Multiple Classifiers Approach for Computational Efficiency in Multi-scale Search Based Face Detection

Hanjin Ryu, Seung Soo Chun, and Sanghoon Sull

Department of Electronics and Computer Engineering, Korea University, 5-1 Anam-dong, Seongbuk-gu, Seoul, 136-701, Korea
{hanjin, sschun, sull}@mpeg.korea.ac.kr

Abstract. The multi-scale search based face detection is essential to use a window scanning technique where the window is scanned pixel-by-pixel to search for faces in various positions and scales within an image. Therefore, detection of faces requires high computation cost which prevents from being used in real time applications. In this paper, we present face detection approach by using multiple classifiers for reducing the search space and improving detection accuracy. We design three face classifiers which take different feature representation of local image[1]: gradient, texture, and pixel intensity features. The designed three face classifiers are trained by error back propagation algorithm. The computational efficiency is achieved by coarse-to-fine classification approach. A coarse location of a face is first classified by the gradient feature based face classifier where the window is scanned in large moving steps. From the coarse location of a face, the fine classification is performed to identify the local image as a face where the window is finely scanned. In fine classification, the output of each face classifier is combined and then used for a reliable judgment on the existence of face. Experimental results demonstrate that our proposed method can significantly reduce the number of scans compared to the exhaustive full scanning technique and provides the high detection rate.

1 Introduction

The detection of face in an image has been intensively studied and a wide variety of techniques have been proposed so far. Among various face detection methods, image based methods recognize face patterns by classifying a local image within a fixed size window into face and non-face prototype classes using statistic models, such as neural network [1, 2, 3], support vector machine [4, 5] and principal components analysis [6]. In order to find faces in various scales and positions within an image, the fixed size window is scanned at all positions for a pyramid of image that is obtained by subsampling the input image. Therefore, the fixed size window that is the basis unit for classifying a face is scanned for multiple images at various scales. Since the fixed size window is exhaustively scanned to find face in images at various resolutions, this method is often referred to as the multi-scale search technique.

Although multi-scale search based face detection methods can provide high detection accuracy on low quality images, they require high computational cost. Therefore,

[1] For convenience, the image within a scanning window is called a local image.

L. Jiao et al. (Eds.): ICNC 2006, Part I, LNCS 4221, pp. 483 – 492, 2006.

in order to reduce the computational cost accompanied by the window scanning procedure on the whole image, some approaches [7, 8] use skin color or object motion to provide prior information on the estimate location of face. Although these approaches can reduce much the computational cost, they cannot be applied to gray scale and static images.

In order to overcome above limitations, coarse-to-fine search approaches have been proposed. The method in [3] proposed a two-stage scheme to overcome the problem of exhaustive full search. In the first stage, a candidate face classifier was used to quickly discard non-face regions, and in the second stage a more complex classifier was used to perform final classification on the local image that passed from first stage successfully. However, the detection rate is lower than the full search process.

Approaches proposed in [9, 10] made use of grid based search method. In each sub-sampled image, each intersection point of a regular grid was tested by a face classifier. If the output value of the face classifier at the intersection points of a grid was greater than a threshold value, the fine search can be started around those points. The grid based search method heavily relies on the grid step.

In this paper, we present multiple classifiers based face detection approach. The multiple face classifiers, which are taken a different feature such as gradient, texture and pixel intensity, are designed to reduce the computational cost while maintaining the high detection rate.

For computational efficiency, we also use coarse-to-fine classification approach. The coarse-to-fine classification is based on improvement of window scanning process which is achieved by increasing the moving step of scanning window. The sub-optimal moving step of scanning window is empirically determined by the sensitivity analysis of each face classifier. Especially, the translation invariant property of adopted gradient feature contributes to improvement of the scanning process in coarse classification stage. The gradient based face classifier is used to find the coarse location of a face where the window is scanned in large moving step. Then, the local image is identified as a face using multiple face classifiers where the window is finely scanned.

The rest of this paper is organized as follows. Section 2 gives an overview of the proposed system. Section 3 addresses the feature representation for multiple face classifiers. Section 4 presents face detection based on coarse-to-fine classification. In order to demonstrate the effectiveness of proposed approach, the experimental results are provided in Section 5. In Section 6, the concluding remarks are drawn.

2 System Overview

The proposed face detection approach is based on multiple classifiers which are composed of three face classifiers. Each face classifier is trained by error back propagation algorithm and taken a different feature representation such as gradient, texture and pixel intensity.

Fig. 1 illustrates the proposed overall system architecture. A pyramid of multi-resolution of the input image is obtained by a scaling factor 1.2. Before classification, each local image is converted to gray image and then pre-processed to reduce the intensity variation. In pre-processing step, a face mask is applied to remove any piece

of the background image. Subsequently, the intensity normalization which consists of a correct lighting [1] and histogram equalization is used to alleviate the variation of lighting condition within local image. After pre-processing, the features of the local image are extracted and then passed to each face classifier. The each face classifier returns a result between 0.0 and 1.0.

For computational efficiency, the coarse-to-fine classification which is based on improvement of window scanning process is utilized. In order to find coarse location of a face, the window is scanned in large moving steps and the local image that might contain a face is examined by the gradient based 1st face classifier. From the coarse location of a face, the other face classifiers identify the local image as a face where the window is finely scanned. As a confidence measure for identifying a face, we apply a weighted sum of the output values from two other face classifiers including 1st face classifier. The identified regions in each scale are mapped back to the input image scale.

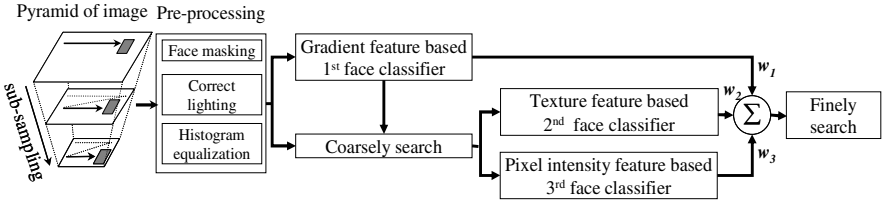


Fig. 1. The overall system architecture based on multiple face classifiers

3 The Feature Representation for Multiple Face Classifiers

Since mixture of various classifiers may give more reliable judgment for a face than using only single classifier, we design the multiple face classifiers which are taken different representations of face patterns. The employed gradient and texture features are represented for global face appearance and the pixel intensity feature is for local face appearance.

3.1 Gradient Feature for 1st Face Classifier

The 1st face classifier is based on gradient feature obtained from the horizontal gradient projection [11]. As shown in Fig. 2, the gradient feature contains the integral information of the pixel distribution, which retains certain invariability among facial features. It is noticeable that the positions of facial features are quite stable even under translating the face center regardless of different amount of gradient strength. This fact provides a clue to improve the window scanning process. That is, if the center of the window falls within permissible bound from the center of face, the 1st face classifier may identify a local image as a face pattern. Therefore, the determination of the permissible bound is a main problem of improving window scanning process, and the solution is described in detail in section 4.

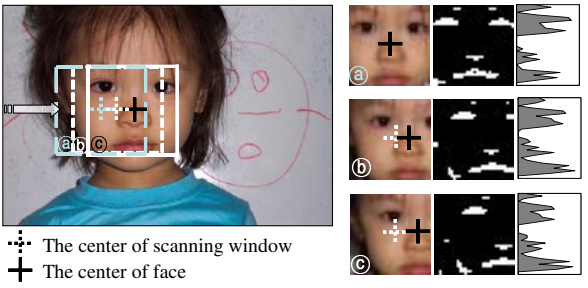


Fig. 2. The gradient feature's characteristic and its translation invariant property

In order to obtain the gradient feature, the horizontal binary edge image ($Edge(i, j)$) is generated by applying the Sobel edge operator with horizontal mask. The gradient feature is defined as equation (1). The $HP(j)$ is the j^{th} entry in the horizontal projection which is formed by summing the pixels in the i^{th} column. The number of edges corresponding to 30 bins is normalized and passed to the 1st face classifier.

$$HP\ (\ j\ ) = \sum_{i=0}^{29} Edge\ \ (i,\ j),\ \ 0 \le\ j \le\ 29\ , \tag{1}$$

3.2 Texture Feature for 2nd Face Classifier

Texture is one of the most important defining characteristics of an image. A face image can be thought of as a regular and symmetric texture pattern. Although a human face has a distinct texture pattern compared to other objects, this property has not been utilized widely in developing face detection. In our system, the texture feature is derived from gray level co-occurrence matrix [12]. The $(i,j)^{th}$ element of the co-occurrence matrix represents the number of times that the pixel with value i occur, in adjacent distance (d) along a direction (θ), related to a pixel with value j in an image.

The texture features are extracted by three measures; correlation, variance and entropy. The correlation is related to the joint probability occurrence of the specified pixel pairs. The variance measures the amount of local variations in an image, whereas the entropy measures the disorder of an image. Feature extraction is processed as follow:

1. The input local image (30×30) is reduced to 10×10 smoothed image by applying the average filter in 3×3 size.
2. The each pixel is quantized into 25 bins for the computational efficiency.
3. Obtain texture features through the following measures;

$$Correlation: \frac{\sum_i \sum_j (i-\mu_x)(j-\mu_y)\,p(i,j)}{\sqrt{\sigma_x \sigma_y}}, \tag{2}$$

$$Variance: \sum_i \sum_j (i-\mu)^2\,p(i,j), \tag{3}$$

$$Entropy := -\sum_i \sum_j p(i,j)\log(p(i,j)), \tag{4}$$

where $p(i, j)$ is the $(i, j)^{th}$ entry of the normalized co-occurrence matrix and $\mu = \mu_x = \mu_y$, because of symmetric matrix.

The extracted 9 texture features (3 measures×3 directions (0°, 45° and 90°)) are passed to the 2nd face classifier.

3.3 Pixel Intensity Feature for 3rd Face Classifier

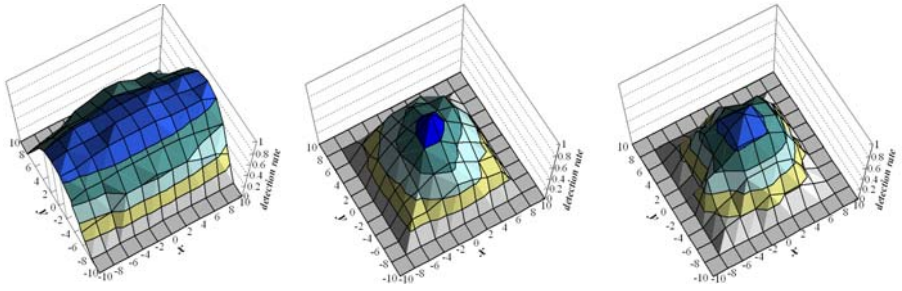
The pixel intensity feature is the most commonly used input vector for neural network based object detection. The methods that use pixel intensity have yielded promising detection performance so far. Especially, the regions of facial features have been proven valuable clues for classifying faces. In our system, the pixel intensity feature is extracted from eye region, because eye region is more reliable than nose and mouth region for determining face pattern.

To extract the feature, a 10×10 smoothed image is first reconstructed from the local image by sub-sampling the local image with 3×3 average-mask. The normalized pixel intensity values of 40 pixels corresponding to eye region (10×4) are finally obtained and passed to the 3rd face classifier.

4 Coarse-to-Fine Classification for Computational Efficiency

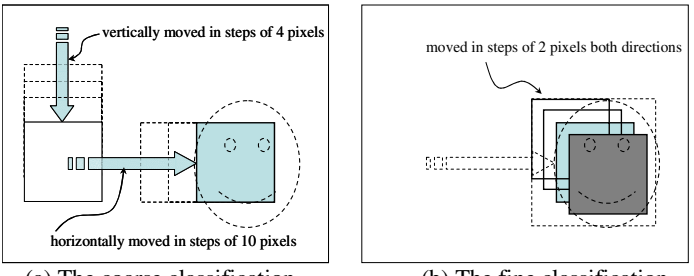
The computational efficiency is achieved by using coarse-to-fine classification approach. In coarse-to-fine classification, the problems that must be solved are related to following two matters. The first matter is how to improve the window scanning process in the coarse classification and the second is how to reliably identify the local image as a face in the fine classification. In order to improve the window scanning process in the coarse classification stage, we use the translation invariant property of the gradient feature that is used in 1st face classifier. That is, if we know the permissible bound of translation, we can easily increase the window moving step to find coarse location of a face. For applying the translation invariant property for the 1st face classifier, we analyze the sensitivity of 1st face classifier with respect to the degrees of shift. That is, we collected a set of 50 images for sensitivity analysis, each of them was cropped around center of face in both x and y directions.

Fig. 3(a) presents the detection rate of the 1st face classifier with respect to shift in both x and y direction when the threshold value was strictly set to 0.8. The detection rate was over 80% when the images were shifted within 10 pixels in x direction and within 4 pixels in y direction. This allows the window moving step for scanning, in the coarse classification stage, to be up to 10 pixels in x direction and 4 pixels in y direction (see Fig. 4(a)). From the coarse location, the fine search is started where the window is shifted by 2 pixels in both x and y directions (see Fig. 4(b)). This is based on observations that the 2nd and the 3rd face classifiers, which are performed in the fine classification stage, have a detection rate of over 80%, when the images were shifted by 2 pixels in both x and y directions as shown in Fig. 3(b) and 3(c), respectively. If the 1st face classifier can not identify the local image as a face, the coarse classification process is applied again.



(a) The 1st face classifier (b) The 2nd face classifier (c) The 3rd face classifier

Fig. 3. The results of sensitivity analysis with respect to shift images



(a) The coarse classification (b) The fine classification

Fig. 4. The moving steps of scanning window

In the fine classification stage, a weighted sum of the results from the multiple face classifiers is utilized to identify the local image as a face. If the weighted sum value is grater than the threshold value (τ), the local image is identified as a face.

5 Experimental Results

5.1 Training Face Classifiers

Each face classifier is trained by error back propagation algorithm and a logistic sigmoid activation function is used in each unit. The numbers of hidden unit of the 1st, 2nd and 3rd face classifier are 10, 4 and 12, respectively.

The 1,056 training images of face pattern came from the benchmark face database (AT&T[2], BioID[3], Stirling[4], Yale [13] dataset) and World Wide Web. The face patterns were manually normalized to 30×30 rectangle including the outer eye corners and upper eyebrows. In addition, we included the mirror-reverse and two rotation angles (5°, -5°) of each image and produced a total of 4,224 examples of faces. The non-face patterns were collected via an iterative bootstrapping procedure [1]. Before training, we used an initial training set of 2,080 non-face patterns from background images. After bootstrapping process, 15,798 non-face patterns were obtained.

[2] http://www.uk.research.att.com/facedatabase.html
[3] http://www.bioid.com/downloads/facedb
[4] http://pics.psych.stir.ac.uk/

5.2 Results on Several Databases

To evaluate the performance of our proposed method, we compared to an exhaustive full scanning method with several databases which were not used in the training process. The test database consisted of four different test sets (IMM[5], Caltech[6], AR database [14] and World Wide Web). The face databases were publicly available on the World Wide Web and often used for the benchmarking of face detection algorithm. The images from the IMM (640×480) and the AR database (768×576) which had a uniform background with various poses, expressions and illuminations, while the Caltech database (896×592) varied a lot with respect to background. The images obtained from World Wide Web varied a lot with respect to image resolution and background.

Table 1 shows a tabulated comparison for the proposed method and the exhaustive full scanning method on test databases. Examples of detection results are shown in Fig. 5 and 6. The applied threshold value (τ) and weight factors (w_1, w_2, and w_3) of the proposed method were empirically set to 0.65, 0.25, 0.35, and 0.4 respectively. As shown in Table 1, the proposed method achieved a detection rate between 93.0% and

Table 1. Experimental results

Test DB	Detection results						Reduction rates of # of scans
	Exhaustive full scanning method			Proposed scanning method			
	Detection rate	# of false	# of scans per image	Detection rate	# of false	# of scans per image	
IMM	96.2 %	28	755,418	95.7 %	8	72,273	90.4 %
Caltech	94.5 %	12	1,369,067	93.0 %	10	176,674	87.1 %
AR	95.7 %	22	1,128,541	95.0 %	6	142,136	87.3 %
WWW	80.7 %	46	4,312,203	83.7 %	12	513,152	88.1 %

Fig. 5. Detection results. The IMM, Caltech, and AR database (from top to bottom).

95.7% which is almost the same as the detection rate achieved by the full scanning method. In terms of the computational cost, we obtained the total number of scans per image used to detect the faces. It can be seen that the proposed method can reduce the number of scans up to 90.4% compared to the exhaustive full scanning method.

[5] http://www2.imm.dtu.dk/~aam/
[6] http://vision.caltech.edu/html-files/archive.html

Fig. 6. Examples of detection results obtained from World Wide Web

5.3 Comparison with Other Methods

To compare to the detection rate of our proposed method with other methods, we used CMU database [3] which is widely used for testing face detectors. Table 2 shows a tabulated comparison of the proposed method and the other methods. It is important to observe that the detection rate of proposed method is similar to the other methods even though the search space is reduced by increasing the moving step of scanning window. Some detection results are shown in Fig. 7.

Table 2. The performance comparison with other methods[7]

	Detection rate	# of false
Rowley method [3]	86.2%	23
Froba method [9]	87.8%	120
Feraud method [10]	86.0%	8
Proposed method (coarse-to-fine search)	86.6%	19
Proposed method (full search)	89.1%	32

6 Conclusions

In this paper, we suggest a way to overcome the computational inefficiency of exhaustive full search which is commonly used in multi-scale search based face detection.

[7] Since we trained 30×30 face pattern differently, the images were scaled up by a factor 1.5 before testing.

Fig. 7. Examples of detection results obtained from CMU database

The proposed multiple face classifiers using coarse-to-fine classification is based on the improvement of window scanning process. In order to improve the window scanning process, we empirically determined the sub-optimal moving step of scanning window by analyzing the detection rate of each classifier for various moving step sizes. Furthermore, multiple face classifiers were designed for the reliable judgment on existence of face. Experimental results shows that our proposed method can reduce a significant amount of computational complexity with a negligible change in detection rate compare to exhaustive full search method.

References

1. Sung, K-K., Poggio, T.: Example based learning for view based human face detection. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20 (1998) 39-51
2. Juell, P., Marsh, R.: A hierarchical neural network for human face detection. Pattern Recognition, Vol. 29 (1996) 781-787
3. Rowley, H., Baluja, S., Kanade, T.: Neural network based face detection, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20 (1998) 23-38
4. Osuna, E., Freund, R., Girosi, F.: Training support vector machines: an approach to face detection. Pro. Conf. Computer Vision and Pattern Recognition, (1997) 130-136
5. Shih. P., Liu, C.: Face detection using discriminating feature analysis and support vector machine. Pattern Recognition, Vol. 39 (2006), 260-276
6. Turk, M., Pentland, A.: Face recognition using eigenfaces. Pro. Conf. Computer Vision and Pattern Recognition (1991) 586-591

7. Yang, J., Waibel, A.: Tracking human faces in real time. Tech. Report CMU-CS-95-210 (1995)
8. Soriano, M., Martinkauppi, B., Hunvinen, S., Laaksonen, M.: Adaptive skin color modeling using the skin lucus for selecting training pixels. Pattern Recognition, Vol. 3 (2003) 681-690
9. Froba, B., Kublbech, C.: Robust face detection at video frame rate based on edge orientation features. Proc. Conf. Automatic Face and Gesture Recognition, (2002) 327-332
10. Feraud, R., Bernier, O. J., Viallet, J-M., Collobert, M.: A fast and accurate face detector based on neural networks. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 23 (2002) 42-53
11. Bebis, G., Uthiram, S., Georgiopoulos, M.: Face detection and verification using generic search. Artificial Intelligence Tools, Vol. 9 (2000) 225-246
12. Peter, R. A., Strickland, R. N: Image complexity metrics for automatic target recognizers. Automatic Target Recognizer System and Technology Conference (1990)
13. Georghiades, A. S., Belhumeur, P. N., Kriegman, D. J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 23 (2001) 643-660
14. Martinez, A. M., Benavente, R.: The AR face database. CVC Tech, Report #24 (1998)