

The Study on the Model of Effective Face Retrieval in Pedestrian Detection

Hua-Yang LI^{1,a}, Pan-Pan XIA^{2,b}, Hui-Lian XU^{3,c}

¹Jiangxi University of Finance and Economics, School of Software & Communication Engineering, China

² Jiangxi University of Finance and Economics, School of Software & Communication Engineering, China

³ Jiangxi University of Finance and Economics, School of Foreign Languages, China

^a49834715@qq.com, ^b398368708@qq.com, ^c81036404@qq.cpm

* Hua-Yang LI

Keywords: Effective Face Recognition, Symmetry Coefficient, Distance Coefficient.

Abstract. Pedestrian detection and retrieval plays an important role in intelligent video surveillance and automobile auxiliary driving system. Video surveillance creates mass monitored data among which a little effective information exists and could be hardly found. To settle the problem, based on pedestrian detection and face identification, the paper proposes a judgement and retrieval model of effective faces in pedestrian detection. The model helps to delete useless data and reserve the pedestrian data with effective faces, therefore, it provides data base for pedestrian image retrieval.

Introduction

It has achieved a lot in the researches of pedestrian detection with a great deal of manpower and funds due to the need for intelligent video surveillance. However, there exist difficulties in pedestrian detection and retrieval because features of pedestrians are changeable and complex in different situations. There is no universal algorithm with high performance for pedestrian detection and retrieval [1].

The histogram of oriented gradients (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection. The feature is composed of statistical histograms of gradient information in the local region of the image [2,3,4]. To accelerate calculation speed, blocked histogram is calculated with integral images.

$$h(x, y) = \sum_{x' \leq x, y' \leq y} h(x', y') \tag{1}$$

Eq. 1 defines integral image calculation. The integral image of $M(x, y)$ is a gradient histogram of all the pixels in the upper left corner. $h(x, y)$ refers to calculated integral image, $h(x', y')$ is the gradient histogram of all the pixels in the upper left region. The concept of the integral histogram is shown in Fig. 1.

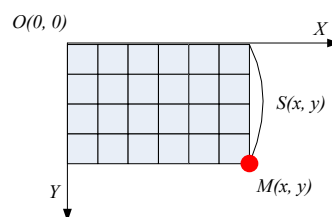


Fig. 1. A diagram of integral images

It is time consuming to view substantial data in surveillance videos. The purpose of the paper is to implement human retrieval in videos to construct database information for relevant departments to retrieve and query.

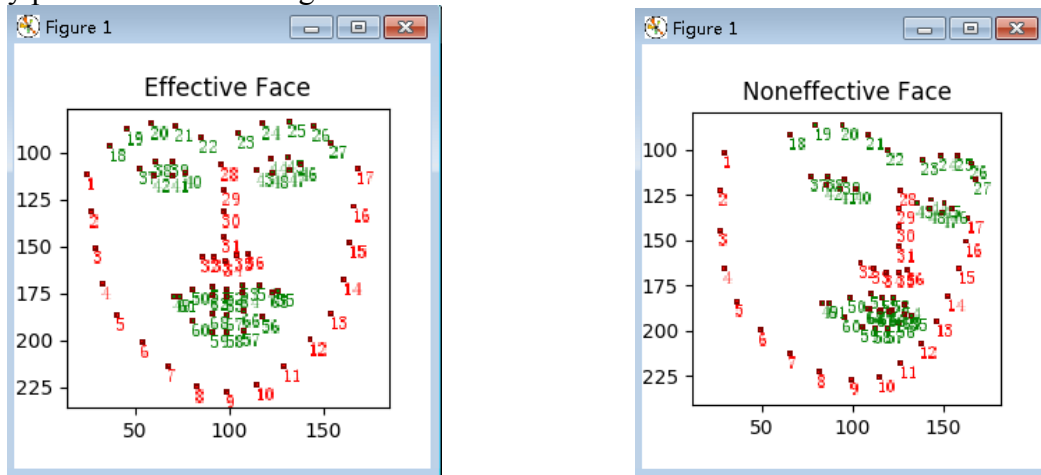
Our Work

It is essential to establish retrieval system to obtain relevant helpful images of pedestrians to greatly reduce the workload of some staff to settle problems. Images of people used to be retrieved should be the clearest facial images or shapes. Therefore, in the paper Effective Face refers to the clearest pictures of faces of people in a video. It is difficult to accurately judge the face posture with a single evaluation indicator[5,6]. To achieve accurate judgement of face images, the paper proposes two effective indicators, face area coefficient and face double symmetry coefficient. The effective coefficient of face symmetry is composed of distance symmetry and area symmetry of partial key points.

Moreover, the imaging area of an object is smaller in the distance than that nearby for a fixed surveillance device. Thus the overall area of a human face is an important parameter as well.

Distance symmetry of partial key points

The paper achieves face recognition for pedestrian images with Dlib toolkit which marks 68 key points. Suppose the symmetry coefficient of face key points is 1. If the face facing the lens squarely begins to move gradually towards the other side, its symmetry coefficient of face key points will decrease to 0 little by little. Not all parts of a face, such as the contour of the eye and mouth, own good distance symmetry features in the 68 key points marked in the face recognition. Therefore, a nose and the contour of the left and right side of a face are outlined as partial key points. The sequence number of the face contour key points is 1-17 and that of the nose is 28-36. The state of distance symmetry of selected key points is shown in Fig. 2.



(a)The distance between the key points is relatively symmetrical (b)The distance between the key points is not symmetrical

Fig. 2. Selected Key Points in Distance Symmetry

Steps for calculating the distance symmetric coefficient c_d of the partial key points is as follows.

Step 1 With the distance symmetric key point vector D , the contour feature vector D_1 and the nose feature vector D_2 are obtained. D'_1 , the feature vector of dimension 8, is obtained by recognizing the contour features D_1 according to symmetry.

$$D_1 = (P_1(x, y), P_2(x, y), P_3(x, y), \dots, P_{17}(x, y)) \quad (2)$$

$$D_2 = (P_{28}(x, y), P_{29}(x, y), P_{30}(x, y), \dots, P_{36}(x, y)) \quad (3)$$

$$D'_1 = ((P_1(x, y), P_{17}(x, y)), \dots, (P_i(x, y), P_{18-i}(x, y))), i = 1, 2, 3, \dots, 8 \quad (4)$$

Step 2 $\bar{P}(x, y)$, the center of the nose, is calculated according to D_2 . $\bar{P}(x, y)$ is regarded as a reference point when the contour symmetry is calculated.

$$\bar{P}(x, y) = (\bar{x}, \bar{y}) = ((\sum_{i=28}^{36} P_i^x)/9, (\sum_{i=28}^{36} P_i^y)/9) \quad (5)$$

Eq. 5 \bar{x} is the abscissa of the center point, \bar{y} is its ordinate. P_i^x is the abscissa of the marked point i , P_i^y is the ordinate of i .

Step 3 The corresponding distance symmetry coefficient c_d^i is calculated for each reference point $\bar{P}(x, y)$ in D_1' .

$$d_i = \sqrt{(\bar{x} - P_i^x)^2 + (\bar{y} - P_i^y)^2}$$

$$d_{18-i} = \sqrt{(\bar{x} - P_{18-i}^x)^2 + (\bar{y} - P_{18-i}^y)^2}$$

$$c_d^i = \frac{\min(d_i, d_{18-i})}{\max(d_i, d_{18-i})} \tag{6}$$

Eq. 6 d_i is the distance from the key point i to the center point \bar{P} . d_{18-i} refers to the distance from the key point $18-i$ to the center point \bar{P} .

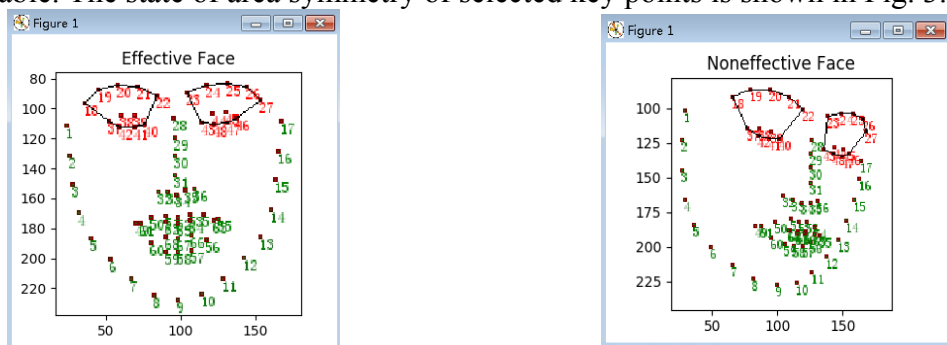
Step 4 Each key point is assigned with a weight and the distance symmetry coefficient c_d is given after the result of step 3 is average-weighted

$$c_d = (\sum_{i=1}^8 c_d^i) / 8 \tag{7}$$

Known from the above steps, the closer the value of c_d is to 1, the more symmetrical the face contour is to the nose. And the probability of having an effective face is greater. Due to a certain errors in marking key points of a face, the distance symmetry coefficient c_d is not possibly to be equivalent to 1. Therefore c_d is used to support an effective face in the process of ultimately determining which faces are effective or not.

Area symmetry of partial key points

The two eyes, highly symmetrical on a human face, are key reference factors in judging effective faces of pedestrians. The method of applying distance symmetry of key points to judge the features of a face is not suitable because key points in the eye region are highly concentrated without good distance symmetry. The paper puts forward the calculation method based on area symmetry of partial key points. First the left eye region and the right eye region are constructed with key points of the two eyes. Then the area of the two regions are calculated, with which their area ration is obtained to get area symmetric coefficient c_s . The closer c_s is to 1, the more similar the area of the left and the right parts of eyes will be and the more squarely the face faces to the lens, from which judging effective face is more probable. The state of area symmetry of selected key points is shown in Fig. 3.



(a) the key points of area are relatively symmetrical (b) the key points of area are not symmetrical

Fig. 3. Selected Key Points in Area Symmetry

Steps for calculating the area symmetric coefficient c_s of the partial key points is as follows.

Step 1 Once S , a key point vector of the area symmetry, is decomposed, S_1 (a key point vector of the left eye) and S_2 (a key point vector of the right eye) are obtained. In calculation of the area of eye part,

two key points in the left region or the right region are included in the eye polygon, so the four key points are omitted when S_1 and S_2 are selected to calculate the area of eye region.

$$S_1 = P_i(x, y), i = 18,19,20,21,22,40,41,42,37 \quad (8)$$

$$S_2 = P_i(x, y), i = 23,24,25,26,26,43,46,47,48 \quad (9)$$

Step 2 R_1 and R_2 are respectively constructed. R_1 is the polygon region of the left eye, connected by points in S_1 and R_2 is the one of the right eye, connected by points in S_2

Step 3 The area of R_1 and R_2 are respectively calculated. The area of R_1 is A_1 , The area of R_2 is A_2 .

Step 4 c_s , the area symmetry coefficient, is calculated.

$$c_s = \frac{\min(A_1, A_2)}{\max(A_1, A_2)} \quad (10)$$

A polygon is divided into several triangles whose area is calculated with the Helen formula. The sum of the area of the triangles is the polygon area.

The coefficient of judging effective faces

c , the coefficient of judging effective faces, is composed of the combination of distance symmetry coefficient and the area symmetry coefficient of partial key points.

$$c = f(c_d, c_s) = \frac{c_d + c_s}{2} \quad (11)$$

$$h(c, \theta) = c - \theta = \begin{cases} -1 & c - \theta \leq 0 \\ 1 & c - \theta > 0 \end{cases}, 0 \leq \theta \leq 1 \quad (12)$$

Eq. 12 $h(c, \theta)$ is a judging function and θ is a threshold parameter. Effective face and ineffective face are the result of judging faces in the paper. If $h(c, \theta) = -1$, the target image is judged to be ineffective face. If $h(c, \theta) = 1$, the target image is effective one.

Effective face detection

Based on effective face judgement, images are marked with facial geometric position features and image similarity is measured by the similar coefficient of double distance which is combined by weight Euclidean distance and average cosine distance. The steps of calculating facial geometric position features are as follows.

Step 1 Data of 68 key points are read and marked as P which are divided to be p_a and p_b . p_a is the feature points of the eye, contour and the mouth. p_b is feature points of a nose.

$$P = (P_1(x, y), P_2(x, y), P_3(x, y), \dots, P_{68}(x, y)) \quad (13)$$

$$p_a = P - p_b \quad (14)$$

$$p_b = (P_{28}(x, y), P_{29}(x, y), P_{30}(x, y), \dots, P_{36}(x, y)) \quad (15)$$

Step 2 $\bar{P}(\bar{x}, \bar{y})$, the facial center, is calculated by making reference to all the key points in p_b .

$$\bar{P}(\bar{x}, \bar{y}) = ((\sum_{i=28}^{36} P_i^x)/9, (\sum_{i=28}^{36} P_i^y)/9) \quad (16)$$

Eq. 13 \bar{x} is the abscissa of the center point, and \bar{y} is its ordinate. P_i^x is the abscissa of the marked point i and P_i^y is i 's ordinate.

Step 3 The maximum distance d is obtained by calculating the distance d_i from the feature point to the center point.

$$d_i = \sqrt{(\bar{x} - P_i^x)^2 + (\bar{y} - P_i^y)^2} \quad (17)$$

$$d = \max(d_i) \quad (18)$$

Eq. 14 \bar{x} is the abscissa of the center point, and \bar{y} is its ordinate. P_i^x is the abscissa of the key point i and P_i^y is i 's ordinate, d_i presents the distance from the key point i to the center point, d is the maximum value of distance.

Step 4 d is the denominator of distance normalization. The distance of all feature points is normalized to the same unit length, and the 59 normalized features are constructed as a facial geometric feature vector F .

$$F = \left(\frac{d_1}{d}, \frac{d_2}{d}, \frac{d_3}{d}, \dots, \frac{d_{59}}{d} \right) \quad (19)$$

Similarity calculation Weighted Euclidean distance

First, the similarity coefficient of two images is defined to be D . Then WD_e and MD_c are obtained by calculating similar distances of facial geometric position features with the improved method. Where, the weighted Euclidean distance with the value range $[0, +\infty)$, the smaller the value is, the more similar the two images are. The mean cosine distance with the value range $[-1, 1]$, the bigger the value is, the more similar the two images are. To combine the two coefficients, the similarity coefficients are defined as follows.

$$D = \frac{1}{1+WD_e} + MD_c \quad (20)$$

Eq. 20 the value range of $\frac{1}{1+WD_e}$ is $[0, 1]$, and the value range of MD_c is $[-1, +1]$. Thus the value range of the similarity coefficient D is $[-1, 2]$, which shows the bigger the value is, the more similar the two images will be.

Analysis of the experiment

The hardware applied in the experiment is as follows.

Table 1 Experimental Hardware Condition

Operation System	Window7 64bits Version
Memory(RAM)	8.00GB
Processor	Intel(R) Xeon(R) CPU E7-4820 v2 @ 2.00GHz
Hard Disk	500G
Toolkit	Opencv-3.2.0-vc14, Dlib-19.2
Development Language	C++, Python

Analysis of pedestrian detection

INRIA Database of pedestrian detection and recorded video data are applied in the experiment. INRIA Database is used to train pedestrian detection classifier. With the trained classifier pedestrian recorded in the experimental data are detected. INRIA Database contains test and training data sets. In test data, 288 samples are positive and 453 samples negative. In testing data, 613 samples are positive and 1218 samples negative. The experimental data are practical in actual application situations because the data acquisition is closely influenced by the complex background, light conditions and human gestures.

Table 2 Comparison of detection performance with different HOG feature methods

Feature method	Recognition rate	Error detection rate	Average detection time
HOG feature	72.1%	40%	2.3s
HOG features of 3D linear interpolation	80.2%	27%	7s
Integral histogram accelerated HOG feature	70.6%	43%	0.38s

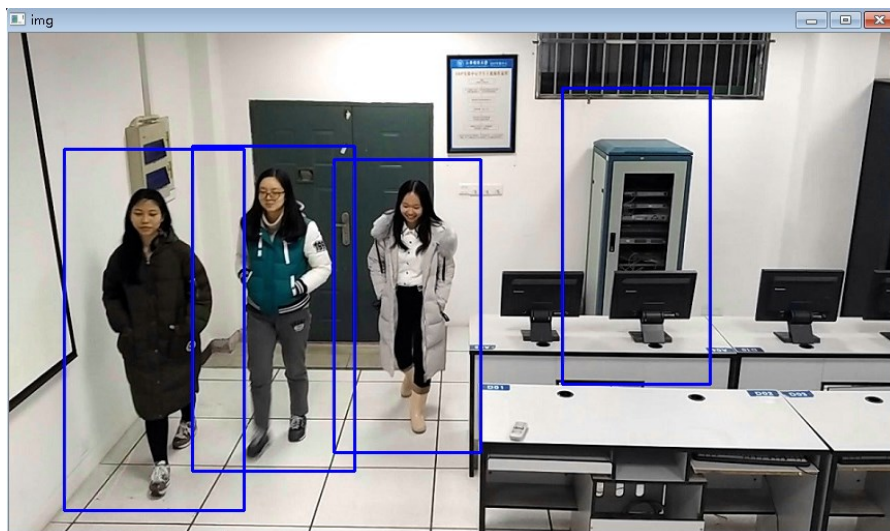


Fig. 4. Effect detected picture with Integral histogram accelerated HOG feature

In the analysis of the detection result, higher recognition rate exists in the detection with 3D linear interpolation and faster detection speed appears with integral images. A pedestrian target will appear in several continuous frames in a video file and in the stage of judging effective faces some images of pedestrians will be filtered, so pedestrian can be well detected in a video with Integral Histogram as shown in Picture 1.

Analysis of effective face judgement

Based on pedestrian detection, effective faces are judged in the pedestrian data. In the process face key points mark and



Fig. 5. Pedestrian images



Fig. 6. Corresponding images of face key points

Effective coefficients of images are calculated according to marked result. Judgement coefficients of 6 images in Fig. 6 are shown in Table 3.

Table 3 Judgement coefficients of 6 images

No.	Distance symmetry coefficient of partial key points	Area symmetry coefficient of partial key points	Double symmetry coefficient
A1	-1	-1	-1
A2	-1	-1	-1
A3	0.598937384795	0.599156118143	0.599046751469
A4	0.742306122578	0.754166666667	0.7482363946225
A5	0.896432752106	0.884955752212	0.890694252159
A6	0.938131843406	0.911016949153	0.9245743962795

Known from the experimental result, distance symmetry coefficient and area symmetry coefficient of partial key points are much closer to 1 when pedestrians face the lens more squarely.

Summary

The paper proposes a model of judging and retrieving effective faces based on the technology of pedestrian detection and face recognition. The model optimizes the feature extraction method with integral images by detecting pedestrians in videos and retrieve effective faces with double symmetry coefficient to obtain more effective pedestrian data.

Acknowledgement

The paper is sponsored by visiting Scholar Project of Faculty Development of Jiangxi Province.

References

- [1] D. T. Nguyen., W. Li, P. O. Ogunbona. Human Detection from Images and Videos[J]. *Pattern Recognition*, 2016(51): 148-175.
- [2] N. Dalal. Finding People in Images and Videos[D]. Grenoble: Institut National Polytechnique de Grenoble, 2006.
- [3] Xu Yuan, Xu Xiaoliang, Li Cainian, etc. Pedestrian Detection Combining with SVM Classifier and HOD Feature Extraction. [J]. *Computer Engineering*, 2016, 42(1): 56-60.
- [4] K. Dohi, K. Negi, Y. Shibata, et al. FPGA Implementation of Human Detection by HOG Features with AdaBoost[J]. *Ieice Transactions on Information & Systems*, 2013, 96(8): 1676-1684.
- [5] Liu Jinkui Researches on Recognition Methods of Face Frontalization [D]. Shenzheng: Zhongshan University, 2009.
- [6] Zhang Qian Comparative Research of Image Retrieval Based on Text and Content [J]. *Information Retrieval*, 2012(1): 111-113.