Adaboost Face Detection Based on Improved Covariance Feature

Rui Li
Lanzhou University of Technology//School of Computer and Communication, Lanzhou, China
Email: ruililut@126.com

Changfeng Li
Lanzhou University of Technology//School of Computer and Communication, Lanzhou, China
Email: changfengli2013@sina.cn

***Abstract*—Excessive number of Haar-like features and the complex threshold calculation of covariance matrix feature are two key issues of Adaboost face detection. In this paper, an efficient feature named covariance feature is proposed. The novel method divides the face image into several regions and it calculate covariance feature of any two regions. Then optimal weak classifiers will be picked out by Adaboost algorithm and they will be composed to a strong classifier. The experiments result in MIT+CMU data sets shows that the feature extraction times of the novel method is slightly slower than covariance matrix feature. However, the feature threshold is obtained much faster than covariance matrix feature, leading the significant reduction of the training time of Adaboost algorithm. Comparing with the Haar-like feature, the detection rate and speed improved obviously.**

***Index Terms*—face detection, covariance feature, Adaboost, feature extraction**

I. INTRODUCTION

Face detection(FD) refers to the process of defining the position, size, posture of face(if it exists) on the input image. As a special case of object detection, face detection problem has attracted extensive attention of researchers in last two decades due to its widely application in various fields. The fields are human-computer interfaces, content retrieval, digital video processing, visual monitoring[1] and so on. In recent years a lot of face detection methods appeared[2-3], and the statistical learning methods has become mainstream in pattern recognition field gradually. In 2001, Viola proposed Adaboost face detection based on Haar-like feature(HLF)[4]. This method can obviously improve the detection rate and speed .

Adaboost is an iterative algorithm and the core idea is training weak classifiers from a training set, then gather some suitable weak classifiers to a strong classifier. The calculation of sample weight is according to classification is correct or not and the overall classification accuracy of last time. Then training the modified new data sets with the next layer. Finally grouping the best classifiers of each layer into the final decision classifier. Using Adaboost classifier can eliminate some unnecessary features and focus on key data training.

HLF calculate face feature by integral image which can boost computational efficiency. However, HLF is a coarse feature, it is sensitive to edge, line and can only describe specific graph structure. Recently, covariance matrix feature(CMF) is proposed[5], it can reflect the intrinsic relevance of image pixels, and overcome the shortcomings of HLF . Nonetheless, the detection time is three times slower than HLF because the threshold calculation is too complex.

Therefore, this paper proposed the covariance feature (CF) to improve detection speed based on CMF. Comparing with CMF, the detection speed was greatly improved and CF is about CMF detection rate. Comparing with HLF, the face detection rate was improved and elapsed time was less than HLF. This novel feature is concerned with two regions instead of matrix feature of four regions. It vastly reduces features and simplify the calculation of matrix feature threshold, thereby it improve speed of Adaboost face detection .

II. RELATED WORKS

In classical Adaboost algorithm the weak classifier used Haar-like features which named after similarity to Haar wavelet. HLF is a kind of simple rectangle feature and it extract face feature fast. HLF value refers to the difference of gray level sum between the black rectangle and white rectangle on the test image. It reflects the differences of local image gray.

Figure 1 gives five basic templates of HLF feature. We can matching the feature template in every position of detection and achieve a HLF feature. The matrix in image window should meet the S-T criteria[6]. The formula of quantity calculation is as the follow:

$$\Omega_{(s,t)}^{n}=\left(\left\lceil\frac{m}{s}\right\rceil+\left\lceil\frac{m-1}{s}\right\rceil+\cdots\left\lceil\frac{s+1}{s}\right\rceil+1\right)\bullet\left(\left\lceil\frac{m}{t}\right\rceil+\left\lceil\frac{m-1}{t}\right\rceil+\cdots\left\lceil\frac{t+1}{t}\right\rceil+1\right) \quad (1)$$

If we just use this five feature templates as show , there are 78460 matrix features in 20x20 pixel image. However, we may use more expansion HLF templates[7] in practical applications. The detector with slightly better performance may contain thousands of HLF.

doi:10.4304/jcp.9.5.1077-1082

Figure 1. Five basic kinds of HLF template

HLF can describe the differences of local gray, most of the time it can achieve a fine detection result. But HLF describe the structure coarsely, it is sensitive to certain graph structure, such as edges, lines,etc. Figure 2 embodies the defect of HLF, Figure 2(a) shows two basic HLF and in Figure 2(b) the two features can match the facial image ideally, thus the face image can detect by the rectangle feature. However, the image in Figure 2(c) may greatly mistaken for a face because the features also can match the image.

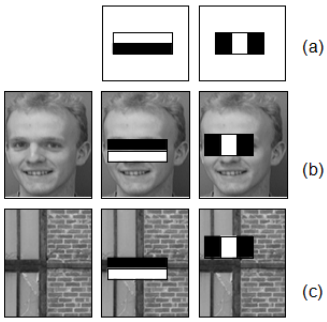


Figure 2. Example of matching HLF in two typical images

With the above problem, the CMF was proposed by Hua et.al.[5]. CMF itself is a covariance matrix[8], each value of matrix is a covariance between each sample subregion. The detection process is as follow, firstly, the covariance between each sample are calculated, then combine a portion of covariance value into covariance matrix according to some rule. Each covariance matrix is a CMF.

Hua divide the image into 5×5 same size matrix region as shown in Figure 3. Taking any four different sub-regions and extract the CMF, 12650 covariance matrix features are extracted if we don't consider the positional relationship. Every CMF is a feature template, The appropriate weak classifiers are selected by Adaboost algorithm, then a face classifier is combined by the weak classifiers. As displayed in Figure 4, we select four typical regions p(1,2), p(2,4), p(4,3), p(5,4) and calculate the covariance between each region pixel, then it constitute a 4×4 covariance matrix, it also call a CMF. The matrix value is shown in Table 1.

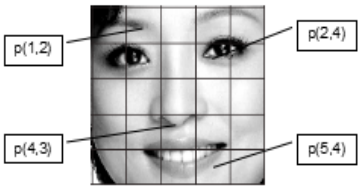


Figure 3. Dividing face into 5x5 regions

Figure 4. Face sub-regions

TABLE 1.
VALUE OF A CMF AS THE SUB-REGIONS OF FIGURE 4 SHOWN

866.8	-758.1	-231.4	23.15
-758.1	3905.8	14.32	603.5
-231.4	14.32	494.7	-176.6
23.15	603.5	-176.6	224.3

CMF need calculate the threshold of weak classifier, a CMF corresponds to a week classifier. The calculation of CMF threshold is complex, firstly, calculating the Euclidean distance between feature matrix s and template matrix t, the Euclidean distance is as follow:

$$d_j^a(s,t) = sqrt\left[\sum_{i=1}^{n}(s_j^a[i]-t[i])^2\right] \qquad (2)$$

In the formula(2), j is the current CMF and a is the current sample, $s_j^a[i]$ is the jth covariance value of sample a, t[i] is the corresponding template value. Then calculating the distance mean value of every CMF, the mean value is the final threshold for judgment. Variable k is the sum of samples. The threshold formula of feature j is as follow:

$$\mu(j) = \frac{1}{k}\sum_{a=1}^{k} d_j^a(s,t) \qquad (3)$$

After the threshold of a test sample is achieved, if the Euclidean distance between covariance matrix and template matrix is less than the threshold, the sample is considered to be a face,otherwise it considered a non-face. The weak classifier is as follow:

$$h_j(x) = \begin{cases} 1, f_j(x) < \mu(j) \\ 0, otherwise \end{cases} \qquad (4)$$

The process of training CMF weak classifiers is complex, furthermore, a 5×5 image contains 12650 CFMs, the Adaboost algorithm based on CMF leads to a time-consuming result.

There are 78460 HLFs in a 20×20 pixel image, it is time-consuming even we calculate the feature by integral image. There are 12650 CMFs in a 5×5 sub-region image and the CMF threshold calculation is complex, the entire training process is more time-consuming than HLF. Focus on these problems, this paper proposed the covariance feature based on CMF, it can reduce the number of feature and overcome the time-consuming problem, and remain the detection rate steady. In the experiment we use the 8×8 sub-region covariance feature and extract only 4064 features, the number of feature is far less than HLF and CMF. The feature threshold calculation is similar to HLF, it is more simple and faster than CMF threshold.

CMF and CF all have to calculate the covariance between each sub-region in common, they all can describe correlative degree of each sub-region. The difference of CMF and CF is as follows:

(1)The representation of CF value is a two-relative region covariance, but the representation of CMF value is Euclidean distance between feature matrix and template matrix.

(2)CF only need to calculate a covariance matrix to express all the CF value in each test image. However, a covariance matrix is a CMF. Four relative sub-regions covariance matrix combine into a CMF.

III. ADABOOST FACE DETECTION BASE ON CF

A CF is a covariance coefficient[9], and it is a low dimensional feature that can calculate easily. On a given W×W pixel image, dividing sample image into d×d same size matrix sub-region, then we can calculate the covariance coefficient C(x,y) of any two matrix. x,y represent the ith, jth matrix respectively. The number of pixels in each matrix is n, and $\mu(j)$ represent the mean value of all ith matrix pixels:

$$\mu(i) = \frac{1}{n}\sum_{k=1}^{n} f_k(i) \qquad (5)$$

The covariance coefficient is similar to covariance formular and defined to be:

$$C(i,j) = \frac{1}{n-1}\sum_{k=1}^{n}(f_k(i)-\mu(i))(f_k(j)-\mu(j)) \quad (6)$$

In the formula (6), the ith, jth sub-region is interpreted as a two dimension random variable and evaluate the covariance coefficient C(i,j). In virtue of the covariance symmetry and the covariance of sub-region with itself is insignificant, we can extract $[(d^4+d^2)/2]-d^2$ significative CFs on d×d matrices. For illustration purposes, as shown in figure 5, dividing a test image into 3×3 sub-region. We can obtain a covariance matrix and the covariance of any two different coordinate region is a CF. The 3×3 specification only generate 36 CFs as the table 2 shown. For example, p(1,3) and p(3,2) represent the 3rd and 8th sub-region, and Via formula (6) we can calculate the (3,8) coordinate value of the CF matrix is 14.76 .

TABLE 2.
3×3 MATRIX OF CF VALUE

1575.2	140.56	1268.8	-182.7	-85.57	-51.36	-52.16	-107.0	-94.62
140.56	364.77	241.23	20.62	101.78	2.66	343.67	79.82	-60.09
1268.8	241.23	1819.1	-45.76	182.63	-57.37	328.69	14.76	57.39
-182.7	20.62	-45.76	277.02	136.82	143.35	176.14	133.89	104.17
-85.57	101.78	182.63	136.82	283.69	65.49	255.91	96.61	8.20
-51.36	2.66	-57.37	143.35	65.49	164.15	154.25	65.78	-21.84
-52.16	343.67	328.69	176.14	255.91	154.25	1148.0	162.27	-73.69
-107.0	79.82	14.76	133.89	96.61	65.78	162.27	307.38	135.62
-94.62	-60.09	57.39	104.17	8.20	-21.84	-73.69	135.62	463.62

Figure 5. 3×3 cutting face image

The experiment result showed that the extracting speed of CF is faster than CMF with the same specification, but there is a large gap with the desired detection. Boosting the detection performance is necessary. The purpose of Adaboost algorithm is improve the classification accuracy of any given learning algorithm, the next process is discussing how to improve the performance of CF classifier through Adaboost algorithm.

The example of 3×3 specification is too rough and can't primely describe the internal relation of face sub-region. To refining the extracting feature, we choose the 8×8 specification to extract the CF in the experiment. Figure 6 shows the 8×8 diving face, choosing any two sub-region and calculating the CF value C(x,y) by formula (6), we can extract 4064 CFs $C_n(1,2,\cdots,4064)$ in total. Through the Adaboost algorithm, we can training a appropriate number of CFs from the 4064 CFs that constitute a strong classifier.

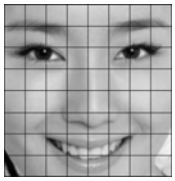


Figure 6. 8×8 cutting face image

Adaboost is a self-adapting Boosting algorithm, the training process is greatly associated with sample probability distribution. When the sample classified correctly, reduce the sample weight, and when the sample classified falsely, increase the sample weight while the classifier receive more attention. Then combining the optimum weak classifier of each round into a strong classifier and output the value of judgment. So long as the classify accuracy of CF classifiers is greater than 50% and training enough CF weak classifiers, we can obtain a strong classifier which the error rate approach 0. The form of the jth weak classifier is given as follows:

$$h_j(x) = \begin{cases} 1, p_j f_j(x) < p_j\theta_j \\ 0, otherwise \end{cases} \qquad (7)$$

$h_j(x)$ is the judge value of sample *x*, the threshold offset is p_j, the value only can be ± 1, the threshold is θ_j, $f_j(x)$ is the jth CF value of sample *x*.

Following are the procedure of Adaboost classifier learning:

(1)Given training example $(x_1,y_1),(x_2,y_2)\cdot\cdot(x_n,y_n)$ where y_i =0,1 for negative and positive examples respectively.

(2)Initialize weights $\omega_{1,i} = 1/2m, 1/2t$ for y_i=0,1 respectively, where *m* and *l* are the number of negatives and positives.

(3)For $t=1,\cdots T$

(a)Normalize the weights, $\omega_{t,i} = \omega_{t,i} / \sum_{j=1}^{n}\omega_{t,j}$, so that ω_t is a probability distribution.

(b)For each feature *j*, training a classifier h_j which is restricted using single feature. The error is

evaluated with respect to ω_t, calculate the error rate: $\varepsilon_j = \sum_{j=1}^{n} \omega_{t,j} \left| h_j(xi) - y_i \right|$.

(c)Choose the classifier h_t, with the lowest error ε_t.

(d)Update the weights: $\omega_{t+1} = \omega_{t,i} \beta_t^{1-e_i}$, where e_i =0 if example x_i is classified correctly, e_i =1 otherwise, and $\beta_t = \varepsilon_t / (1-\varepsilon_t)$.

(4)The final strong classifier is:

$$h(x) = \begin{cases} 1, \sum_{t=1}^{T} \alpha_t h(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0, otherwise \end{cases} \quad (8)$$

Where $\alpha_t = \log \frac{1}{\beta_t}$.

We chose 356 classifiers of higher detection rate from 4064 weak classifiers after training, the detection sub-regions are distribute mainly beside eyes, eyebrows, nose, mouth. These region can distinguish the face and non-face effectively.

Figure 7. 6 CFs which are easiest to distinguish.

Below is the detection algorithm:

Dividing the test image $k(1,2,\cdots,n)$ into 8×8 sub-region and mark for 1, then detect the image in each layer classifier of the cascade classifier[10-11] orderly.

For $i = 1:T$ //the ith layer classifier

For j=1:M//the jth weak classifier of ith layer

Calculating the CF value of two sub-region appointing by order jth, from the jth classifier we can obtain the judgment value of the current sample, and the value is H_{ij}.

End

The weight of jth weak classifier in layer ith is α_{ij}.

$$h(x) = \begin{cases} 1, \sum_{j=1}^{M} \alpha_{ij} H_{ij}(x) \geq \frac{1}{2} \sum_{j=1}^{M} \alpha_{ij} \\ 0, \ otherwise \end{cases} \quad (9)$$

Mark 0 if $h(x)$ =0 and eliminate the sample which is unnecessary to detect on the next layer. Otherwise, mark 1 and detecting sequentially.

End

After T layers selection, the samples which are marking 1 are estimated faces and remain are non-faces.

Ⅳ.EXPERENENT RESULTS

The training samples chose from MIT face test set, we selected 2106 faces and 2781 non-faces for training CF classifiers. In the pre-process, we normalized the sample into 40×40 pixels, and enhanced the image by histogram equalization. We divided the normalized image into 8×8 specification and every sample image can extract 4064 CFs. The test set select from MIT+CMU, there are 1236 face samples and 1276 non-face test samples. Furthermore, face image from Internet were used, and preprocessed the image with skin detection method[12]. The operating environment is AMD 3.0GHZ PC, Windows XP operating system. Our experiment compared the 20×20 pixels HLF, 5×5 specification CMF and 8×8 specification CF.

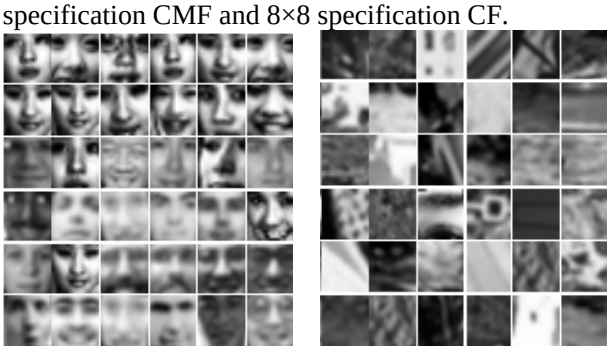


Figure 8. Part of face and non-face samples from MIT set

The experiment results of feature extracting are as follow in table 2. A large number of HLF lead time-consuming extraction. Though the covariance calculation is more complicated, the 5×5 specification CMF only need calculate covariance value 300 times per sample image, then combine any four different value into a CMF. There are 12650 combinations and only spend 143s. The 8×8 specification CF calculate the covariance values 4064 times and the calculation complexity of covariance value is same as the CMF. CF spend 473s in feature extraction. The speed of CMF extraction is slightly higher than CF. However, the speed of CMF threshold calculate is too more complex than CF. It is the biggest defect of CMF and this problem lead to a time-consuming detection process.

TABLE 2.
EXTRACTING TIME-CONSUMING OF THREE FEATURE

Feature	Number of feature	Time-consuming(s)
HLF	78460	1584
CMF	12650	143
CF	4064	473

In the testing process, , the detection rate is not always increscent while the weak classifiers increase. Comparing the three curves in the figure 9, CMF reach the maximum firstly, CF second and HLF last. When the weak classifiers increase unceasingly, HLF detection rate will declines obviously because the sample weight has become distorted and HLF has degraded. Adaboost algorithm based on CMF and CF can avoid the degradation preferably, CMF and CF enhance the detection rate and lower the false positive rate.

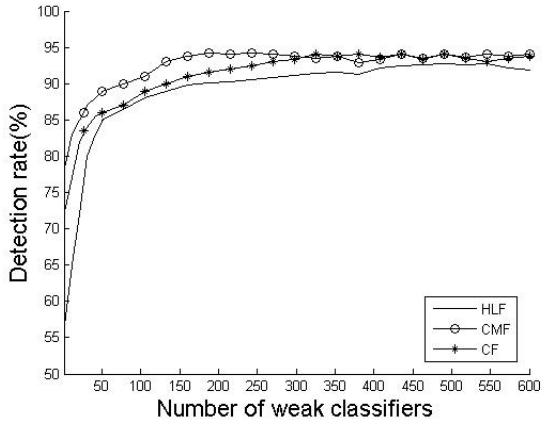


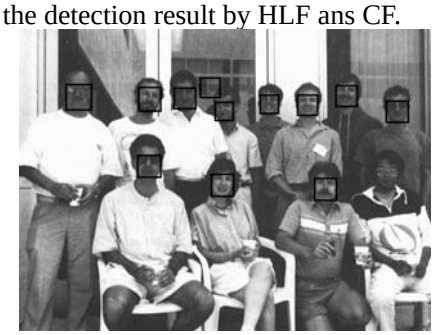
Figure 9. The relation of detection rate and number of weak classifiers

The CF can overcome the shortcomings of time-consuming as shown in table 3, the time spending of CF training and detection is 2562s and significantly less than CMF and HLF. The CF detection rate is 93.7% which is nearly CMF and higher than HLF. After tests, using 356 CF weak classifier can reach the best detection rate.

TABLE 3. DETECTION RATE AND TIME SPENDING OF WHOLE PROCESS

Feature	Detection rate(%)	Number of weak classifier	Average time spending(s)
Haar-like	92.5	528	13647
CMF	93.6	237	37459
CF	93.7	356	2562

The following picture is from CMU test set, and we can see the detection result by HLF ans CF.


(a) using HLF

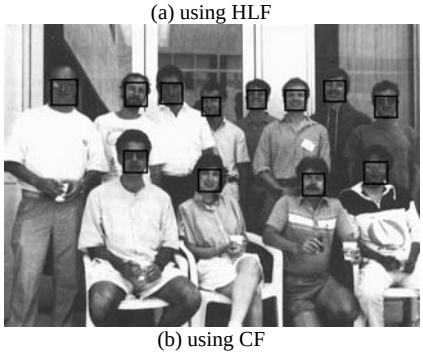

(b) using CF

Figure 10.Face detection result

V .CONCLUSIONS

In this paper, the covariance feature is proposed. It combines the advantages of traditional Haar-like feature and covariance matrix feature. This paper also related the combination of CF and Adaboost algorithm. The test results on MIT+CMU test set showed CF can greatly improve training and test speed while the detection rate is similar to CMF. In recent years, many researches[13-14] focus on multi-view face feature and the research results showed these improvements can detect non-front face preferably. In the future, the multi-view CF is a novel direction worthy of studying.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China(No. 61263019), and supported by the Natural and Science Foundation of Gansu Province(No. 1208RJZA212), and supported by the scientific research project of Gansu Finance Department(No. 1114ZTC144).

REFERENCES

[1] L.H. Liang, H.Z. Ai, G.Y. Xu and B. Zhang, “A survey of human face detection,” *Chinese Journal of Computers*, vol.25, No.5, pp.132-137, May 2002.

[2] M.H. Zhang and L.H. Gang, “A Detection Method of Driver’s Face Orientation Based on Visual Cues and SVM,” Journal of Software, Vol 8, pp924-931, Apr 2013.

[3] Y. Ouyang and N. Sang, “Robust Automatic Facial Expression Detection Method, ” Journal of Software, Vol 8, pp1759-1764, Jul 2013.

[4] Viola P, Jones M, “Rapid object detection using a boosted cascade of simple feature,” in Proceedings of the IEEE *Computer Society Conference on Computer Vision and Pattern Recognition*, pp.511-518, 2002.

[5] Y. Hua, H.J. Peng and J.L. Gu, “Face detection method by boosting covariance feature,” *Computer Engineering and Applications*, Vol.46, No.18, pp.186-189, 2010.

[6] B. Wu, C. Huang, H.Z. Ai and S. Lao, “A multi-view face detection based on real adaboost algorithm.” *Journal of computer research and development*, Vol.42, No.9, pp.1612-1621. 2005.

[7] Lienhart R, Maydt J, “An extended Set of Haar-like Features for Rapid Object Detection,” [C]//Proc. Of ICIP’02. New York, USA:[s.n.], pp.155-162, 2002.

[8] Moakher M, “A differential geometric approach to the geo-metric mean of symmetric positive-definite matrices,” *SIAM J Matrix Anal Appl*, Vol.26, No.3, pp.735-747, 2005.

[9] Tuzel O, Porikli F, Meer P, “Region covariance: A fast descriptor for detection and classification,” //Proc 9th *European Conf on Computer Vision*, Graz, Austria, pp.589-600, 2006.

[10] Viola P, Jones M, “Robust real-time face detection,” *Computer Vision*, Vol.57, No.2, pp.137-154, 2004.

[11] W.H. Li, Y.F. Lin, B. Fu, M.Y. Sun and W.T. Wu, “Cascade Classifier Using Combination of Histograms of Oriented Gradients for Rapid Pedestrian Detection,” Journal of Software Vol 8, pp.71-77, Jan 2013.

[12] Erdem C, Ulukaya S, Karaali A and Erdem A, “Combining Haar Feature and skin color based classifiers for face detection,” In *Acoustics.Speech and Signal Processing (ICASSP)*, 2011 IEEE International Conference, pp.1497-1500, may 2011.

[13] X.K.. Liu, X.H. Sun and Y.X. Zhou, “Multi-angle Face Detection Based on New Haar-like Feature,” *Computer Engineering*, Vol.35, No.19, pp.195-197, 2009.

[14] Z.X. He, X.Q. Ding, C. Fang and D. Wen, "Multiview face detection based on LBP and CCS-Adaboost," *Journal of Zhejiang University(Engineering Science)*, Vol.47, No.4, pp.622-629, 2013.

Rui Li was born in Gansu Province, China in 1971. She received the Master of Science degree in computer application in 2005. She is currently a professor at school of computer and communication of Lanzhou University of Technology. Her research interests include pattern recognition, picture processing and digital watermarking.

Changfeng Li was born in Hubei Province, China in 1991. He received the B.E degree in 2011.He is currently a postgraduate student at School of Computer and Communication of Lanzhou University of Technology. His research interests include pattern recognition and picture processing.