MATEC Web of Conferences **44**, 01001 (2016)
DOI: 10.1051/matecconf/20164401001

RESEARCH for face detection on improved algorithm of AdaBoost

Hong CHEN

Department of Engineering Technology and Information, Guang'an vocational and technical college, Guangan 638000,China

Abstract. Face detection is basic research in computer visual field, it has important application value in the fields of cameras and surveillance and Automatic face recognition. Because traditional algorithm of AdaBoost has the problem that it needs more features and the detection speed is slow when it is used to detect face, so a improved algorithm of AdaBoost is presented. The experiment results shows that compared with traditional algorithm of AdaBoost, lower features are used in the improved algorithm and higher accuracy rate can be get and the speed of detection is improved Significantly.

keywords: face detection; algorithm of AdaBoost;double threshold

1 introduction

With the quick development and wide application of computer,face detection is more and more attented by Science and technology personnels, it is widely used in the fields of Real-time monitoring, identification and Image database retrieval.Face detection is the key step in the system of face automatic recognition,certain strategy is used to search some of the pictures or videos,and as a result, it can be judged that if there is face.If the faces exit then their position,size and pose will be ensured[1].

Viola and his friends[2] used algorithm of AdaBoost to establish cascading face detector,the method is important progress on face detection.The basic ideology is on the training set that is given,important classification characteristics are selected to expanse the weak classifier through repeated training,and finally the strong classifier can be get afer the weak classifier through a course of linear combination.After each round of training, AdaBoost must resolve two core problems that how to update sample weight and how to choose appropriate weak classifier to constitute strong classifier. The traditional AdaBoost exists a certain problem that the algorithm weight update rule will be handled well when difficult samples appear.But sample weight becomes very big if very few difficult samples exist,as a result,the weigh will be distorted and the trainning will be stop.The search strategy of AdaBoost is ordered to go forward, local optimal principle is used to every iteration,but the weak classifier and coefficient that constitute strong classifier is not the most optimal,as a result,the face detection effect is reduced.

2 improved AdaBoost

Additional AdaBoost has some insufficient,a improved AdaBoost is proposed in the article,the improved AdaBoost changes single threshold value weak classifier to double threshold value weak classifier.There are many factors are considered in the course of selecting characters,for example,the classification ability of characteristics of their own. Information relevance is used to measure the degree of correlation among

Article available at http://www.matec-conferences.org or http://dx.doi.org/10.1051/matecconf/20164401001

characters and redundant information is reduced among characters.

Supposing f1,f2,⋯,fm are selected in the course of selecting characters. In order to avoid redundant information among selected characters and will be selected, relevance should be evaluated from selected characters and will be selected. Supposing x is a character that will be selected,the relevance of R(x) with character which is selected is as follows:

$$R(x)=\max(x,fi)\ i\text{-}1,2,\cdots,m \quad (1)$$

A appropriate threshold value of ε is assumed,if R(x)>ε ,it will be considered that there exists very big relevance between x and selected character,new information will not generate and x can't be selected as character.On the contrary,the relevance between x and selected character is small and it can generate enough new information,x can be selected as character.

Contraposing the defects of traditional AdaBoost, some improvements are arised in the article.Improved algorithm main bring in character relevance and is used to chatacer selection,the method reduces the redundant information among selected characters to minimum. A certain number of sample(label to Positive or negative) are selected as training set,then improved Adaboost is used to be feature selection. The following steps are used to set up strong classifier:

(1) Sample training set{(x1,y1),⋯,(xn,yn)} is input,n is number of sample,xi is NO. of i in the Sample training set,yi is the category lable(1 or 0,their mean are positive samples or Negative samples);{h1(x),h 2(x),⋯,hk(x)} has k number of weak classifier,each weak classifier has a character that will be choosed, t is the number of characters which will be choosed, ε is the Correlation threshold.

(2)Sample weigh is initialized:

$$w_{1,j}=1/n. \quad (2)$$

(3)Character is choosed and classifier is established:if the number of character which is choosed less than t then the circulation occurs:

①Sample weigh is normalized:

$$W_{t,i}\leftarrow p_{t,i}=\frac{W_{t,i}}{\sum_{j=1}^{N} W_{t,i}} \quad (3)$$

Weak classifier which is h(j) is designed for each character:

$$h(j)=\begin{cases}1 & \delta 1 < fj(x) < \delta 2\\ 0 & \text{others}\end{cases} \quad (4)$$

In the formula (4),fj(x)is the character value of sample. Weighted classification error is:

$$\varepsilon_t=\sum_{i=1}^{N} W_{t,i}|h_t(x_i)-y_i| \quad (5)$$

The minimum ofε_t and δ1 and δ2 should be computed.

②Among the all weak classifiers,the h*m* whose Weighted classification error is minimum is choosed,it's character value is f*m*.If M(f*m*)> δ then it will exit the current cycle,otherwise f*m* is the character of the current cycle.

③Output:

$$H(x)=\begin{cases}1 & \sum_{t=1}^{T} a_t h_t(x) \geq \frac{1}{2}\sum_{t=1}^{T} a_t\\ 0 & \text{others}\end{cases} \quad (6)$$

In the formula (6),a*t*=log(1/β_t).

X is extracted from image that is detected input into the classifier and get discriminant set,the result in the discriminant set is voted and the final classification result is get.

3 Experimental environment and the simulation results

In order to verify the improved algorithm's feasibility,the training sample relevant data which is choosed when the simulation result is carryed through:there are 3486 pieces of face sample. There are 1686 pieces of positive face images are selected from Feret face database,these face images have been Image processed.There are 5590 piecesof image that are not face,they are from Internet and 45898 Harr features are extracted from each image.There are 4605 pieces of face images and 6415 pieces of images that are not face when test,these images are from Internet and the size are set to be 240×320 pixel.The test environment are Visual c++.NET,4G internal storage, Inter(R) Core(TM) i5-3230M 2.60 GHz.True Positive Rate(TPR) and False Positive Rate(FPR) are used to be test standards.The test result is follows in figure 1 and figure 2.

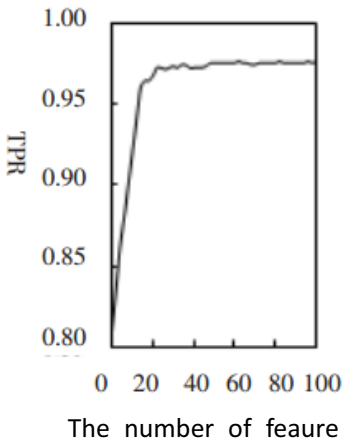

Figure 1. the ratio that face divide to be non-face

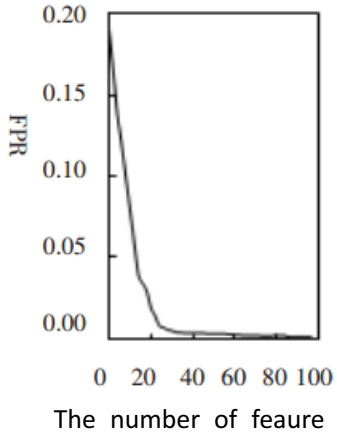

Figure 2.the ratio that non-face divide to be face

It shows from the figure 1 that good effect is get if the selected feature number reach 25,the reason is that the accuracy rate is stable afer feature number reach 25.

The same experimental samples that in the simulation results are used to be compared,they direct at the algorithm in the article,traditional AdaBoost algorithm, the algorithms in reference[3]and in reference [4].Same training set is used and Harr feature is extracted,the four method are used to trained and same testing set is used to test.The threshold value of feature relevance is ε =0.28.The judgement of the four algorithms's property is described by ROC curve.It is assumed that sample is described by two method: positive(face) and feminine(non-face). Positive sample is named true positive if it is classified to positivity, otherwise it is named false feminine. Negative sample is named true feminine if it is classified to Negativity ,otherwise it is named false positive.ROC curve of four al-gorithms is in figure 2, false positive ratio is expressed by x axis,true positive ratio is expressed by y axis.The data of feature number,TPR,FPR,test time is given in table 1 when the four algorithm s's property is stable.

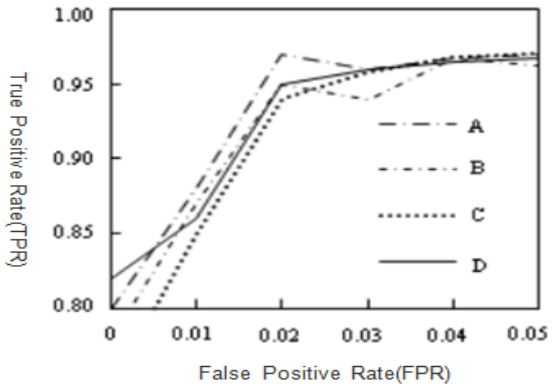

Figure 3. ROC curve of four algorithms

Table 1.experimental result comparation of four algorithms

Algori-thm	AdaBoost	Reference [3]	Reference [4]	the algorithm in the article
Feature number	90	60	50	25
True Positive Rate(%)	97.0	96.8	97.0	97.1
False Positive Rate(%)	2.6	3.3	2.7	2.4
Average time(s)	0.309	0.209	0.168	0.094

It shows from figure 3 that the detection performance are same to four algorithms,but it shows from table 1 that the four algorithms need feature number are 90,60,50,25 if stable detection accuracy is reached. The algorithm in the article needs minimum of feaure,so the detection rate is fastest.

4 Conclusion

Traditional AdaBoost algorithm of face detection has some defection,so improved algorithm of AdaBoost is created.The simulation result shows that improved algorithm of AdaBoost can reduce the feature number which is used to detection,so the detect rate is improved Significantly,the accuracy rate of face detection improved obviously.

Referance

1. Yang Ming-Hsuan David J Kriegman, Narendra

Ahuja.Detecting Faces in Images ASurvey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2002,**24** (1):34~58

2.P Viola M Jones Rapid Object Detection using a Boosted Cascade of Simple Features [C]. IEEE Proceedings of the Computer Vision and Patton Recognition Conference December 11~13, Hawaii USA,2001.

3.Tian jian Liu.the improved AdaBoost algorithm that base on character selection of entropy[J].Min jiang academic journal, 2009,4,**30**(2): 60~64

4. Rongye Liu. An Unsupervised Feature Selection Algorithm:Laplacian Score Combined with Distance -Based Entropy Measure, Third International Symposium on Intelligent Information Technology Application, 2009. Nov.21~22 2009, 65~68