Acta Technica *62* No. 2A/2017, 293–302

Matching technique of image extraction and machine learning based on face recognition

Yunfei Li[2,3], Zhaoyang Lu[2], Jing Li[2]

Abstract. In order to design a kind of new technology used for feature extraction, a matching technique of image extraction and machine learning based on face recognition is proposed. As a biometric identification technology, human face recognition involves many fields such as artificial intelligence, pattern recognition, computer vision, image processing and analysis, image coding, computer graphics and so on. Compared with other identification methods, face recognition technology has some advantages, so it has a wide range of applications. It is mainly used in the field of national security, military security and public security, intelligent access control, intelligent video surveillance and other fields. A video based real-time face recognition system is designed in this paper. In this system, we adopt video capture technology. At the same time, the location and size of the human face can be detected by the function provided by the visual library, and the human face image can be extracted. Finally, the device can achieve 5 frames per second of recognition speed, and fully meet the application requirements. The experiment results show that the recognition rate of facial expression is improved after illumination normalization. After the improved histogram equalization method is used, the recognition rate is higher than that of the traditional histogram equalization method. Based on the above finding, it is concluded that the information security technology based on face information recognition will be widely used and benefit mankind.

Key words. Face recognition, feature extraction, video capture technology.

1. Introduction

Face recognition is a technique that uses computer to analyze face images and extract valid information from them so as to identify people's identity [1]. Compared with other mature biometric methods, the face recognition system has the advantages of low cost, no invasion and no human participation. But there are also

[1]The authors acknowledge the National Natural Science Foundation of China (Grant: 51578109), the National Natural Science Foundation of China (Grant: 51121005).

[2]School of Telecommunications Engineering, Xidian University, Xi'an, 710071, China

[3]School of Network Security and Information, Weinan Normal University, Weinan, 714099, China

some disadvantages: for example, it is vulnerable to the environment, hairstyles, age, make-up and jewelry. The human face is represented by a geometric feature vector, and the classifier is designed according to the hierarchical clustering in pattern recognition. This requires that the selected geometric feature vector has a certain unique, and can reflect the difference between different faces. At the same time, it also has a certain flexibility to eliminate time span, light and other effects. Geometric feature vector is a feature vector based on the shape and geometric relationship of human facial organs. Its components usually include the Euclidean distance and curvature angle between two points specified by the human face. Deformable template face has a better method of describing active contour, which makes the facial organ model parameterized. However, there are two problems [2]. The selection of weighting coefficients at various costs is empirical and difficult to generalize. The process of generalization is time-consuming and not suitable for real-time applications. he main work of this paper is to study the basic theory and key techniques of face detection and recognition. The problems of face detection, geometric normalization, feature extraction, feature selection and face recognition in video based color images and gray images are discussed. And the design and implementation of a video based real-time face recognition system is proposed.

2. Literature review

The earliest research of face recognition technology can be traced back to 1950s, and the researchers at that time were mainly involved in the field of social psychology. By 60s, a number of engineering documents had been published. But the research of automatic face recognition in real sense started from Kanade and Kelly in 1970s. The techniques used at that time were basically typical pattern recognition techniques, such as classifying and identifying the distances between important facial feature points. With the development of computer technology, from the 80s to the early 90s, face recognition technology has been greatly developed, and has entered the field of practical application. At this stage, statistical recognition methods based on facial appearance have been greatly developed. Experiments proposed by Eigenfaces and Fisherfaces on large scale face databases have obtained fairly good results. At the same time, the recognition method based on facial features has gradually developed. This method is not sensitive to the changes of light and angle of view, and the location of human faces. It is beneficial to improve the recognition rate, but the feature extraction methods used are not mature enough and reliable. Since the late 90s, a number of commercial face recognition systems have gradually entered the market. In recent years, face recognition has developed rapidly in the world as a computer security technology [3]. At this stage, more research has been focused on video based face recognition. The image is captured by a computer camera, then the position of the human face is detected from the video, and the captured face image is pretreated with light, size and position. Then, the features of human face are extracted and useful features are selected. Finally, the selected features are trained and identified to determine the face category of the feature and identify the identity of the person.

3. Methods

3.1. Overview of face detection

A standard face template is first predefined and parameterized. Then, an input image is provided, respectively, to calculate the facial contours, and the correlation between the eyes, nose, and mouth and the template. If the correlation is within a certain threshold, the input image is judged to be human face. Otherwise, it can be ruled as a non-face. Sakai and other researchers modeled with the eye, nose, mouth, and human face contours and detected the positive faces in photos. Each sub template is defined according to the line segmentation, and the line of the input image is extracted based on the maximum gradient. Then, the extracted line is matched with the sub template, and the relation between the sub image and the contour template is computed to detect the candidate area of the face, and the other sub templates are used to match the candidate region [4].

Face detection is to determine whether the human face exists in the input image. If there is a human face, then the number and the location and size of each face are determined. As a key technology in face information processing, face detection is the first part of face recognition system. Whether the location of a human face can be detected from an image or video stream correctly has a significant impact on the subsequent feature extraction and recognition. At the same time, the research of face detection is of great academic value. Human face is a kind of natural structural object with complex details. The challenge of this kind of target detection problem is that the difference position of human face and camera will lead to difference face image, such as the acquisition of human face image in front, and side and 45-degree angle. There may be glasses, beards and other appendages on the face, including their different shapes, sizes, and colors. At the same time, it is also influenced by people's appearance, facial expression and skin color, and the conditions of photography, such as intensity of light, light distribution and characteristics of cameras. Therefore, the methods to solve these problems need to be found to construct a face detection and tracking system successfully, which will provide important implications for other similar complex pattern detection problems. The method of detecting human faces from an image can be divided into the following categories: skin color based methods, template matching methods, feature based methods, and neural network based methods.

3.2. Traditional histogram equalization algorithm (HE)

The gray histogram of the image represents the number of pixels with each gray level in the gray image, and reflects the frequency of each gray level in the image. It is one of the basic statistical features of the image. Histogram equalization is the use of histogram statistics to modify the histogram. The gray values of each point in the image are changed by some corresponding relation, so that the histogram of the transformed image is relatively uniform and flat with respect to the original image, and the gradation is clear. Thus, the purpose of image enhancement and reducing

the influence of illumination on recognition are achieved. If the variable r represents the gray level of the pixel in the image, the r is normalized. That is, r is a decimal between (0, 1), where $r = 0$ stands for black and $r = 1$ stands for white. For a given image, each pixel obtained in [0, 1] intervals within the gray scale is random [5]. R is a random variable, which can be used to represent the gray distribution of the original image by the probability density function $Pr(r)$. The gray distribution feature of an image can be seen from the distribution of gray level. If the gray values of most pixels in the image are in the region close to 0, the whole image is dim, otherwise, the whole image is brighter. The ideal histogram equalization process is designed to make the histogram of the processed image be flat. That is, each gray level has the same frequency of occurrence, so that each gray level has a uniform probability distribution, so that the image looks clearer. Therefore, it is necessary to find a transform relation $S = T(R)$, so that the new gray histogram after transformation is more straight than the histogram before transformation. The traditional histogram equalization adopts the cumulative probability distribution function of R as the transfer function, that is,

$$S = T_r = \int_0^r Pr(w)\,\mathrm{d}w\,. \tag{1}$$

In formula (1), w is the integral variable, and $\int_0^r Pr(w)\,\mathrm{d}w$ is the cumulative probability distribution function of r. Let r_k is used to represent discrete gray values, and $Pr(r_k)$ is used to represent $Pr(r)$. Then

$$Pr(r_k) = \frac{n_k}{n}, \quad (0 < r_k < 1, \quad k = 0, 1, 2, \ldots, L-1)\,. \tag{2}$$

In the formula, n_k is the number of pixels in this image and n is the total number of pixels in the image. Symbol L represents the gray value of the image, which is divided into L classes. The discrete form of the transformation function is

$$S_k = T_{rk} = \sum_{j=0}^{k} \frac{n_j}{n} = \sum_{j=0}^{k} Pr_{rj}, \quad 0 < r_k < 1, \quad k = 0, 1, 2, \ldots, L-1)\,. \tag{3}$$

It can be proved that when the histogram of the image is uniformly distributed, the amount of information contained is the largest.

3.3. Improved histogram equalization algorithm (NHP)

In formula (3), for a 256×100 with 256-gray level image, if the pixel number of grayscale k in original image is less than 100, then, in the converted image, there will be a combination of the adjacent gray levels, and it is impossible to have its own gray value. This results in that the gray range of the transformed image is difficult to reach the maximum gray range permitted by the image format. At the same time, because of the "gray phagocytosis", it is also easy to cause the loss of image information. Especially when the histogram distribution of original image is very

uneven, such as the human face image acquired in bright or very dark environment, the level of image is worse and the information loss is more serious [6].

Jiang Duan and GuoPing Qiu proposed an improved histogram modification algorithm. Based on the RGB color space of image, this method uses high resolution gray value, which floats to indicate the original image pixel gray value, so as to effectively reduce the influence caused by the annexation of gray. Algorithm principle: the lower form is used as the gray transformation formula of RGB color space, in which Lu is the floating-point number type. In our experiment, Lu takes two points after the decimal point

$$\mathrm{Lu}(x,y) = 0.299 \times R_{(x,y)} + 0.587 \times G_{(x,y)} + 0.144 \times B_{(x,y)}\,. \tag{4}$$

Local histogram equalization (LHE): Histogram equalization is the transformation of the distribution of pixel values across the entire range of images. However, the best results achieved in the global range may not be optimal in local areas. The local histogram equalization adopts the same algorithm as the traditional global histogram equalization. It changes the image area of the algorithm, and decomposes the image into a number of $m \times n$ windows. The region of each window is balanced by histogram, so that the images in each region can get the best results. In the experiment, the original image is divided into M rows and N columns, and the $m \times n$ regions are balanced by histogram respectively.

Adaptive histogram equalization (AHE): The adaptive histogram equalization algorithm uses the histogram of the local correlation region with the pixels as the center. It applies the result of histogram equalization in this region to the pixel point method instead of the histogram of the whole image area to modify the pixels. That is to say, for each pixel of the image, the first thing to do is to find a region (such as the window of the $m \times n$ centered at that point). In this area, the mapping function is calculated according to the histogram equalization algorithm described above, and the new values of the pixels are calculated according to the mapping function. Adaptive histogram equalization needs to compute the histogram of each pixel of the image. Therefore, the amount of computation is considerable. However, since the corresponding regions of adjacent pixel points are related, their correlation can be used to optimize the computational process [7].

4. Results and discussion

In the experiment, the Gabor filter is used to extract the facial features, and the K mean distance classifier is used to recognize the features. PCA and PCA+LDA are used to reduce the features of the two methods. The training and testing samples are derived from the AR face database, and the recognition rate is shown in Fig. 1. 134 people in the AR database are selected, each of whom contains 7 frontal images with no mask on the face [8]. The recognition rates using various feature selection methods are shown in Figs. 2 and 3.

The experimental results show that the recognition rate of facial expression is improved after illumination normalization. After the improved histogram equalization

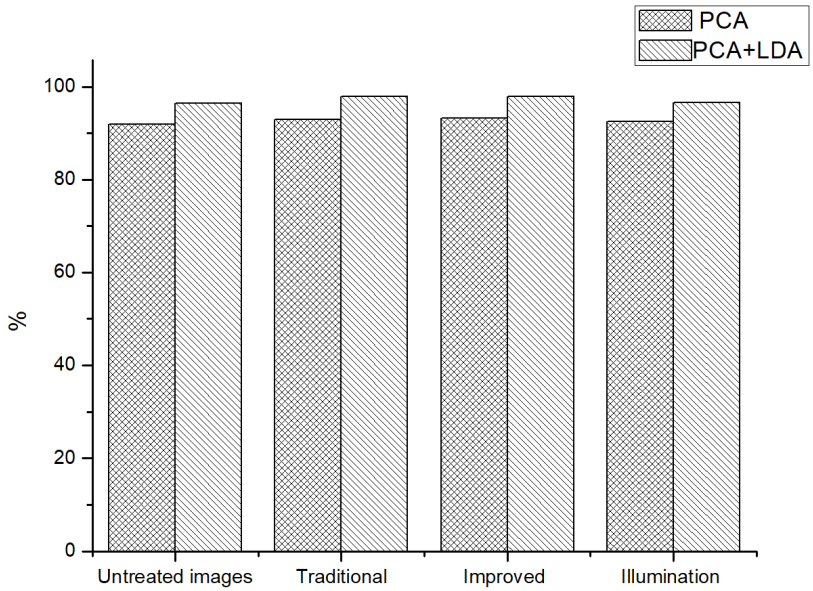

Fig. 1. Recognition rate of different illumination normalization in face recognition system

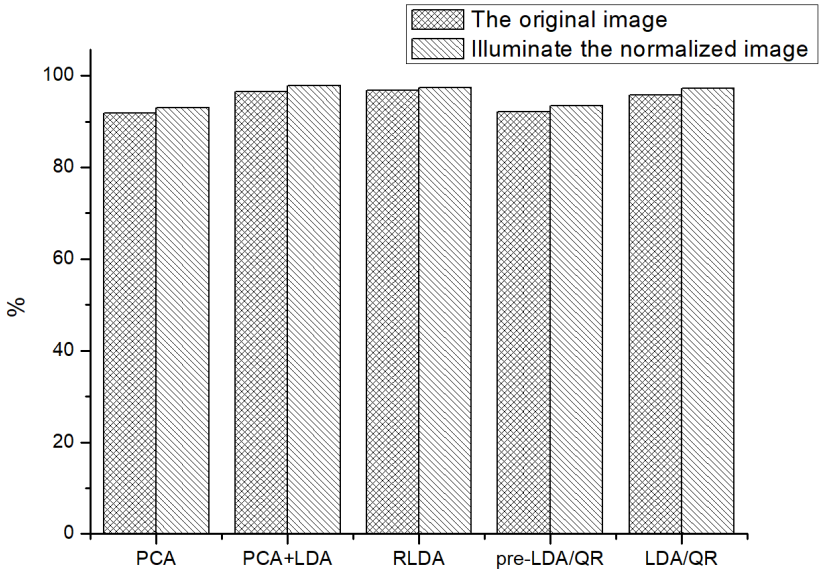

Fig. 2. Comparison of different feature selection with global Gabor

method is used, the recognition rate is higher than that of the traditional histogram equalization method. However, the improved histogram equalization requires that

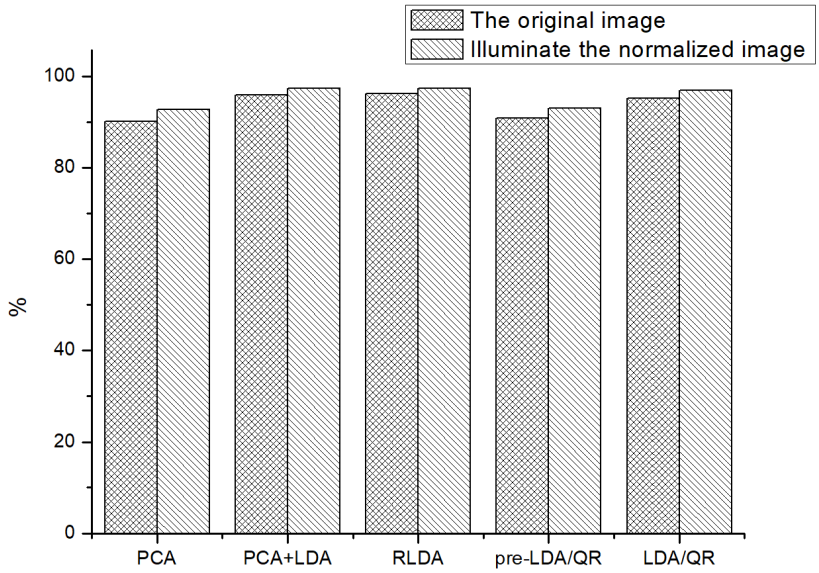

Fig. 3. Comparison of different feature selection with local Gabor

the original image contains RGB color space information, which limits its range of applications. However, the illumination normalization method based on affine transformation illumination model is poor in real-time performance, and the processing speed increases with the complexity of the filter used. At the same time, in the experiment, the Gabor feature is reduced to dimension. This process may remove the feature vector that represents the illumination property of the image. Therefore, the effect of different illumination normalization methods has some limitations. The use of PCA+LDA, RLDA5 and LDA/QR feature selection method can get higher recognition rate [9]. Among them, RLDA needs to spend more computing time, and PCA+LDA is the second, LDA/QR has the lowest computational complexity. Since the number of training samples in our system is relatively small, it can be seen from the above analysis that the computational complexity of PCA+LDA is very different from the computational complexity of LDA/QR. And the recognition rate of PCA+LDA feature selection method is 0.2 to 0.75 percentage points higher than that of LDA/QR. Therefore, PCA+LDA is adopted as dimension reduction feature selection method in this system.

A real-time face recognition system based on video is designed and implemented. In this system, DirectShow video capture technology is adopted. After capturing an image, firstly, the location and size of the human face are detected by the function provided by Intel OpenCV visual library, and the human face image is extracted. Then the face image is pre-processed, including eye location and geometric normalization. Then the feature extraction and feature reduction is combined to extract the low dimensional Gabor features of human face. Finally, the feature is input into the classifier to classify the feature, and the classifier outputs the recognition

result. The classifier used in this system is Euclidean distance classifier. The system diagram is shown in Fig. 4.

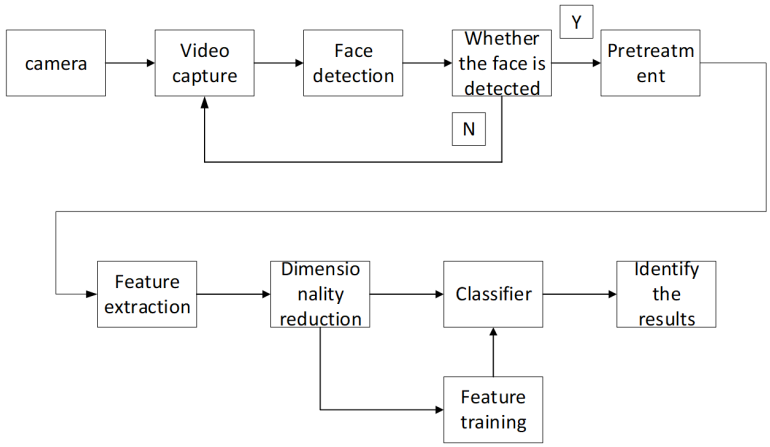

Fig. 4. Framework of real-time face recognition system

In the real-time face recognition demo system based on video, a common USB camera is used, and the resolution of the image is 320×240. In order to improve the speed of face detection, we stipulate that only the face not less than 20×20 is detected. In the face detection part, multiple faces is detected at the same time, but for the convenience of identification, the first face is only recognized. At the same time, in order to improve the accuracy of recognition, multi frame recognition results voting strategy is adopted (as shown in Fig. 5). The recognition results of 5 consecutive face images are classified into the categories with the highest number of recognition results. As the final recognition results, the input feature classification results are output.

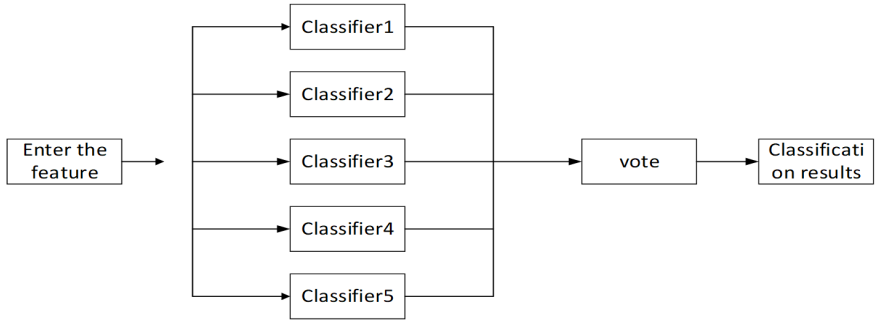

Fig. 5. Strategy of classifier

Microsoft Visual C++NET is selected as the development platform to complete the face image acquisition, detection, preprocessing, feature extraction and selection, identification of the whole process. Because C++ has the characteristics of high code efficiency, convenient memory and powerful function, it can make up for the

inconvenience of large amount of data, complex operation and long time-consuming in image processing. Matlab is selected as a tool for feature selection training, and Matlab's powerful matrix computing function is used for offline computation. The establishment of real-time face recognition system is image acquisition and regular USB camera. The image separation rate is 320×240 (that is, the size of the camera monitor area). The CPU size is P4 3OG. Memory is 512 MB. The software development platform is Microsoft Visual C++ NET. Table 1 presents the average processing speed of a real-time face recognition system. Since image capture and subsequent processing processes are synchronous and interactive, the processing speed given in the previous table is the average processing speed for each process separately. As can be seen from Table 1, the entire process of the system takes about 200 ms time, and has basically met the requirements of real-time systems.

Table 1. Processing speed

Image capture	Face detection	Preprocessing	Feature extraction and selection	Classification recognition
$<$29 ms	$\sim$100 ms	$<$29 ms	50$\sim$ 100ms	$<$10 ms

5. Conclusion

The structure of face recognition system is divided into image capture, face detection, image preprocessing, feature extraction and feature selection, and face recognition. The problems of face detection, geometric normalization and feature selection are discussed. Real time face recognition demo system based on video is designed. It uses a common USB camera, and real-time recognition of the front of the camera face is realized through pretrained templates It has the advantages of fast recognition speed and high accuracy. It can achieve real-time monitoring effect, and can also be used in practical applications.

With the development of micro electronics, computer and network technology, traditional identification methods such as magnetic cards, IC cards and passwords are facing severe tests. The appearance and rapid development of face recognition technology are expected to solve these problems in the near future. Because any person's facial information features are always followed by their own, they do not worry about losing face and being stolen. In addition, as face features are relatively complex and difficult to copy, it has higher reliability and security as a security key. Therefore, with the rapid progress of people's in-depth research for face recognition technology and computer science and technology, in the near future, the information security technology based on face information recognition will be widely used and benefit mankind.

References

[1] J. Cao, Y. Zhao, X. Lai, M. Eng Hock Ong, C. Yin, Z. X. Koh, N. Liu: *Landmark recognition with sparse representation classification and extreme learning machine.* Journal of the Franklin Institute *352* (2015), No. 10, 4528–4545.

[2] J. Lu, G. Wang, P. Moulin: *Localized multifeature metric learning for image-set-based face recognition.* IEEE Transactions on Circuits and Systems for Video Technology *26* (2016), No. 3, 529–540.

[3] Z. Lei, M. Pietikäinen, S. Z. Li: *Learning discriminant face descriptor.* IEEE Transactions on Pattern Analysis and Machine Intelligence *36* (2014), No. 2, 289–302.

[4] J. Lu, V. E. Liong, X. H. Zhou, J. Zhou: *Learning compact binary face descriptor for face recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence *37* (2015), No. 10, 2041–2056.

[5] G. Deepak, J. Lee: *Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines.* Sensors *13* (2013), No. 6, 7714–7734.

[6] J. Tompson, M. Stein, Y. Lecun, K. Perlin: *Real-time continuous pose recovery of human hands using convolutional networks.* Journal ACM Transactions on Graphics (TOG) *33* (2014), No. 5, Article No 169.

[7] R. Girshick, J. Donahue, T. Darrell, J. Malik: *Region-based convolutional networks for accurate object detection and segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence *38* (2016), No. 1, 142–158.

[8] Y. T. Yu, G. H. Lin, I. H. R. Jiang, C. Chiang: *Machine-learning-based hotspot detection using topological classification and critical feature extraction.* IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems *34* (2015), No. 3, 460–470.

[9] S. Puttinaovarat, P. Horkaew: *Urban areas extraction from multi sensor data based on machine learning and data fusion.* Pattern Recognition and Image Analysis *27* (2017), No. 2, 326–337.

Received July 12, 2017