Facial Expression Recognition Based on Gabor Multi-orientation Feature Fusion

Xuechen Liang*, Lei Yang and Shengda Luo
Macau University of Science and Technology, Macau, China

*E-mail: 977717682@qq.com

Abstract. Facial expression is a key to nonverbal communication, which has been confirmed by many different research projects. A change in intensity or magnitude of even one specific facial expression can cause different interpretations. With the continuous and fast development of computer vision and pattern recognition, facial expression recognition has received significant attention recently due to the wide range of commercial and law enforcement application and the availability of feasible technology during 30 years of research. In this thesis, facial expression recognition is studied by applying several commonly used methods in the whole process. By numerical experiment, we find that our approach with Gabor based transformation for face expression feature extraction, combining the advantages of various algorithms, Gabor wavelet transform and non-negative matrix decomposition of facial expressions are used to obtain features, and CNN is used to classify and apply them to facial expression recognition.

1. Introduction

With the development of computer technology, computer vision technology has emerged and has been successfully applied in the fields of industrial automation and agricultural product inspection and testing. Among them, the use of computer vision technology for the commercialization of fruit and vegetable represented by automated collection and grade classification has a very broad development space.

A system for measuring the facial expression based on machine vision is designed for facial expressions. The use of digital image processing technology combined with computer technology grading means can realize facial expression images, storage and reproduction, and has important significance for the establishment of facial expression image processing, and is of great significance to the development of the transportation industry.[1] In order to accurately detect various parameters of the facial expression, the camera needs to be calibrated to eliminate the influence of camera distortion on the detection accuracy. The position of the surface point of the space object on the image is related to the geometric position of the point. The relationship between these positions is determined by the camera imaging geometric model. The parameters of the geometric model are called camera parameters. These parameters must be determined by experiments and calculations. The process of experimentation and calculation is called camera calibration. Image collection and recognition is a more important topic. Before the level judgment of the collected facial expression image is judged, a series of pre-processing of the selected image is first performed. First, a grayscale process is performed to convert the acquired colour image into a grayscale image. Second, for higher quality
image support for subsequent analysis, the image needs to be noise removed. Third, extract the Gabor wavelet feature, and fourth, perform CNN recognition.

2. Feature Extraction Based on Gabor Wavelet Transform and Non-negative Matrix Factorization

Fourier transform is a method of signal analysis that only reflects the global characteristics of the signal. In order to obtain information about the change of signal within a certain time, Gabor proposed a new method, that is, introduced a window function to localize time in Fourier transform function, to a certain extent, overcome the shortcoming that Fourier transform can’t realize localization analysis. Therefore, it belongs to Fourier of Gaussian window Transform. In 1946, the Gabor transform was formally proposed. Subsequently, Gabor transform is widely used in signal analysis and image processing and other fields[2]. In 1966, Lee successfully characterized the features of the image using Gabor wavelet transform, emphasizing that Gabor wavelet transform has a good biological background and has the biological characteristics of the same species. Since Gabor wavelet has a similar biological which can effect to the human eye, it is often used in texture feature extraction.

Gabor wavelet amplifies the change of grayscale like a magnifying glass, and local features in some key functional areas of human face (eyes, nose, mouth, eyebrows, etc.) are enhanced to distinguish different face images. The Gabor wavelet kernel function has the same characteristics as the two-dimensional reflex region of the simple cells in the cerebral cortex of nursery animals, that is, it has strong spatial position and direction selectivity and can capture the local structure information corresponding to the space and the frequency [3]. The Gabor filter will change the brightness and contrast of images and the change of face pose have strong robustness, and it expresses the most useful local features for face recognition. Gabor wavelet is a good approximation to the neurons in the visual cortex of the higher vertebrate and is a compromise between the accuracy of the time domain and the frequency domain.

Gabor function is a complex sinusoidal function modulated by Gaussian function, which can extract local frequency domain features in a given region [4]. A typical 2-D Gabor function $h(x, y)$ and its Fourier transform $H(u, v)$ has the following form [5]:

$$
\begin{align*}
g(x,y) &= \frac{1}{2\pi\sigma_x\sigma_y} \exp\left( -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right) \\
h(x,y) &= g(x,y) \cdot \exp(2\pi j w_c) \\
H(u,v) &= \exp\left( -\frac{1}{2} \left( \frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right) \right)
\end{align*}
$$

While $g(x,y)$ is the Gaussian function used to modulate; $\sigma_x$ and $\sigma_y$ are their standard deviations on both axes, which determine the size of the filter’s area of action; $W$ is the complex sine function on the horizontal axis Frequency of the Gabor function is decomposed into two components, the real part $h_R(x, y)$ and the imaginary part $h_I(x, y)$, and the image filtered is this[6].

$$
S(x, y) = \sqrt{(h_R * I)(x,y)^2 + (h_I * I)(x,y)^2}.
$$

While $(h * I)$ represents the convolution of image $I$ and filter $h$, and $S(x, y)$ is Gaussian smooth, which is the feature image extracted by Gabor filter.

If we use $h(x, y)$ as the mother wavelet, we can get a set of self-similar filters called Gabor wavelet through proper scaling and rotation transformation.

$$
hmn(x, y) = a \cdot mh(x', y'), a > 1, m, n \in Z.
$$

While $x' = a - m (xcos\theta + ysin\theta)$, $y' = a - m (-xcos\theta + ysin\theta)$, $\theta = n \pi / K$; $a$-$m$ is a scale factor; $S$ and $K$ are the number of scales and directions, By changing the values of $m$ and $n$, a set of Gabor filters with different directions and scales can be obtained. Suppose the wavelet
family contains $S$ scales, $K$ directions, and frequency range $[ Ul, Uh ]$. The parameter selection method is as follows:

$$
\begin{align*}
    a &= (U_h/U_l)^{-1/(S-1)} \\
    U_{(n)} &= a^n U_l \\
    \sigma_{u(n)} &= (a-1)U_{(n)}/[ (a+1)\sqrt{2\ln2} ] \\
    \sigma_{v(n)} &= \tan(m\pi/2K) [ U_{(n)} - 2\ln2(\sigma^2_{u(n)}) ]/\sqrt{2\ln2 - (2\sigma_{u(n)}\ln2/U_{(n)})^2}
\end{align*}
$$

Plural:

$$
    g(x,y;\lambda,\Theta,\psi,\sigma,\gamma) = \exp\left(-\frac{x^2 + \gamma^2 y^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right).
$$

Real part:

$$
    g(x,y;\lambda,\Theta,\psi,\sigma,\gamma) = \exp\left(-\frac{x^2 + \gamma^2 y^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \psi\right).
$$

The imaginary part:

$$
    g(x,y;\lambda,\Theta,\psi,\sigma,\gamma) = \exp\left(-\frac{x^2 + \gamma^2 y^2}{2\sigma^2}\right)\sin\left(2\pi\frac{x'}{\lambda} + \psi\right).
$$

One of the important is

$$
    x' = x\cos\theta + y\sin\theta, y' = -x\sin\theta + y\cos\theta.
$$

In the formula: $\lambda$: sine function wavelength; $\theta$: direction of Gabor kernel function; $\psi$: phase offset.

In general, the problem we present in this paper is that its dimension is too large. Two-dimensional $p \times q$ gray-scale image expansion is a $m = p \times q$ dimensional vector space. Then a 256 × 256 pixel image is in a 65536-dimensional in the space. The problem is: not all the dimensions are useful. We can only make decisions by looking for the components of the information that make up the majority of the image by calculating the mean square error of the data. PCA was independently proposed by Karl Pearson and Harold Hotelling [7] in transforming a set of potentially relevant variables into a set of smaller independent variables. The idea is that high-dimensional data sets are often described by related variables, so that only a few meaningful dimensions take up most of the information. The PCA method calls the direction of the largest variance in the data the principal component.

Algorithm Description

For any random vector $X = \{x_1, x_2, ..., x_d\}$, we calculate its mean:

$$
    \mu = \frac{1}{n} \sum_{i=1}^{n} x_i
$$

Calculate its covariance matrix $S$:

$$
    S = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^	op
$$

Calculate the eigenvalue $\lambda_i$ and eigenvector $v_i$ of $S$:

$$
    Sv_i = \lambda_i v_i \quad i = 1, 2, ..., n.
$$

The eigenvectors are arranged in decreasing order of Eigen values. The k principal components are the eigenvectors corresponding to the k largest Eigen values. The k principal components of $X$ are

$$
    y = W^T(x - \mu)
$$

PCA can be used for image reconstruction.

$$
    x = \hat{W}y + \mu
$$
3. Convolutional Neural Networks
Convolutional Neural Network (CNN) is a kind of end to end mode, artificial neural network (ANN), which is inspired by the study of the cat's primary visual cortex: simple cells respond by receiving stimuli from specific edges of the local receptive field as input to complex cells. The network structure is shown in Figure 1[8]. The most especial of CNN is local connectivity and weight sharing. This makes the network parameters less, the training speed is faster. There is a certain regularization effect. CNN is a complex neural network. The loss function is an evaluation criterion for evaluating the degree of agreement between the predicted value and the true value. The commonly used loss function is the mutual entropy loss, hinge loss, and L2 of the soft ax classifier, Norm loss, L1 norm loss [9].

The key to CNN's ability to solve nonlinear problems is to activate the function. The function of the "activated neuron" is retained and mapped by the function. The commonly used activation function is tanh. Sigmoid, soft sign, ReLU, LReLU, etc. In order to prevent the network layer. The over-fitting phenomenon caused by the deeper number increases the generalization ability, and CNN generally needs to perform regularization operations. Commonly used regularization methods include data amplification, dropout, L2 regularization, and drop connect. Moreover, CNN uses the pooling (pooling layer) operation to minimize network complexity and keep CNN a certain local invariance, thus closer to the animal visual cortex mechanism.

![Convolutional neural network schematic](image)

**Figure 1.** Schematic diagram of the convolutional neural network

4. Experiments
In this chapter, the experiments will be carried out by using two methods. In the first method, the filter image of each filter is decomposed by PCA and classified by two-level classifier (Gabor + NMF + Libsvm). Method two, In the second method, feature extraction is first performed by Gabor, and then classified using a convolutional neural network (Gabor + NMF + CNN). We apply the two methods to the classification and recognition of expressions. The JAFFE facial expression database, YALE facial expression database and AR facial expression database were selected for related tests, and the ideal recognition rate was obtained.

4.1 Design and implementation of two-layer classifier
In the design process of the two-layer classification mode, the nearest neighbour classifier based on the Euclidean distance is used as the first layer classifier, and the probability statistical classifier based on the probability and statistics idea is used as the second layer classifier. The feature image of the filtered sub graph of each filter is the input of the nearest neighbour classifier, the output of the nearest neighbour classifier is the input of the probability and statistical classifier, and the output of the probability and statistical classifier is the final classification result of the image. The two-layer classifier design is shown in Figure 2.
5. Experimental results and analysis

In the classification and recognition stage of the image, the main work is to train the classifier and complete the classification, and output the classification result. Finally, according to the classification results, the recognition accuracy rate is judged, and the experimental results are analyzed. The experiment was carried out in the MATLAB 2015b programming environment. The experimental database selected JAFFE Japanese female expression database, YALE white male expression database and AR adult expression database. In this database, crop the image containing the emoticon and normalize its dimensions to. In the feature extraction stage, 40 different filters consisting of $\mu = 0, 1, 2, 3, 4$, $\mu = 0, 1, 2, 3, 4, 5, 6, 7$, and 5 scales are used to test 416 images in the database.

The division of experimental data is divided into two steps, namely vertical division and horizontal division. Firstly, the vertical division is performed. In the feature extraction stage, the filtered image of each filter is divided into a group of 40 groups to realize vertical division of data. Each group of filtered images after longitudinal division is further divided into three groups to realize horizontal division of data. In the first group, the first image of each expression of each person is selected as the test image, the remaining images are used as the training image, and so on. In the third group, the second image is selected as the test image for the expression image set of only two images, and the rest is used as the training image. In this way, there are 70 images in each set of test sets, and 143 images in the training set, which are cyclically tested to achieve cross-validation of data.

When the original non-negative matrix is decomposed by NMF, the selection of the number $r$ of feature bases determines the dimension of the sample image matrix projected in the feature subspace, which directly affects the feature selection of the image. When $r = 30$, the variance and mean of the number of features of the expression are shown in Figure 3 to Figure 6.

![Figure 2. Design of two-layer classification](image-url)
In the case of a small number of training samples, the class features extracted by them are relatively weak[10]. From table 1, we can found that when the number of training samples increases gradually, the categories extracted by the face images are more characteristic, and the facial expression recognition rate increases accordingly.
Table 2. Result of different algorithms

| Way           | Recognition rate (%) |
|---------------|----------------------|
| Image + NMF + NNC | 88.75               |
| Gabor + NMF + NNC   | 90                  |
| Gabor+ NMF + SVM     | 93.75               |
| Gabor+ NMF+CNN       | 97.75               |

Compared with Gabor + NMF + NNC, the recognition rate of this method is obviously better than that of Gabor + NMF + NNC, but the idea of this method is simple intuitive. From table 2, we can found that it is easier to understand than the Gabor + NMF + NNC method. Compared to the other three methods, Gabor+NMF+CNN is the best way.

Because of the images in the three facial expression databases are relatively small, filtering by Gabor wavelet transform increases the number of images and alleviates the defects of the sample shortage. The non-negative matrix factorization algorithm is used to decompose the image matrix to ensure the non-negativeness of the base matrix and the coefficient matrix, which is more interpretable. Each image can be regarded as a linear superposition of a set of base images. In the process of non-negative matrix factorization, it is very important to choose the appropriate r, not only to extract the characteristics of the reflected category information, but also to reduce the interference caused by redundant information and noise information.

6. Summary
Facial expression recognition is an emerging topic with great development potential and commercial value [11]. Currently, although it has been proposed very different algorithm models have achieved good recognition results, but further improving the accuracy and robustness of facial expression recognition has been the goal that researchers are striving for.

A new method of facial expression recognition based on the combination of Gabor feature extraction and convolutional neural network is proposed in this paper. In the JAFFE database, the correct recognition rate of the face expression was 97.75%. In the Yale database, the best recognition rate of the face expression was 98.57%. And compared to other state-of-art methods, the proposed method performs better than other methods. Experimental show that this method can be used for facial expression recognition, which can improve the recognition performance.

In the JAFFE database, this method can get the best face facial expression recognition rate of 98%; in the Yale database, the best 93.75% facial expression correct recognition rate has been achieved. Its recognition rate is much higher than other classification methods. It can be seen that this method is suitable for facial expression recognition and can better improve the recognition performance.

Because of the limited training samples and the absence of face samples such as different gender, age, face angle and face profile, the real-time facial expression recognition system designed in this paper has a certain degree of misleading rate in practical application. And for some similar expressions with only slight differences, it is difficult to distinguish them effectively. The next step is the need to collect a wider range of training samples, and improve the algorithm model to enhance the system’s real-time.

We investigated existing hardware and software for facial expression recognition and found that the correct rate of recognition of Caucasian expressions was significantly higher than that of yellowish facial expressions. Later on, we will improve all aspects of yellow facial expression recognition and improve the accuracy of yellow facial expression recognition.
References

[1] Zhou, J., Zhang, S., Mei, H., & Wang, D. (2016). A method of facial expression recognition based on Gabor and NMF. *Pattern Recognition and Image Analysis*, 26(1), 119-124.

[2] Zhang, L. J., Zhang, L., Li, Z. M., & Cui, S. Y. (2014). Study of Gabor Features and Heart Sound Signal Recognition by the Principal Component Analysis. In *Applied Mechanics and Materials* (Vol. 644, pp. 4452-4454). Trans Tech Publications.

[3] Xing Y, Luo W. Facial expression recognition using local gabor features and adaboost classifiers[C]//Progress in Informatics and Computing (PIC), 2016 International Conference on. IEEE, 2016: 228-232.

[4] Kalsi, K. S., & Rai, P. (2017, January). A classification of emotion and gender using approximation image Gabor local binary pattern. In *Cloud Computing, Data Science & Engineering-Confluence, 2017 7th International Conference on* (pp. 625-628). IEEE.

[5] Zhang, L., Hu, C., Wu, S., Wang, T., Cui, J., & Qiu, J. (2017, July). An improved non-local means image denoising algorithm. In *Information and Automation (ICIA), 2017 IEEE International Conference on* (pp. 781-786). IEEE.

[6] Zhang, S., Li, L., & Zhao, Z. (2012). Audio-visual emotion recognition based on facial expression and affective speech. In *Multimedia and Signal Processing* (pp. 46-52). Springer, Berlin, Heidelberg.

[7] Gottumukkal, R., & Asari, V. K. (2004). An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letters*, 25(4), 429-436.

[8] Takahashi, T., & Kurita, T. (2002, August). Robust de-noising by kernel PCA. In *International Conference on Artificial Neural Networks* (pp. 739-744). Springer, Berlin, Heidelberg.

[9] Papa, J. P., Rosa, G. H., Marana, A. N., Scheirer, W., & Cox, D. D. (2015). Model selection for discriminative restricted boltzmann machines through meta-heuristic techniques. *Journal of Computational Science*, 9, 14-18.

[10] Zhou, L., Wang, L., Ge, X., & Shi, Q. (2010, March). A clustering-Based KNN improved algorithm CLKNN for text classification. In *Informatics in Control, Automation and Robotics (CAR), 2010 2nd International Asia Conference on* (Vol. 3, pp. 212-215). IEEE.

[11] Ma, W., Li, T., Zhang, X., Fan, J., Yao, Y., & Yang, J. (2018, July). The Effect of Parameters on PI Control for Hydraulic Looper Control in Hot Rolling Mill. In *IOP Conference Series: Materials Science and Engineering* (Vol. 381, No. 1, p. 012063). IOP Publishing.