

Facial Feature Point Location in the Neural Network with Few Training Samples

Zhengyong Chen^{1, a}, Wei Xiang^{1, b, *}, Xiaohang Zhang¹, Jiaqi Shi¹,
Shuang Zheng¹, Xiaodan Zou¹

¹College of Electrical and Information Engineering, Southwest University
for Nationalities, Chengdu, Sichuan, China 610041

^a627973269@qq.com, ^b3730544@qq.com

(*Corresponding Author)

Key Words: Facial feature point location, Small sample training, Neural network, Sub-network

Abstract. This paper investigates a method to locate important facial feature points with small training samples. Firstly, the facial feature points are divided into several categories, then these various feature points are trained by LMBP neural network to get each sub-network. Outputs of these sub-networks can be combined to locate the important facial feature points. The experimental results show this kind of method based on neural network performs well, especially its calculation speed is fast, it can be applied in analysis of the facial expressions, facial reconstruction and other aspects.

Introduction

Facial feature point location is very important in face recognition and 3D face model reconstruction. It can provide important geometry information for face image processing and analysis, using the results of the facial features location can analyze facial expressions and track the face images automatically. Facial features location can also be applied in facial image encoding based on model; In facial 3D reconstruction, positioning is indispensable in the facial feature points.

Facial points positioning methods can be divided into two categories, one is based on the search space of facial locations method, the method is to determine whether the local areas are the key points, such as facial feature points positioning according to the priori rules, Kawaguchi used the improved Hough transformation to detect human eyes [1]; Another example is eyes locations method based on Adaboost [2-4], this method firstly builds multiple weak classifier for the same training, then weak classifier constitute a strong classifier to finish locating the eyes by decision analysis together. The other is the key point locations prediction method, such as ASM [5-7] and AAM [8]. The thoughts of such methods aim to form learning sets by using hand-marked face key points information. When testing, for any given new face image, it is estimated the similarity between new images and training samples, and the same similarity will be used in face shape reconstruction.

At present, these two methods of the key points of face orientation have achieved some results. However, there are still problems of high computational complexity and high cost of training samples resulting in less training samples.

ANN is composed of a lot of simple neurons connected to each other that become the adaptive nonlinear dynamic systems; It can carry out large-scale parallel processing and distributed storage of information. It has a strong ability of self-study, association and fault tolerance, good adaptability and self-organizing. These characteristics make ANN is able to approach more input to the more output of the nonlinear system.

This paper uses the BP neural network respectively to faces the important feature points to train sub-network, so as to realize the input image feature point positioning of the output.

Training of the Main Features of the Human Face

The Definition of Main Feature Points of Human Face. The realization of the important feature point locating method is to select the appropriate face image training sample set. In this paper, the famous ORL(Olivetti research laboratory) face database [9] is the Olivetti Research Institute of University of Cambridge in the UK. The database consists of 40 different people, each of the 10 images, a total of 400. Each original image is 256 gray level, with a resolution of 112*92. The ORL face image was taken at different time, from different perspectives, a variety of expressions with eyes closed (/ eyes, smiling / surprised / angry / anger / happy) and different facial details (wear glasses / did not wear glasses, beard / no beard, hair style is different) conditions.

After that, the selected face sample image feature points, the general manual marking, manual marking based on 'Artificial marking basis'.

The feature points of the feature points can reflect the feature of the sample, therefore, the complex object, the feature is relatively large, the feature points of the tag to include the eyes, nose, mouth, chin and facial contour. This paper chooses corner of the eye, nose, nose, lips, cheeks, chin, as the main face feature points. Fig. 1 is the location of the facial feature points. After the completion of the facial feature points, the feature points of the image are trained to construct the feature points and the neural network model is constructed.



Fig. 1 The location of the facial feature points.

Selection of Sub Neural Network.

LMBP neural network. LMBP neural network is a feed forward neural network, the main feature of the network is the signal forward pass, error back propagation. In forward pass, the input signal is processed from the input layer to the output layer by layer by layer. Each layer of neurons only affects the state of the next layer of neurons. If the output layer can not get the desired output, then the BP neural network is predicted by the network weights and thresholds, which are based on the prediction error to adjust the network weights and thresholds. The topological results of BP neural network are shown in Fig. 2.

In Fig. 2, x_1, x_2, \dots, x_n is the input value of LMBP neural network, y_1, y_2, \dots, y_n is the prediction value of LMBP neural network, V, W is the neural network weights. As can be seen from the graph, the BP neural network can be considered as a nonlinear function, the network input value and the predicted value are the independent variables and dependent variables of the function. When the input node number is n and the input node number is LMBP, the M neural network expresses the function mapping relation from n independent variable to m dependent variable.

The LMBP neural network can learn and store a large number of input and output Object Relational Mapping without prior revealing the mathematical equations describing the mapping relationship. Its learning rule is the use of Levenberg-Marquardt algorithm, the Levenberg-Marquardt algorithm is the deformation of Newton method, through the back propagation to continuously adjust the weights and thresholds of the network, so that the sum of the squared errors minimum. For a moderate number of network parameters, even if a large number of calculations are done, the LMBP algorithm is still the fastest neural network training algorithm [6].

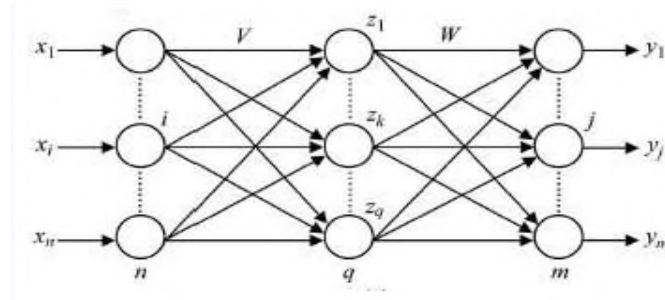


Fig. 2 BP neural network topology.

The research shows that Multilayer networks can be used for any classification problem. A two layer network is provided with sufficient neurons in the hidden layer, and the transfer function of the hidden layer neuron is S type, and can be any real function.

As the LMBP neural network has the above characteristics, so in this paper, we use the image as input, and the location of the feature points as the output, using LMBP neural network to approximate the desired input and output functions.

Sub-network partitioning. In the work of this paper, the important feature points are classified, and a neural network is trained for each class. The input image, the output of the total network sub network division all feature points, divide and rule.

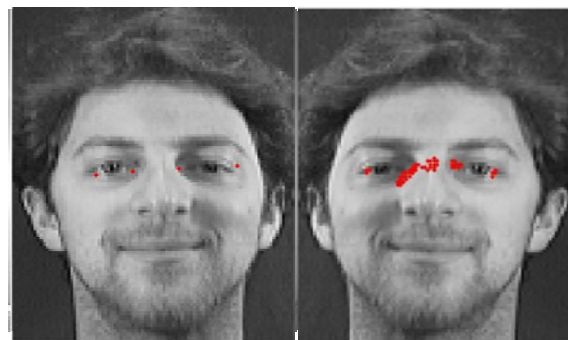
If you use a network to describe the relationship between input and output, then the difference between the output and the output of the network will be very large. Then Neural network is required to learn more input and output mode mapping relations, for the case of small training samples, the effect of network training is difficult to achieve the desired effect.

Training Sample Processing.

Input data format. This paper attempts to convert the image into two types of input data:

- (1) to the image matrix into a row vector;
- (2) in the image, each pixel is started in the domain of thirteen house, and the gray value of the pixel domain is converted to a row vector.

In order to train the eye network as an example, Fig. 3 shows the training effect of the 2 kinds of input data. You can see the results of the 1 input data is close to the ideal, while the input data of the network output of 2 training results are patchy, the effect is poor, so we use the input data in the form of 1 training neural network.



(1) Data form 1

(2) Data form 2

Fig. 3 Training effect of the 2 kinds of input data.

Artificial marking basis. The main features of this paper are as follows:

On both sides of the corner for whites in contact with the skin, eye for eye point in the middle (four points);

Nose: the nose wings curve inflection point (two points);

The tip of the nose: nose curve is also turning point (point);

Mouth: upper and lower lip two points (four points);

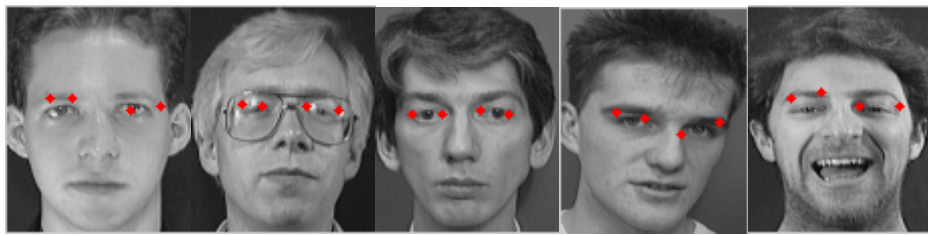
Cheeks: the outline of the two sides of the cheek (two points);

Chin: Chin Center (a little);

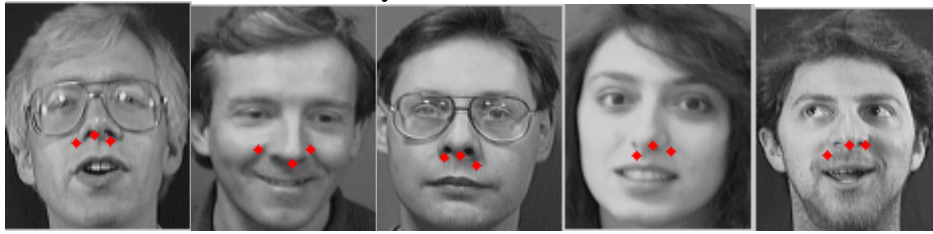
14 main characteristics.

Sub-network Training Results. Using the LMBP algorithm, the neural network is trained by MATLAB software. The 210 images are randomly selected as the training set. The 45 images are used as the test set. 100 images are used to prevent over fitting test set. 45 as a complete test set, the results are obtained as follows in Fig. 4.

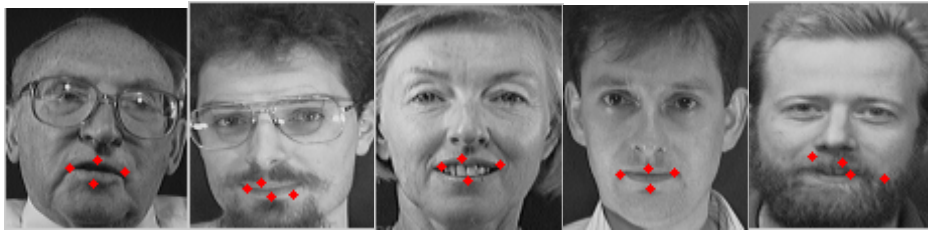
We can see that because of the training sample is too small, and the training sample is in different time, different perspective, various facial details of the conditions, the difference is huge, especially with the beard cover up, the difference between the samples is greater, the network test effect is not ideal. The calculation results of these cases slightly worse.



Eye sub network.



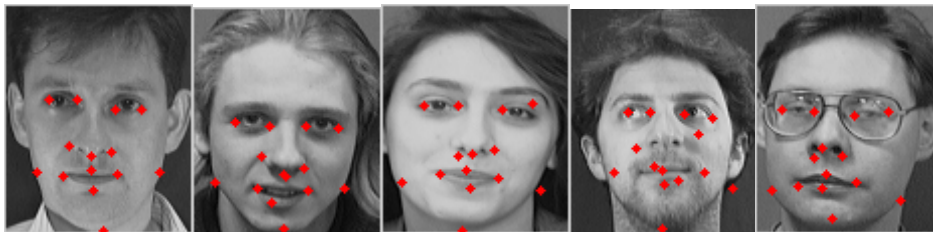
Nose sub network.



Mouth sub network.



Cheek sub network.



Overall effect.

Fig. 4 Subnet training results.

The Result Analysis

Randomly selected a picture of a OLR picture after 2.4 of sub-network calculated each feature point and marked on the image, the effect is shown in Fig. 5:



Fig. 5 Calculates feature points through sub-network.

On the picture most of the points are close to the face feature point we set, but there are individual large deviation point because of the influence of shadow. Calculation of feature point location in MATLAB, a picture of just 0.124s, so in practical application is almost equivalent to not spend time.

Discussion

The above results show that the method of dividing sub network can effectively to each feature point is positioned directly without the need for image processing in the case of less training samples and there is a particularly fast calculation speed after the trained neural network model. Artificial neural network for facial feature points localization has a good effect, but there are still some feature points due to the influence of shadow, face, beard and hair caused by inaccurate positioning, so we can processing to the picture or use the algorithm to modify the output results the future.

Summary

This paper proposes the method of neural network dividing sub network in less training samples, feature points are located on the face. Experimental results show that the artificial neural network has a good effect on the face feature point positioning, and its calculation is simple after building the model. This method can be applied to facial pose and expression analysis, face image encoding and face 3D reconstruction. Effect of the lack of training samples, illumination, hair and other factors make the artificial neural network calculated result is not accurate enough, so we need process to the picture or improved the algorithm to make the results more accurate positioning.

Acknowledgements

This work was financially supported by Innovative Research Team of the department of Sichuan Province (15TD0050) and the Fundamental Research Funds for Central University, Southwest University for Nationalities(13NZYQN04), also was financially supported by students innovation training project of SWUN(S20151065006).

References

- [1] T. Kawaguchi, D. Hidaka, M. Rizon, Detection of eyes from human faces by Hough transform and separability filter, P. IEEE Int. Conference Image P. IEEE Comput. Soc. Press, 1 (2000) 49-52.
- [2] W. C. Zhang, H. Chen, P. Yao, et al. Precise eye localization with AdaBoost and fast radial symmetry // Proc of International Conference on Computational Intelligence and Security, 2006: 725-730.
- [3] X. M. Liu, Generic face alignment using boosted appearance model //Proc of IEEE Conference

Computer Vision and Pattern Recognition, 2007: 1-8.

[4] L. Lin, X. Rong, W. Fang, et al. Face alignment via component-based discriminative search //Proc of the 10th European Conference on Computer Vision, 2008, 72-85.

[5] J. N. Hu, Y. Li, W. H. Deng, et al. Locating facial features by robust active shape model //Proc of the 2nd IEEE International Conference on Network Infrastructure and Digital Content, 2010: 196-200.

[6] P. F. Xiong, L. Huang, C. P. Liu, Initialization and pose alignment in active shape model / //Proc of the 20th International Conference on Pattern Recognition, 2010, 3971-3974.

[7] H. C. Lu, W. G. Shi. Skin-active shape model for face alignment //Proc of Computer Graphics, Imaging and Vision Conference, 2005, 187-190.

[8] T. Cootes, G. Edwards, C. Taylor. Active appearance models. IEEE Trans Pattern Anal. Mach. Intelligence. 23(6) (2001) 681-685.

[9] <http://am.ac.uk/research/dtg/attarchive/facedatabase.html>