Human face detection algorithm via Haar cascade classifier combined with three additional classifiers

Li Cuimei[1], Qi Zhiliang [2], Jia Nan [2], Wu Jianhua [2*]
1.School of Communication and Electronics, Jiangxi Science & Technology Normal University Nanchang 330013, China
2.School of Information Engineering Nanchang University Nanchang 330031, China
Email: *jhwu@ncu.edu.cn*

***Abstract*—*Human face detection has been a challenging issue in the areas of image processing and patter recognition. A new human face detection algorithm by primitive Haar cascade algorithm combined with three additional weak classifiers is proposed in this paper. The three weak classifiers are based on skin hue histogram matching, eyes detection and mouth detection. First, images of people are processed by a primitive Haar cascade classifier, nearly without wrong human face rejection (very low rate of false negative) but with some wrong acceptance (false positive). Secondly, to get rid of these wrongly accepted non-human faces, a weak classifier based on face skin hue histogram matching is applied and a majority of non-human faces are removed. Next, another weak classifier based on eyes detection is appended and some residual non-human faces are determined and rejected. Finally, a mouth detection operation is utilized to the remaining non-human faces and the false positive rate is further decreased. With the help of OpenCV, test results on images of people under different occlusions and illuminations and some degree of orientations and rotations, in both training set and test set show that the proposed algorithm is effective and achieves state-of-the-art performance. Furthermore, it is efficient because of its easiness and simplicity of implementation.*

***Keywords*—*Human face detection; Haar-like features; face skin hue histogram match; eyes detection; mouth detection; cascade classifier; weak classifier*

I. INTRODUCTION

Human face detection has been playing an important part in human-machine interaction and computer vision based applications [1]. As a human identity information, human face has the advantages of uniqueness and non-replicability [1]. However, due to a large variations such as different illuminations, expressions and backgrounds and other uncertainties, human face detection remains a challenging issue in real world applications.

In 2002, Ojala utilized local binary pattern (LBP) for classification of image textures [2]. LBP is a grayscale irrelevant texture operator with powerful discrimination and has also been used in human face detection and recognition. However, although LBP features have great discriminative power, they miss the local structure under some certain circumstances. In 2004, Viola proposed a cascade classifier algorithm and used Haar-like features for human face detection [3-4]. Since then many researchers have worked in this area based on Haar-like feature or its modifications and achieved much progress [5-6]. The detection speed has approached the objective of practical use due to the simplicity and effectiveness of Haar-like features. However, the detection speed is obtained at the cost of some wrong acceptance of non-human faces although the rejection error for human faces is significantly small. The reason is that in essence, Viola-Jones' rejection cascade of classifiers are combined by AdaBoost to form a strong classifier with each node being a set of weak classifiers using Haar-like features. The cascaded classifier is a supervised classifier where sub-windows of multi-resolutions over an image are sequentially tested against all nodes in the cascade. A window which passes all nodes is regarded as candidate of a human face. In Viola-Jones' cascade each node is designed to have a high (say 99.9%) detection rate (low false negatives) at the cost of a low rate of true positives (near 50%, a little better than guess). This means a high rate of false positives. At any node a decision of rejection terminates the detection while the detection process continues when a temporary acceptance is made. Thus computation time is greatly saved because sub-windows containing human faces are extremely much lesser than non-face sub-windows from an image.

Therefore, emphasis should be laid on subsequent weak classifiers as long as they keep a sufficiently low rate of false negatives (wrong rejection for real human faces), guaranteeing that almost all sub-windows containing human faces pass all nodes of cascaded classifiers. In this paper, a new human face detection algorithm is proposed on a basis of this idea. Three additional weak classifiers are subsequently appended to the primitive Haar-like features based cascaded classifiers. One is a decision node based on human skin hue histogram matching because the skin tone for a specific race of people has a relatively determined distribution. Of course it is not sure that a sub-window from an image contains a human face if this sub-window is hue-distribution matched. However it would be certain to assure a rejection if the sub-window does not match the face skin distribution. The prototype of face skin histogram is obtained in this paper by a statistic analysis of 30 images of people in which 344 human faces are included. The second and the third weak classifiers are based on eyes and mouth detections. Because eyes and

978-1-5090-5035-2/17/$31.00 ©2017 IEEE

mouth detections are also implemented with Haar-like features based cascade classifiers, both of them have a sufficiently high detection rate, satisfying conditions of weak classifiers. Experimental results show that the proposed human detection algorithm compensates the shortcomings of the primitive Viola-Jones' cascade classifier and makes the whole human face detection rate higher while keeping nearly zero wrong rejection.

The remainder of this manuscript is organized as follows. Section II reviews some related work including basis for Haar-like features, cascade of classifiers and color model of HSV and skin tone histogram matching. A new human detection algorithm based on the primitive cascaded classifier and three additional weak classifies is proposed in detail in section III. Experimental results are given in section IV and finally in section V, conclusions are drawn.

II. RELATED WORK AND SOME BASICS

A. Viola and Jones' Haar-like features and cascade classifiers

The typical cascade classifier is the very successful method of Viola and Jones for face detection [3-4]. Generally, many object detection tasks with rigid structure can be addressed by means of this method, not limited to face detection. The cascade classifier is a tree-based technology, in which Viola and Jones used Haar-like features for human face detection.

The Haar-like features by default are shown in Figure 1 [7], which can be used with all scales in the boosted classifier and can be rapidly computed from an integral version of the image to be detected in.

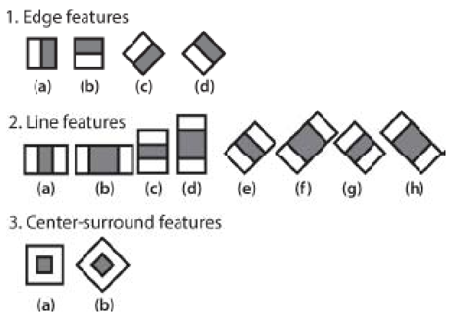

Fig. 1 Haar-like features from the OpenCV source distribution. In this wavelet representation, the light region is interpreted as "add that area" and the dark region as "subtract area.

There is also a good feature, namely LBP, which was originally proposed by Ojala for use as a kind of texture descriptor [2]. It was later adapted in the boosted cascade environment of the Viola-Jones' object detection. But in this paper we are only interested in Haar-like features.

The Viola-Jones' detector uses AdaBoost, called a rejection cascade, which is a series of nodes, with each node being a definite multi-tree AdaBoosted classifier. Figure 2 shows the Viola-Jones' rejection cascade, composed of many boosted classifier groups of decision trees trained on the features from faces and non-faces or other training objects to be detected. In case of faces, almost all (99.9%) of the faces are found, but many non-faces (about 50%) are wrongly regarded as positive ones. This is feasible, because a sufficiently large number (say 20) of such nodes will yield a false positive rate of only $0.5^{20} \approx 0.0001\%$ and a face detection rate of $0.999^{20} \approx 98\%$.

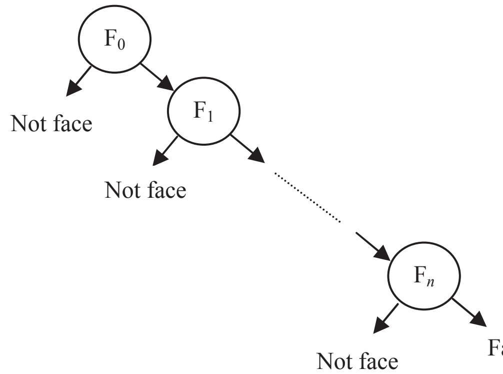

Fig. 2 Rejection cascade by Viola-Jones, each node being a multi-tree of boosted classifiers trained in such a way that almost all non-faces are rejected at the last node nearly without missing a real human face.

B. Color Model HSV and histogram matching

The HSV color model is an ideal tool for developing image processing algorithms based on color descriptions, which are natural and intuitive to human observers of images. The HSV model decouples the intensity (V) from color dimensions, hue (H) and saturation (S). After the RGB-HSV transform, hue is a color attribute that tells the observer what color is perceived (pure yellow, green, or red). For people of a specific race, say yellow race, the face skin color follows specific distribution, which does not change a lot under different light conditions according to principle of color constancy. Thus the hue is a relatively robust feature that carries skin tone information. The RGB-HSV transform formula are given by [8]:

$$H = \begin{cases} \theta & \text{if } B \le G \\ 360 - \theta & \text{if } B > G \end{cases} \tag{1}$$

with $$\theta = \cos^{-1}\left\{ \frac{[(R-G)+(R-B)]/2}{[(R-G)^2 + (R-B)(G-B)]^{1/2}} \right\}$$

$$S = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)] \tag{2}$$

$$V = (R+G+B)/3 \tag{3}$$

One cannot expect different people to have exactly the same skin tones, however, the skin tone is subject to a relatively fixed distribution. The skin tone model has been used in human skin segmentation and also, used in human face detection [9]. In this paper, for a human face candidate already detected by the rejection cascade at initial phase, we use skin hue histogram to characterize the skin tone's distribution and append a decision tree.

C.Eyes detection and mouth detection within a human face candidate

By means of Haar-like features and taking the advantage of conception of cascade classifiers, one can design and implement eyes and mouth detections. Similarly to rejection cascade for human face, the classifiers for eyes and mouth detections are used as weak classifiers, making the whole classification system stronger.

III. PROPOSED HUMAN FACE DETECTION METHOD

On account of the aforementioned background and discussions, a new human face detection algorithm is proposed to implement a stronger whole detection system by appending three weak classifiers, i.e., a classifier based on skin tone histogram matching, a classifier based on eyes detection, and third, a classifier based on mouth detection. The proposed framework is shown in Figure 3. Some interpretation are noted as follows.

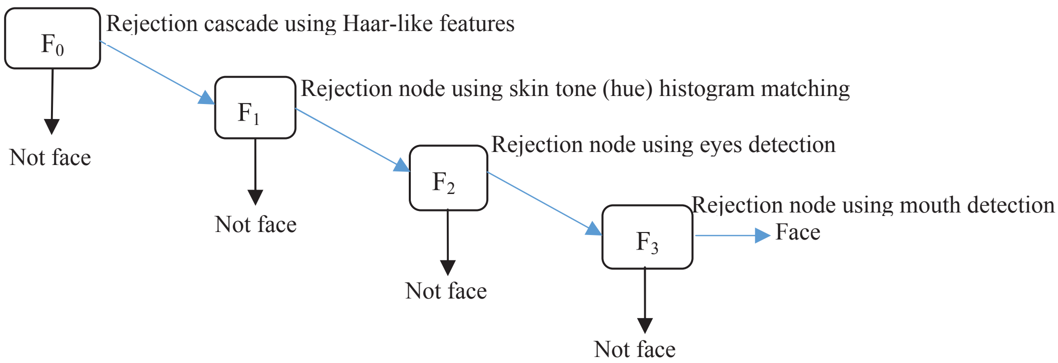

Fig. 3 Proposed framework of human face detection cascade

Note 1: Training for obtaining prototype of skin hue histogram

The proposed human face detection is implemented with the help of OpenCV. However, at the stage of skin hue histogram matching, the prototype of histogram should be obtained before the test phase of human face detection.

Note 2: Implementation of initial face detection

The block F_0 is itself a cascade of classifiers using Haar-like features. The cascade classifier supplied OpenCV 3.3.0 will be used, so it needs not additional training in the work.

Note 3: Implementation of eyes and mouth detections

The eyes and mouth detections are also implemented by calling modules in OpenCV, so it needs not a training process before the stage of test either.

The histogram matching means comparison of two histograms so as to measure the difference between a human face candidate's skin hue histogram and the prototype of hue histogram of training (real) human faces. There are quite a lot of distance measure to implement this task, 3 of which are as follows (given two histograms denoted as H_1 and H_2):

(1) Correlation:

$$d(H_1,H_2)=\frac{\sum_i (H_1(i)-\bar{H}_1)(H_2(i)-\bar{H}_2)}{\sqrt{\sum_i (H_1(i)-\bar{H}_1)^2 \sum_i (H_2(i)-\bar{H}_2)^2}} \tag{4}$$

where an hyphen above H means average.

(2) Chi-square:

$$d(H_1,H_2)=\sum_i \frac{[H_1(i)-H_2(i)]^2}{H_1(i)} \tag{5}$$

(3) Histogram intersection:

$$d(H_1,H_2)=\sum_i \min[H_1(i),H_2(i)] \tag{6}$$

(4) Bhattacharyya distance:

$$d(H_1,H_2)=\sqrt{1-\frac{1}{\bar{H}_1\bar{H}_1 N^2}\sum_i \sqrt{H_1(i)H_2(i)}} \tag{7}$$

where N is the number of bins of the histograms. In this paper, the metric of correlation is chosen. This is an empirical choice.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experiments in this work is completed under an environment of Visual Studio 2017, incorporated with modules of recently released OpenCV 3.3.0, in a computer of desktop with a processor of Intel(R)

Core™ i7-2630QM CPU @ 2.00 GHz and an internal RAM of 8 GB. It takes minutes for detecting and verifying 344 faces in 30 images of people with different backgrounds and light conditions.

A.Results for training set

We select 30 color images of people of our friends, students, colleagues, and family members, with their consent. All people are of yellow race, including 344 human faces. These are manually verified from initially detected faces (440 true and false faces). After removing 96 wrongly detected faces in initial phase, the prototype of skin hue histogram is determined by remaining 344 skin hue histograms of true faces. Then, these 30 color images of people are again processed by the proposed human face detection system. A statistic of human face detection results are shown in Table I, where PPV means positive prediction value, defined as:

$$PPV = TP / (TP + FP) \times 100\% \qquad (8)$$

where *TP* means true positives, i.e., the number of correct detected faces, and *FP* means false positives, i.e., wrongly detected faces (non-faces as faces). Because the true negatives (*TN*) is very large and the false negatives (*FN*) is very small, comparison of other measures, such as sensitivity and specificity partly in terms of *FN* or *TN* is of little significance. Thus, only the results of measure of *PPV* are given in Table I.

Table I. Results for training images

Number of images	30	
Real human faces	344	
Faces detected by primitive cascade of classifiers	440	PPV: 78.18%
Faces detected by proposed detection system	351	PPV: 98.01%

From Table I, it can be seen that the proposed detection method greatly improves the detection performance from a positive prediction value of 78.18% to 98.01%, by an increment of near 20%. The reason is that by skin hue histogram matching, one can remove a proportion of false human faces and by detecting eyes and mouths, one can get rid of another proportion of false positives. An example of images is shown in Figure 4.

Fig. 4 An example of face detection by the proposed algorithm

In Figure 4, it can be seen that the initial detector at the former part of the proposed detection system detected 13 faces, including 7 true faces and 6 false faces. 5 black circles indicate those false faces rejected by the weak classifier based on skin tone histogram matching, and the little white circle at the middle-right side indicates a false face rejected at the last node based on eyes and mouth detections.

B.Results for test set

We have randomly downloaded several dozens of images of people from Internet. These images, include about 400 human faces, are used as test set for further verifying the proposed algorithm. Detection results for the test set have been obtained and the PPV values are nearly the same to Table I, which are then not shown here for simplicity. The results are reasonable and predictable, because the cascade classifiers we used in both the stages of initial rejection cascade and rejection cascades of eyes and mouth detections are all ones from the package of OpenCV 3.3.0. Only the prototype of skin tone (hue) histogram was trained from our selected images of people. Due to principle of color constancy, the skin hue is subject to a relatively fixed of distribution (Figure 5). One can expect that the prototype of histograms of trained set would generalize to other human faces well, as long as new human faces to be tested are of the same race with people in the training images, say yellow race, regardless of different light conditions.

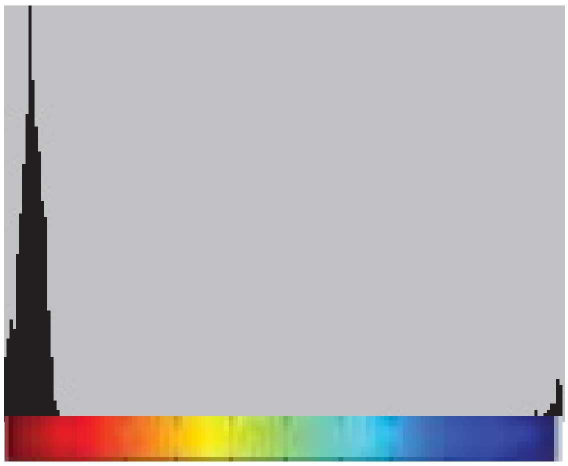

Fig. 5 Prototype of human face skin tone histogram of trained images

V. CONCLUSION AND FUTURE WORK

In this paper, a new human face detection algorithm is proposed on a basis of cascade classifiers using Haar-like features. Three additional weak classifiers are subsequently appended to the primitive Haar-like features based cascaded classifiers. One is a decision node based on human skin hue histogram matching. The second and the third weak classifiers are based on eyes and mouth detections, respectively. Because eyes and mouth detections are also implemented with Haar-like features based cascade classifiers, both of them have a sufficiently high detection rate, satisfying conditions of weak classifiers. Experimental results show that the proposed human detection algorithm compensates the shortcomings of the primitive Viola-Jones' cascade classifier and makes the whole human face detection rate higher while keeping nearly zero wrong rejection.

The contributions of this work can be concluded as below. (1) A weak classifier based on human face skin tone histogram can reject a big proportion of non-faces wrongly detected by the primitive Viola-Jones' Haar-like features-based cascade classifiers. (2) 2 additional classifiers based on eyes and mouth detections further remove those non-faces whose colors happen to be in accordance with the human skin color, but there are probably no eyes- and mouth-like objects in it. (3) The proposed human face detection system is simple to implement due to availability of modules in OpenCV.

In future, more research work should continually focus on human face detection for people of different races, instead of faces of single race as in our work. Computation time should also be further saved for real world applications.

ACKNOWLEDGMENT

This work is in part financially supported by National Natural Science Foundation of China under Grant No. 61662047 and by the Teaching Innovation Program of Jiangxi Provincial Department of Education under Grant No. JXJG-16-10-11).

REFERENCES

[1] Zhang K, Zhang Z, Li Z. Joint face detection and alignment using multitask cascaded convolutional networks [J], IEEE signal Processing Letters, 2016, 23(10):1499-1503.

[2] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary pattern [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.

[3] Voila P, Jones M J. Rapid object detection using a boosted cascade of simple features [C], Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (ICCVPR 2001), vol. 1, pp. I-511-I-518, 8-14 Dec. 2001, Kauai, USA.

[4] Viola P, Jones M. Robust real-time face detection [J], International Journal of Computer Visio , 57(2):137-154.

[5] Jin H, Liu Q, Lu H, Tong X. Face detection using improved Bayesian framework [C], Proceedings of the Third International Conference on Image and Graphics (ICIG'04), pp. 306-309, 2004.

[6] Mandal B. Face recognition: Perspectives from the real world [C], IEEE 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV), pp. 1-5, 13-15 Nov. 2016, Phuket, Thailand.

[7] Kaehler A, Bradski G, Learning OpenCV 3 [M], O'reilly Media, Inc., 2016, Sebastopol, CA 95472, USA.

[8] Gonzalez R C, Woods R E. Digital Image Processing (3rd Edition) [M], Prentice Hall, 2008, New Jersey 07458, USA.

[9] Moallem P, Mousavi B S, Monadjemi S A, A novel fuzzy base system pose independent faces detection [J], Applied Soft Computing, 2011, 11:1801-1810.

AUTHORS' BIOGRAPHY

Li Cuimei was born in December 1971 and received the B. Sc. and M. Eng. degrees in Jiangxi Polytechnic University, Nanchang, China, respectively in 1989 and in 2003. Now she works as a professor in School of Communication and Electronics, Jiangxi Science and Technology Normal University, Nanchang, China. Her research interests include image processing, patter recognition, computer applications, etc.

Qi Zhiliang received the B. Sc. degree in Xinyu University, Xinyu, Jiangxi, China, in 2015. He is currently a master student in School of Information Engineering, Nanchang University, Nanchang, China, majoring in information and communication engineering. His research interests include image processing, machine learning based object detection and localization.

Jia Nan received her B. Sc. degree in Changzhi University, Changzhi, Shanxi, China, in 2015. She is currently a master candidate in School of Information Engineering, Nanchang University, Nanchang, China, majoring in information and communication engineering. Her research interests include image processing, deep learning based neural network and medical imaging.

Wu Jianhua received his B. Sc. degree in Harbin Institute of Technology, Harbin, China, in 1982, and M. Sc. degree in South China Institute of Technology, Guangzhou, China, in 1985. He received his Ph.D. in the University of Poitiers, Poitiers, France, in 2005. He is currently a professor with Department of Electronic Information Engineering, Nanchang University, Nanchang, China. He also works as a supervisor for Ph.D. candidates. He has published more than twenty papers in journals such as Optics Communications, Optics and Laser Technology, etc. His research interests include image and signal processing, power quality signal analysis, pattern recognition, etc. His current research work is partly financed by National Natural Foundation of China.