, 17.46, 7.05, and 1.43 seconds respectively. Image resizing times are excluded from the average times. For both algorithms we recorded a hit, when at least one correct detection was made. Detection rates of both algorithms are compared in Table 1. Note that our algorithm uses the same Matlab implementation. Nonetheless, the running time and the accuracy are significantly improved by our face candidate detection technique.

We also tested our algorithm on non-realistic face images such as symbolic faces, cartoons, and caricatures which showed that our approach can be applied both realistic and non-realistic images (similar to the human visual system).

Table 1. Comparison results face detection for 250 subjects, 3 lightings, 2 poses, and 3 rotation angles, in percentage

	Lighting			Rotation			Pose		Avg. Accuracy	Avg. Time
	Low	Bright	Left	0°	45°	-45°	Frontal	45° Left		
Our Algorithm	**93.6**	**94.8**	**93.6**	92	84	72.8	**93.6**	**48.8**	**84.2**	**1.48**
Viola-Jones	84	92	46	NA	NA	NA	84	31.2	67.4	145.11

4 Conclusion

Many face Many face detection algorithms tend to exhaustively search for faces with all possible positions, scales, orientations, and poses in an image. This method of searching requires a substantial amount of time and calculation. In contrast, the human visual system responses significantly fast to the existence of faces and is able to instantly locate facial components [11]. In this paper, we presented a novel method to enhance existing face detection algorithms by avoiding exhaustive searching while locates facial components. We use an ordered search and belief propagation by which face position, scale, and rotation angle as well as facial component positions are deduced. Our approach requires training only for a set of fixed size, upright faces and it can cope with a wide range of skin colors and is robust to illumination variations. Our results on testing both realistic and non-realistic faces show that our method can dramatically reduce the running time and increase the detection rate of many proposed face detection algorithms such as the Viola-Jones face detector.

References

1. Chan, Y., Abu-Bakar, S.: Face detection system based on feature-based chrominance colour information. In: CGIV 2004, pp. 153–158 (2004)
2. Chen, W., Sun, T., Yang, X., Wang, L.: Face detection based on half face-template. In: ICEMI 2009, pp. 4–54 (2009)
3. Choraś, M., Andrysiak, T.: Symmetry-Based Salient Points Detection in Face Images. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Żurada, J.M. (eds.) ICAISC 2006. LNCS (LNAI), vol. 4029, pp. 758–767. Springer, Heidelberg (2006)
4. Frakas, L.G., Munro, I.R.: Anthropometric Facial Proportions in Medicine. In: Thomas, C.C. (ed.) , Springfield, IL (1987)
5. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: illumination cone models for face recognition under variable lighting and pose. TPAMI 23, 643–660 (2001)
6. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. In: FG, pp. 1–8 (2008)
7. Hawkins, J., Blakeslee, S.: On Intelligence. Times Books, New York (2004)
8. Li, F., Wechsler, H., Tistarelli, M.: Robust fusion using boosting and transduction for component-based face recognition. In: ICARCV 2008, pp. 434–439 (2008)
9. Sadr, J., Mukherjee, S., Thoresz, K., Sinha, P.: The fidelity of local ordinal encoding. In: Advances in Neural Information Processing Systems 14, MIT Press, Cambridge (2002)
10. Siagian, C., Itti, L.: Biologically-inspired face detection: Non-bruteforce-search approach. In: IEEE Int. Workshop on Face Processing In Video, pp. 62–69 (2004)
11. Sinha, P., Balas, B., Ostrovsky, Y., Russell, R.: Face recognition by humans: Nineteen results all computer vision researchers should know about. Proc. of IEEE 94, 1948–1962 (2006)
12. Viola, P., Jones, M.: Robust real-time face detection. IJCV 57, 137–154 (2004)
13. Won, W.-J., Jang, Y.-M., Ban, S.-W., Lee, M.: Biologically Motivated Face Selective Attention Model. In: Ishikawa, M., Doya, K., Miyamoto, H., Yamakawa, T. (eds.) ICONIP 2007, Part I. LNCS, vol. 4984, pp. 953–962. Springer, Heidelberg (2008)
14. Zuiderveld, K.: Contrast limited adaptive histogram equalization. In: Graphics Gems IV, pp. 474–485. Academic Press Professional, San Diego (1994)