

Study on the Method of Face Detection Based on Chaos Genetic Algorithm Optimization AdaBoost Algorithm

Yu Chai¹, Yang Yang^{2,*}

^{1,2}School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an, 710054, China

Abstract. Aiming at the problems that the traditional AdaBoost algorithm has complex feature computation, long training time and low detection rate, a method of face detection based on chaos genetic algorithm optimization adaBoost algorithm was proposed. Firstly, this algorithm uses the image color segmentation for coarse screening on the face image, in order to determine the human skin area. Secondly, the adaptive median filtering is applied to denoise the face image to improve the quality of the face image. Finally, the chaotic genetic algorithm is used to optimize the AdaBoost algorithm to achieve higher detection rate and detection speed. Compared with the traditional AdaBoost algorithm, the experimental results showed that the face detection method based on chaos genetic algorithm optimization AdaBoost algorithm proposed in this paper has a significant improvement in detection rate and detection speed.

1 Introduction

Face detection uses a certain strategy or a set of image sequences to search for any given image to determine the position and region of the face, to detect the presence of a face in a variety of different images or image sequences, and to determine the number of faces and the process of spatial distribution [1]. Face detection is a key technology of face information processing which has a very broad application [2], such as: face recognition; face tracking; pose estimation; facial expression recognition; image retrieval and digital video etc [3].

At present, face detection methods are mainly based on the feature detection method and statistical based detection method [4]. Among them, the feature-based detection method is intuitive and easy to implement, but there are great limitations. However, the statistical detection method is not easy to implement, but it has better robustness and universality [5]. In this paper, the AdaBoost algorithm based on statistical detection method is used as the core algorithm of face detection.

Firstly, the skin color segmentation is used to screen the face image. Secondly, the adaptive median filter is used to denoise the face image. Finally, the chaotic genetic algorithm is used to optimize the AdaBoost algorithm. The method reduces the detection time and improves the detection efficiency.

2 Face detection

2.1 Image color segmentation

In order to reduce the search time of AdaBoost algorithm, this paper adopts the color segmentation of the image to

rough screen the face image, and then determine the skin area of human body, thus it can avoid the traversal search of AdaBoost algorithm.

Because the skin color has better clustering characteristics in YCbCr color space, the conversion relation with RGB color space is simple, and there is no singular point [6], so YCbCr color space is chosen as the skin segmentation model. In the YCbCr color space, the Y component represents brightness information, and the Cb component represents the blue chromaticity information, and the Cr component represents the red chromaticity information. The transformation formula of RGB color space and YCbCr color space:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.168 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.0813 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

According to the distribution statistics:

$$\begin{cases} 77 \leq Cb \leq 127 \\ 133 \leq Cr \leq 173 \end{cases} \quad (2)$$

If the pixels fall within this range, the pixels of the skin color are considered. The raw image and skin segmentation results are shown in Figure 1 (a) and (b).

2.2 Image denoising

Normally, the image information obtained by the image acquisition device contains various noises and distortions [7]. Therefore, after the segmentation of the skin color, we need to denoise the face image so as to improve the quality of the face image.

Because of noise and distortion, the image is distorted and degraded, so denoising is the basic step in

* Corresponding author: 18792906858@126.com

image processing. In the denoising process, filtering is necessary in face detection [8]. The commonly used image filtering methods include smoothing filtering and sharpening filtering, etc. smoothing filtering is more widely used in face image denoising processing. In this paper, an adaptive median filtering algorithm is used to denoise the image. The denoising results of the face image are shown in Figure 1 (c).

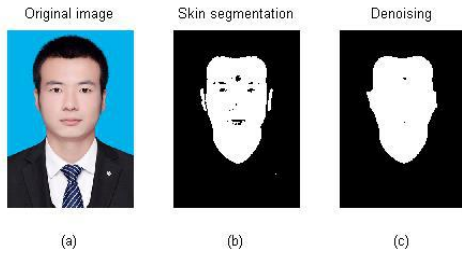


Fig. 1. Skin segmentation and denoising of face images

2.3 AdaBoost algorithm

Freund and Schapire improved the Boosting algorithm in 1995, named it AdaBoost. The algorithm is efficient and does not need to know any prior knowledge about the weak learning algorithm. Paul Viola and Michael Jones in 2001 proposed a face detection algorithm based on AdaBoost, this method combines several strong classifiers into cascade classifiers. When the sub-window is recognized, the sub-window is sent to the subsequent classifier to continue processing when the classifier of the previous stage is judged to be positive, otherwise it is judged as negative directly, and the sub-window will be judged to be positive only if all classifiers are judged to be positive [9-10].

AdaBoost algorithm specific description follows [11]:

1) The input is N training samples.

$(x_1, y_1), \dots, (x_i, y_i), \dots, (x_N, y_N)$, where $i = 1, 2, \dots, N$, $x_i \in X$, $y_i \in Y = \{-1, 1\}$ are non-face samples and face samples respectively.

2) Initialize weights, $\omega_{1,i} = \frac{1}{2m}, \frac{1}{2l}$, where m, l are training sample set of non-face samples and face samples respectively.

3) Start iteration loop, $t = 1, 2, \dots, T$.

a. Normalized weight $\bar{\omega}_{t,i}$.

$$\bar{\omega}_{t,i} = \frac{\omega_{t,i}}{\sum_{j=1}^n \omega_{t,j}} \quad (3)$$

where n is the characteristic number.

b. For each feature j of the weak classifier h_j , the error rate of the current weight is calculated,

$$\varepsilon_j = \sum_{i=1}^n \bar{\omega}_{t,i} |h_j(x_i) - y_i|.$$

c. The decision function h_t corresponding to the minimum error rate ε_t is used as the weak classifier obtained by this iteration.

d. Judge the weight of each training samples according to the decision function h_t , that is $\omega_{t+1,i} = \omega_{t,i} \beta_t^{1-e_i}$. Among, if h_t judge the sample i correctly, $e_i = 0$; otherwise, $e_i = 1$, $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$.

4) Generating a strong classifier.

$$h(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq 0.5 \sum_{t=1}^T \alpha_t \\ 0 & \text{Others} \end{cases} \quad (4)$$

where $\alpha_t = \log_2 \frac{1}{\beta_t}$.

3 Optimization of AdaBoost algorithm based on chaos genetic algorithm

3.1 Improved genetic algorithm

Genetic algorithm is an adaptive global optimization probability search algorithm [12], which mimics the genetic and evolutionary processes of organisms in the natural environment. The idea of solving the genetic algorithm is to map the solution space of the problem into the coding space, and use the fitness function value of the individual as a measure of the quality of the individual. In the process of evolution, the new optimal individuals are generated by continuous selection, crossover and mutation, and finally the global optimal solution is decoded [13].

Genetic algorithm is widely used in many fields, because of its relatively simple and low demand characteristics of the objective function, but the search for the optimal individual is more difficult, the slow speed of the search in the vicinity of the global optimal solution, and may fall into the local optimal solution, slow convergence speed and premature convergence is also related to the lack of. Therefore, in order to improve the above shortcomings of genetic algorithm, a new algorithm combining genetic algorithm with other algorithms is proposed.

The theoretical support of chaotic optimization algorithm is the ergodic property of chaos, so it is a global optimization algorithm. Chaos is a class of random phenomena which appear to be chaotic and exist fine rules in a definite system [14]. Because chaos has the ergodic property, it is better to use chaotic variables to optimize the search than blind random search, and prevent the evolutionary algorithm from being trapped into the local optimum. Chaos optimization algorithm and genetic algorithm are fused here.

3.2 Chaotic genetic algorithm design

3.2.1 Encoding mode

For function optimization, real number coding is the most effective, so we choose real coding to determine individual encoding.

3.2.2 Fitness function

Fitness function is the only basis for natural selection, and it is the criterion used to judge the quality of individuals in a group, which is usually transformed by objective function [15]. In this paper, the decision function h_t in AdaBoost algorithm should have the minimum error rate ε_t . That is, the objective function is

$$\min \varepsilon_j = \sum_{i=1}^n \bar{\omega}_{t,j} |h_j(x_i) - y_i| \quad (5)$$

The objective function is to find the minimum error rate ε_j , that is, the reciprocal of the objective function is used to judge the fitness value of the individual. Individuals with small function values have large fitness values and are individually superior. Therefore, the fitness function is chosen as

$$F(\varepsilon_j) = \frac{1}{\varepsilon_j} \quad (6)$$

3.2.3 Initial population

In this paper, Logistic mapping is adopted, and its mathematical model is as follows:

$$x_{k+1} = \mu x_k (1 - x_k), \quad k = 1, 2, \dots \quad (7)$$

Where μ is control parameter, x_k is chaotic variable, k is the ordinal number of the chaotic variable. When $\mu = 4$, $0 < x_0 < 1$, the system has no stable solution and is in chaotic state. It is a full mapping process, so it can improve the genetic algorithm to initialize population diversity.

3.2.4 Selection

It is assumed that there are N populations that produce N offspring individuals after crossing and mutation. According to the fitness function value, from large to young, sub parent generation of $2N$ groups for sorting, take the start of $N/2$ individuals, from $2N$ individuals last selected $N/2$ individuals were matched. In this way, the population is kept converging to the optimal solution, and the diversity of the population is reduced slowly.

3.2.5 Cross variation

In order to simplify the experimental process, the single point crossover operator as the crossover operator in individual encoding string random sets a crossover point, then part of gene transformation; the basic bit mutation as the mutation operator, the mutation probability of

individual gene encoding string on the base of value of gene mutation operation.

In this paper, the flowchart of the chaotic genetic algorithm is shown in Figure 2.

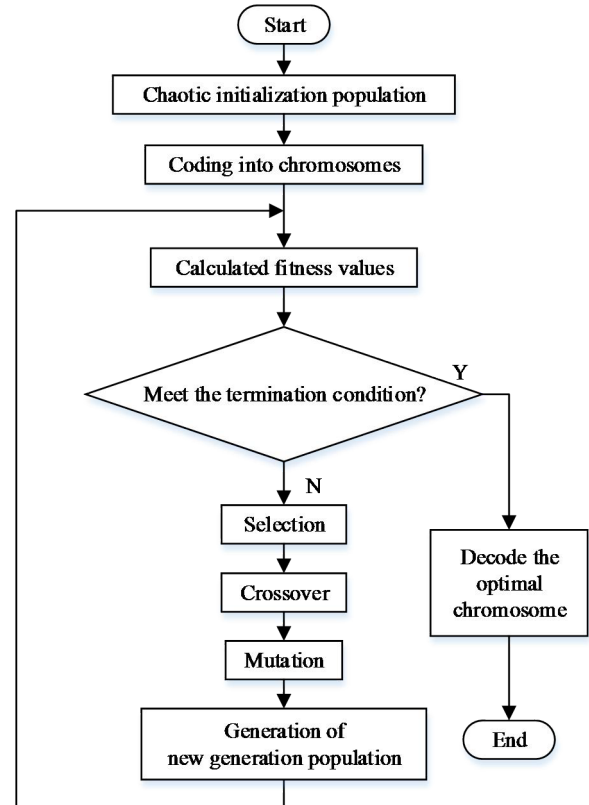


Fig. 2. Flowchart of the chaotic genetic algorithm

3.3 Optimization algorithm flow

The optimization of AdaBoost algorithm using chaotic genetic algorithm is as follows:

1) The input is N training samples.

$(x_1, y_1), \dots, (x_i, y_i), \dots, (x_N, y_N)$, where $i = 1, 2, \dots, N$, $x_i \in X$, $y_i \in Y = \{-1, 1\}$ are non-face samples and face samples respectively.

2) Initialize weights, $\omega_{1,i} = \frac{1}{2m}, \frac{1}{2l}$, where m, l are training sample set of non-face samples and face samples respectively.

3) Start iteration loop, $t = 1, 2, \dots, T$.

a. Normalized weight $\bar{\omega}_{t,i}$.

$$\bar{\omega}_{t,i} = \frac{\omega_{t,i}}{\sum_{j=1}^n \omega_{t,j}} \quad (8)$$

where n is the characteristic number.

b. The genetic algorithm population x_i is initialized using chaotic variables and encoded into chromosomes by real coded encoding.

c. The fitness value of each individual is calculated according to the fitness function of the individual $F(\epsilon_j)$.

d. Determine whether or not the condition has been met. If the condition is reached, the chromosome is decoded so that the decision function h_i corresponding to the minimum error rate ϵ_i can be output. If the conditions are not met, the selection, crossover and mutation operations are performed to generate a new generation of population and return to step c to continue searching.

4) Judge the weight of each training samples according to the decision function h_i , that is $\omega_{t+1,i} = \omega_{t,i} \beta_i^{1-e_i}$. Among, if h_i judge the sample i correctly, $e_i = 0$; otherwise, $e_i = 1$, $\beta_i = \frac{\epsilon_i}{1-\epsilon_i}$.

5) Generating a strong classifier.

$$h(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq 0.5 \sum_{t=1}^T \alpha_t \\ 0 & \text{Others} \end{cases} \quad (9)$$

where $\alpha_t = \log_2 \frac{1}{\beta_t}$.

4 Experimental result

4.1 Experimental environment and data set

Experimental environment: 2.50GHz, 4GB memory, VS2010+OpenCV. In this paper, MIT face sample library is selected. The training sample consists of 2000 face samples and 5000 non-face samples. The test set consists of 1000 face samples and 2500 non-face samples. All the samples are 20×20 pixels. Some of the samples are shown in Figure 3.



Fig. 3. Partial face and non-face samples

4.2 experimental result

The related parameters of the algorithm are as follows: population size 25, maximum reproduction algebra 500, cross probability 0.85, variation probability 0.02. In order to verify the effectiveness of this algorithm, the author uses the traditional AdaBoost algorithm, the AdaBoost algorithm based on skin segmentation and this algorithm three algorithms for comparative analysis.

In this experiment, a strong classifier is trained, in which the training data include the average number of features and the training time required by the classifier. Table 1 presents the training comparison results of the three algorithms.

Table 1. The average number of features and training time comparison

Method	Average characteristic number / peace	Training time / h
Traditional AdaBoost algorithm	225	25.7
the AdaBoost algorithm based on skin segmentation	187	8.3
this algorithm	146	4.9

You can get it from table 1, this algorithm is compared with the traditional AdaBoost algorithm and the AdaBoost algorithm based on skin segmentation, the average number of features required in the training process is much less, the training time is reduced a lot.

In the detection process, a trained strong classifier is used to test the sample set, and the statistical results are checked. Table 2 lists the test results of the test set.

Table 2. Comparison of test results

Method	Detection rate / %	Detection error rate / %
Traditional AdaBoost algorithm	94.67	5.85
the AdaBoost algorithm based on skin segmentation	96.15	3.04
this algorithm	99.26	1.29

From the comparison of the detection results, we can see that the detection performance of the proposed algorithm is improved compared with the other two algorithms. The reason is that this paper uses chaotic genetic algorithm to optimize the AdaBoost algorithm, which has more powerful feature search ability when constructing weak classifiers, so as to achieve higher detection performance and detection speed.

The picture uses the author's class photo, a total of 34 people. The actual effect of the traditional AdaBoost algorithm for face detection is shown in Figure 4, detecting a total of 27 faces, time-consuming 1134ms. The actual effect of this algorithm for face detection is shown in Figure 5, and all 34 faces are detected, consuming time 817ms. This shows that the algorithm is better in detection speed and detection accuracy.



Fig. 4. The actual effect of the traditional AdaBoost algorithm for face detection



Fig. 5. The actual effect of this algorithm for face detection

5 Conclusion

In this paper, a method of face detection based on chaos genetic algorithm optimization adaBoost algorithm is proposed. Firstly, this algorithm uses the image color segmentation for coarse screening on the face image, to determine the human skin area. Secondly, the adaptive median filtering is applied to denoise the face image to improve the quality of the face image. Finally, the chaotic genetic algorithm is used to optimize the AdaBoost algorithm to achieve higher detection rate and detection speed. Experimental results show that this method can improve the detection speed and ensure the accuracy of detection. The next step can aim at the real-time, robustness and generalization of the algorithm. Further optimization of the AdaBoost algorithm, in order to reduce the effects of illumination, facial expression and gesture on face detection.

References

1. Zhang Ya-Nan; Zhang Zuo. Survey of human face detection and tracking technology[J]. Computers and modernization,2012,12(2):60.
2. Xu Qian; Zhao De-An; Zhao Jian Bo. Face detection and localization based on improved AdaBoost algorithm[J]. Transducer and microsystem technologies,2010,29(1):94.
3. You Qing-Qing; Shen Hai-Yun; Luo Jun et al. Survey of face detection technology[J]. Wireless internet technology,2017,5(10):137
4. Zhang Yun-Long; Xie Ze-Qi; Zhang Hui-Min et al. A method of face detection in complex background[J]. Transducer and microsystem technologies,2011,30(6):42
5. Liang Lu-Hong; Ai Hai-Zhou; Xu Guang-You et al. Survey of face detection[J]. Journal of Computer Science,2002,25(5):449.
6. Liu Yu-Sheng; Feng Rui. A face detection algorithm based on AdaBoost[J]. Computer engineering and Applications,2016,52(11):209.
7. Liu Zhuo-Ya. Review of image denoising techniques[J]. Science and technology information,2013,15:317
8. Tomasz Marciniak; Agata Chmielewska; Radoslaw Weychan et al. Influence of low resolution of images on reliability of face detection and recognition[J]. Multimed Tools Appl,2015,74: 4329.
9. Zhang Hua-Xun. A classifier training method for face detection based on AdaBoost[C] //Proceedings of the International Conference on Transportation, Mechanical and Electrical Engineering,2011:731.
10. Chen Shi-Gang; Ma Xiao-Hu; Zhang Shu-Kui. AdaBoost face based on haar-like intensity features and multithreshold features[C] //Proceedings of the International Conference on Multimedia and Signal Processing,2011: 251.
11. Zhang Jun-Yi; Zhang Ze. Face detection based on improved AdaBoost algorithm[J]. Computer simulation,2011,28(7):240.
12. Pratibha Sukhija; Sunny Behal, Pritpal Singh. Face Recognition System Using Genetic Algorithm[C] //International Conference on Computational Modeling and Security(CMS 2016),2016,85:410.
13. Zhou Xian-Fu. The principle and application of genetic algorithm[J]. Science and technology,2017,03:265
14. Petra Snaselova; Frantisek Zboril. Genetic Algorithm using Theory of Chaos[C] //ICCS 2015 International Conference On Computational Science,2015,51:316
15. Xuan Deng-Ying; Wang Fu-Lin; Gao Min-Hui et al. A genetic algorithm with improved fitness function[J]. Mathematics in practice and theory,2015,16:232