

Face Detection using Variance based Haar-Like feature and SVM

Cuong Nguyen Khac, Ju H. Park, Ho-Youl Jung

Abstract—This paper proposes a new approach to perform the problem of real-time face detection. The proposed method combines primitive Haar-Like feature and variance value to construct a new feature, so-called Variance based Haar-Like feature. Face in image can be represented with a small quantity of features using this new feature. We used SVM instead of AdaBoost for training and classification. We made a database containing 5,000 face samples and 10,000 non-face samples extracted from real images for learning purposed. The 5,000 face samples contain many images which have many differences of light conditions. And experiments showed that face detection system using Variance based Haar-Like feature and SVM can be much more efficient than face detection system using primitive Haar-Like feature and AdaBoost. We tested our method on two Face databases and one Non-Face database. We have obtained 96.17% of correct detection rate on YaleB face database, which is higher 4.21% than that of using primitive Haar-Like feature and AdaBoost.

Keywords—AdaBoost, Haar-Like feature, SVM, variance, Variance based Haar-Like feature.

I. INTRODUCTION

OBJECT detection is important for a variety of applications, such as visual surveillance, autonomous vehicle, and automatic driver-assistance system. But it is a challenging task because of the wide variability of objects. One of many challenges is feature extraction from image. Quality of an object detection system depends on the feature extraction. There are many kinds of feature which we can use for object detection.

We can see that Haar-Like feature is a good feature. It is used in a lot of applications for object detection. Haar-like features selected by AdaBoost are used to capture subtle structure of license plate for license plate detection [1]. In field of car detection, combination of Haar-Like feature and HOG (Histogram of Oriented Gradient) is a way to encode an input image to obtain a vector of visual descriptors [2]. Motion features are used to given the candidate regions and Haar-Like features are used to detect the vehicles [3]. Haar-Like features and concept of ROI (Region Of Interest) can significantly increase detection rates for car detection [4, 5]. In vehicle safety systems, Haar-Like feature can be used for zebra detection [6], real-time road sign detection [7]. In addition, Haar-Like feature is useful in low cost speech detection [8].

Cuong Nguyen Khac, Department of Information and Communication of Engineering - Yeungnam University - Korea (phone: +82-10-7219-8879; fax: +82-53-810-4742; e-mail: cuongnguyenkhac@hotmail.com).

Ju H. Park, Ph. D, Professor, School of Electrical Engineering and Computer Science, Yeungnam University - Korea (e-mail: jessie@ynu.ac.kr).

Ho-Youl Jung, Ph.D, Associate Professor, School of Electrical Engineering and Computer Science, Yeungnam University - Korea (phone: +82.53.810.3545; e-mail: hoyoul@yumail.ac.kr; http://yu.ac.kr/ hoyoul).

In above systems, Haar-Like feature is useful for object detection. If we consider face detection then Haar-Like feature is used to construct a framework for robust and extremely rapid face detection. The calculation of Haar-Like feature is not complex. However, the necessary quantity of Haar-Like features need to extract is very huge. With size of image is 384x288 pixels and the base resolution of the detector is 24x24 pixels, the exhaustive set of features is 160,000 features, a number far larger than the number of pixels within image [9, 10, 11]. Even though each feature can be computed very efficiently, computing the complete set is prohibitively expensive. So, if use Haar-Like for feature extraction then we have to select a small number of the critical features from these 160,000 features and reject as much as possible the number of non-critical features for real-time application. And AdaBoost can resolves this problem, but the processing time of feature extraction with a large number of features is long.

The contributions of this paper consist of two parts. The first, a new feature is constructed, so-called Variance based Haar-Like feature. The second, an efficient scheme is proposed to combine the Variance based Haar-Like feature and a very strong learning machine SVM for real-time face detection.

The remaining sections of this paper will describe the related theory, proposed method, experiments. Section II will describe basis of variance value and calculation of Variance based Haar-Like feature using integral and squared integral image. Section III will describe our proposed architecture and details of training stage and classification stage. Section IV will describe experiments and results, which we have obtained. Section V will discuss about the efficiencies of proposed method.

II. FEATURES

A. Variance

In many papers, Haar-Like feature can be used with covariance for pedestrian detection [12], with nonorthogonal feature for matching, reconstruction [13], with stochastic context-free grammar for hand gesture recognition [14]. In this paper, we propose the method using Haar-Like feature with variance value for face detection.

In probability theory and statistics, the variance of a random variable or distribution is the expected square deviation of that variable from its expected value or mean. For a given random variable X , the variance value of X is as follow:

$$Var(X) = E(X^2) - \mu^2 \quad (1)$$

Where $E(X^2)$ is expected value of squared of X and μ is expected value of X .

Because of the quantity of feature is small when we use the Variance based Haar-Like feature for feature extraction. So, it is necessary to enhance the representation of relationship between neighbouring regions in image. In this paper, within each rectangle region of feature extraction, we use sum of variance values in each region to encode the existence of oriented contrasts between regions in an image. Next section, we will describe the calculation of Variance based Haar-Like feature.

B. Variance based Haar-Like feature

Based on the method of Viola and Jones [9, 10, 11], we used four simple rectangle features only to learn and classify images.

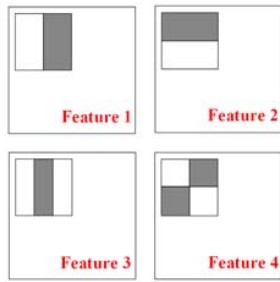


Fig. 1: Four kinds of basis feature are used to extract features of face

The regions within these rectangle features have the same size, shape and horizontally or vertically adjacent (Fig. 1). The value of a two-rectangle feature (*Feature 1* and *Feature 2*) is the difference between the sum of variance values within two rectangular regions. A three-rectangle feature (*Feature 3*) computes the sum of variance values within two outside rectangles subtracted from the sum of variance values in a center rectangle. And a four-rectangle feature (*Feature 4*) computes the difference of the sum of variance values between diagonal pairs of rectangles.

Processing time of feature calculation is very important for real-time face detection. Next section will describe a fast method for feature calculation using integral and squared integral image. We can obtain very quickly the values of Variance based Haar-Like feature at any position in an image.

C. Integral Image and Squared Integral Image

Similar to primitive Haar-Like feature, the Variance based Haar-Like feature also can be calculated very quickly.

For given image $f(x, y)$, we obtain integral image $I(x, y)$ and squared integral image $I^2(x, y)$ as follow:

$$I(x, y) = \sum_{m=1}^x \sum_{n=1}^y f(m, n) \quad (2)$$

$$I^2(x, y) = \sum_{m=1}^x \sum_{n=1}^y f^2(m, n) \quad (3)$$

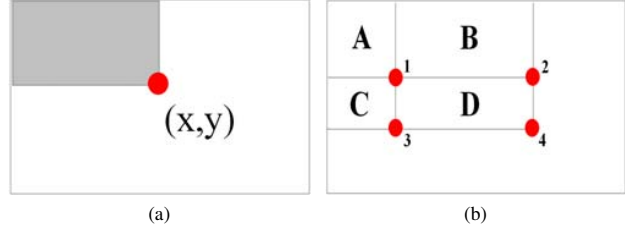


Fig. 2: (a) Integral image and (b) Sum of the pixels within D region is $1 + 4 - (2 + 3)$

Where $I(x, y)$ and $I^2(x, y)$ indicate the sum and the sum of squared of the pixels above and to the left of x, y (Fig. 2a). Using the integral image $I(x, y)$, any rectangular sum can be computed in four array references (Fig. 2b). The sum of the pixels within rectangle D can be computed with values at four position 1, 2, 3, 4. The value of the integral image at location 1 is the sum of the pixels in rectangle A. The value at location 2 is $A + B$, at location 3 is $A + C$, and at location 4 is $A + B + C + D$. So, the sum within region D can be computed as: $1 + 4 - 2 - 3$

We can use the concepts of integral image and squared integral image for calculation of $E(X^2)$ and μ very fast at any position in an image:

$$\mu = \frac{1}{N} (I_1 + I_4 - I_2 - I_3) \quad (4)$$

$$E(f(x, y)^2) = \frac{1}{N} (I_1^2 + I_4^2 - I_2^2 - I_3^2) \quad (5)$$

And, the variance value can be obtained, such as:

$$Var(f(x, y)) = E(f(x, y)^2) - \mu^2 \quad (6)$$

Where N is number of elements within region D.

So, the Variance based Haar-Like feature can be calculated very quickly by using above methods. We propose an architecture, which combines Variance based Haar-Like feature and SVM for face detection. And the basis informations of this architecture will be described in Section III.

III. PROPOSED ARCHITECTURE

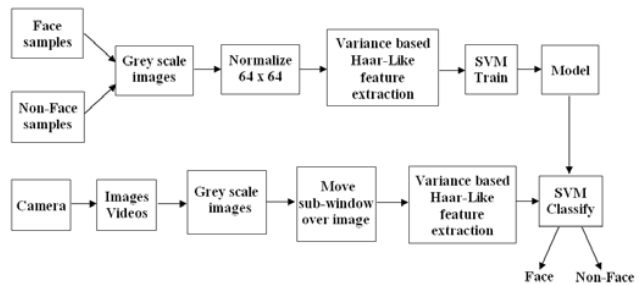


Fig. 3: Architecture of proposed face detection system

We made a database of positive samples consist of 5,000 frontal face images. And a database of negative samples, which

consist of 10,000 non-face images. The frontal face samples just contain eyes, nose, mouth with a lot of differences of size and light condition (Fig. 4).



Fig. 4: Frontal face samples from our face database using for training

Our method consists of two stages. We used SVM for both of training stage and classification stage. We also applied several parameters to improve the correct detection rate. Next two sections will describe more detail about these two stages.

A. Training Stage

All of face and non-face samples are converted to grey scale images and resized to 64x64 pixels. A sub-window with size of 24x24 pixels is used to move 8 pixels for each step over a sample image. 144 feature values are obtained from this processing (each sub-window extracts 4 features). Next, a sub-window with size of 12x12 pixels is used. The number of pixels for each moving step is equal to 4. And 784 feature values are obtained after this processing. Totally, with each sample image, a 928-dimension vector is achieved. In this stage, the Variance based Haar-Like feature is used. Quantity of the features (928 features) is smaller than quantity of pixels within sample image ($64 \times 64 = 4,096$ pixels). This feature extraction method is used to extract the features from 5,000 face samples and 10,000 non-face samples for training to create a model. Next section will describe how to use this model for image classification.

B. Classification Stage

First of all, input image is converted to grey scale image. Then, a detector window with size of 12x12 pixels is created. Moving over input image and increasing the size of detector window using scales which are a factor of 1.25 apart, this processing step is similar to processing step of Viola and Jones [11]. But we did not resize input image. Each sub-image within detector window is resized to 64x64 pixels and extracted features using Variance based Haar-Like feature as described above. After that, the image in each sub-window is classified by using SVM system. We tested our method with many size of input images. The correct detection and processing time depend on size of input images. Section IV will describe the results which we have obtained.

IV. EXPERIMENT

The experimental results were considered on Close test and Open test.

A. Close Test

In this test, we have performed our method with input data, which was 5,000 face images and 10,000 non-face images that we used for training stage. And we obtained the correct detection rate was 100%. It means that our method can recognize exactly all of sample images which we used for training. From this result, we could see that Variance based Haar-Like feature is appropriate for face detection.

To estimate exactly performance of our method, an Open test was considered.

B. Open Test

In this step, we used two World-Test Sets to test our method. The correct detection rate of our method is compared with rate of method based on Haar-Like and AdaBoost [11].

Method 1: Primitive Haar-Like feature and AdaBoost

Method 2: Variance based Haar-Like feature and SVM

TABLE I: Test on CALTECH face database

Method	Number of faces	Number of correct detection faces	Correct detection rate
Method 1	435	311	71.49%
Method 2	435	344	79.08%

TABLE II: Test on YaleB face database

Method	Number of faces	Number of correct detection faces	Correct detection rate
Method 1	5850	5339	91.26%
Method 2	5850	5626	96.17%

TABLE III: Test on non-face database

Method	Number of non-faces	Number of incorrect detection	Error detection rate
Method 1	3021	41	1.35%
Method 2	3021	33	1.09%

Table 1 and Table 2 show that two methods were tested on the small face database (CalTech) and big face database (YaleB) for comparison. The correct detection rate of our method is higher than that of using primitive Haar-Like and AdaBoost. Table 3 shows that two methods were tested on the non-face database. And the error detection rate of our method is less than that of using primitive Haar-Like and AdaBoost.

Figure 5 and 6 show ROC curves of our method and method using primitive Haar-Like and AdaBoost on two face databases. From 100 to 300 samples of CalTech face database, the correct detection rate of two methods is equivalent. Above 400 samples, the correct detection rate of our method is higher. And test on YaleB face database, above 3,000 samples, the correct detection rate of our method is also higher than that of using primitive Haar-Like feature and AdaBoost.

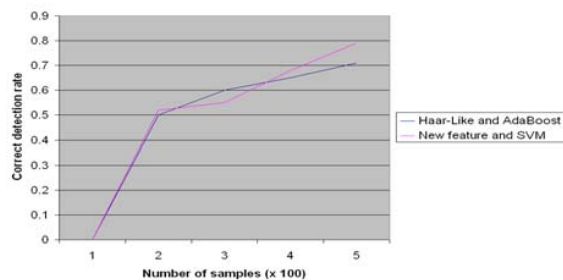


Fig. 5: ROC curves on CalTech face database

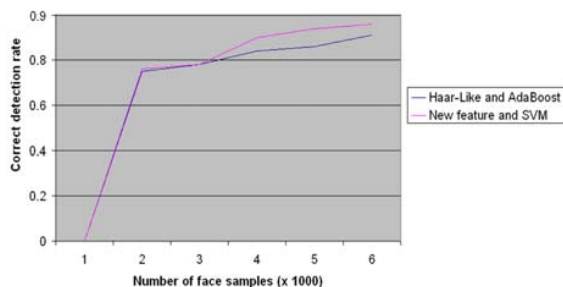


Fig. 6: ROC curves on YaleB face database

V. CONCLUSION

In this paper, we focus on detection of frontal faces. Our method does not need to apply a lot of image preprocessing with input images. The processing time is fast enough which is useful for variety of real-time applications, such as visual surveillance and automatic driver-assistance system.

In this paper, we used four rectangle features for feature extraction and apply SVM for training and classification. Each image is extracted features with two levels, which are high level and low level. These two levels of feature extraction is similar to considering a face at far (24×24 detector) and near (12×12 detector) distances, which are enough to realize the features of a face by using the new feature and SVM. Experiments showed that our method has high correct detection rate in face detection, which was compared with method based on primitive Haar-Like feature and AdaBoost. Fig. 7 show the examples of our test with the images, which contain complex backgrounds and many different poses of faces.



Fig. 7: (a) Our test with outdoor image (b) Our test with indoor image

VI. FUTURE WORKS

We will find the way which can continuously reduce the quantity of feature values (smaller than 928). And we will increase the number of rotated face samples and non-face samples to make a more good model. We can use this new feature to classify faces in a lot of poses with many differences of light condition. Next works, we will test our method with rotated faces and pedestrian objects in daytime and nighttime for intelligent vehicle project. And we believe that the new feature will be useful for real-time applications of object detection.

REFERENCES

- [1] Tan, H, Chen, H, *Selecting Frequency Feature for License Plate Detection Based on AdaBoost*, Proceedings of SPIE-IS and T Electronic Imaging-Visual Communications and Image Processing
- [2] Negri, P, Clady, X, Prevost, L, *Benchmarking Haar and Histograms of Oriented Gradients features applied to vehicle detection*, Proceedings of the Fourth International Conference on Informatics in Control, Automation and Robotics, 2007
- [3] Bai, H, Wu, J, Liu, C, *Motion and haar-like features based vehicle detection*, Multi-Media Modelling Conference Proceedings, 2006 12th International
- [4] Stanculescu, B, Breheret, A, Moutarde, F, *Introducing New AdaBoost Features for Real-Time Vehicle Detection*, Proceedings of COGIS'07 conference on COGNitive systems with Interactive Sensors, held in Stanford University California, Nov 2007
- [5] Haselhoff, A, Kummert, A, Schneider, G, *Radar-Vision Fusion with an Application to Car-Following using an Improved AdaBoost Detection Algorithm*, Intelligent Transportation Systems Conference, ITSC 2007, IEEE, pages: 854-858, Sept 2007
- [6] Lausser, L, Schwenker, F, Palm, G, *Detecting zebra crossings utilizing AdaBoost*, 16th European Symposium on Artificial Neural Networks, Bruges, Belgium, April 23-25, 2008
- [7] Yoon, C, Cheon, M, Kim, E, Park, M, Lee, H, *Real-time road sign detection using Adaboost and Multicandidate*, 2007 International Symposium on Advanced Intelligent Systems, Sokcho, Korea
- [8] Nishimura, J, Kuroda, T, *Low cost speech detection using Haar-like filtering for sensornet*, Signal Processing, 2008. ICSP 2008. 9th International Conference, pages: 2608-2611, Oct. 2008
- [9] Viola, P, Jones, M, *Rapid object detection using a boosted cascade of simple features*, Proceedings of the 2001 IEEE Computer Society Conference, vol 1, pages: 511-518, 2001
- [10] Viola, P, Jones, M, *Robust Real-time Object Detection*, Second international workshop on statistical and computational theories of vision - modeling, learning, computing, and sampling. Vancouver, Canada, July 13, 2001
- [11] Viola, P, Jones, M, *Robust real-time face detection*, International Journal of Computer Vision 57(2), 137-154, 2004, vol 2, pages: 747-747, July 2001
- [12] Paisitkriangkrai, S, Shen, C, Zhang, J, *Fast Pedestrian Detection Using a Cascade of Boosted Covariance Features*, Circuits and Systems for Video Technology, IEEE Transactions, vol 18, pages: 1140-1151, Aug 2008
- [13] Feng Tang, Crabb, R, Hai Tao, *Representing Images Using Nonorthogonal Haar-Like Bases*, Pattern Analysis and Machine Intelligence, IEEE Transactions, vol 29, pages: 2120-2134, Dec. 2007
- [14] Chen, Q, Georganas, N, D, Petriu, E, M, *Hand Gesture Recognition Using Haar-Like Features and a Stochastic Context-Free Grammar*, Instrumentation and Measurement, IEEE Transactions, vol 57, pages: 1562-1571, Aug 2008
- [15] Sonka, M, Hlavac, V, Boyle, R, *Image Processing, Analysis, and Machine Vision*. Boston, MA: PWS-Kent, 1999.
- [16] Andrew Webb, *Statistical Pattern Recognition*, Oxford University Press, New York, 1999.