

Face Expression Recognition Using Artificial Neural Network (ANN) Model Back Propagation

Aris Haris Rismayana¹
{rismayana@poltektedc.ac.id}

Politeknik TEDC¹, Jl. Politeknik - Pasantren KM. 2, Cibabat, Cimahi Utara, Kota Cimahi¹

Abstract. Humans can detect a facial expressions of any person easier, because its naturally directly recognized, but its very difficult to do by machine or computer. There are 3 main stages when designing a facial expression recognition system, that is, face detection (recognizes faces), extraction of the facial expression information (feature extraction which is separate parts of the face that has information about facial expressions) and the last is the classification of the expression[1]. This research facial expression, specially for smile and not smile expression, recognition using Artificial Neural Network algorithm with Back Propagation models and optimization using Principal Component Analysis. The accuracy of the results obtained to predict the smile image is equal to 81.67%, while the accuracy for predicting not smile image is 61.67%.

Keywords: Artificial Neural Network, Back Propagation, PCA, Principal Component Analysis, Facial Expression.

1 Introduction

First and very important steps to recognize an expression is detect faces from the first sequence of images. The next step is to develop a mechanism for extracting facial features from the observed facial image sequences. The facial features are a feature that stands out from the various parts of the eyebrow-face, eyes, nose, mouth, chin, etc. The feature extraction step is often referred to face tracking and its features. The final step is to develop a classifier, which would classify facial expressions into one of the basic facial expressions [2].

The recognized of facial expressions is currently becoming an interesting trend for study [3]. It can be utilized in everyday life. The most obvious example is the use of Intelligence Education System [4], or Intelligent Tutoring System [5], which essentially teaches the learning system to help the teaching and learning process virtually.

The ability of computers to understand, differentiate human expression, and take appropriate action is one of the main areas of research focus in Human Computer Interaction (HCI). Picard states that "Expression plays an important role in rational perceptions of decision making, learning, and various functions" [6]. Therefore, empowering computers to understand human expression will make computer interaction with humans more meaningful and easier. Suppose that during online learning, student acceptance of material will increase if the computer knows the state of student expression. With the knowledge of student expression, the computer can provide appropriate learning in accordance with the conditions of student expression at that time. A psychologist can easily diagnose the disease with knowledge of the patient's emotional state. Applications can be extended to missions involving

very aged, new-born people, patients with Autism etc., who will not be able to express their expressions explicitly [2] [3].

The method of learning and recognition of facial expressions can also be used to see how far humans satisfaction with something, ie, customer satisfaction with products sold by producers, so that in the end can be used as a decision maker, whether to improve the product, if the customer is not satisfied , or more improve again from the other side if the customer has been satisfied. The level of satisfaction can be informed from the introduction of customer facial expression.

To be able to recognize the facial expressions of a person, various methods have been developed. Each of these methods has both advantages and disadvantages, both in terms of accuracy and also in terms of the speed of the system to recognize facial expressions.

Some of algorithms that have been developed in recognizing facial expressions, there is :

1. Fuzzy Neural Network (FNN) algorithm [4]
2. Combination algorithm between Fuzzy Neural Network (FNN) with Hidden Markov Model (HMM) algorithm [3]
3. ActiveAppearance Models (AAM) combination algorithm with Ekman's Facial Action Coding System (FACS) base with Dynamic Bayesian Network (DBN) algorithm with Kalman Filters base. [7]
4. Naive Bayesian classification algorithm (NB) [8]
5. Algorithm of geometric deformation features and support vector machines [9]

2 Face Detection and Feature Extraction

Lajevardi and Lech [8] have proposed a method in which the face area of an image is detected using the Viola-Jones method based on Haar-Like Feature and the AdaBoost learning algorithm. The Viola and Jones Method [11] is an object detection algorithm that provides a real-time competitive object detection rate, and it is primarily designed for face detection. Features derived from pixels are selected from a rectangular area, which is worn over the image and that shows high sensitivity to vertical and horizontal lines.

The same method is also proposed for the detection of features on the face. What distinguishes is the object for detection. The observed detection items are facial features, that is, eyes, nose and mouth.

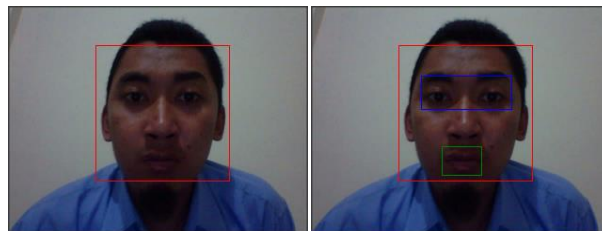


Fig. 1 Face detection and Feature extraction (mouth and eyes)

3 Principal Component Analysis (PCA)

PCA is used to reduce the number of dimensions of the observed features [12]. This method is also widely used in face recognition.

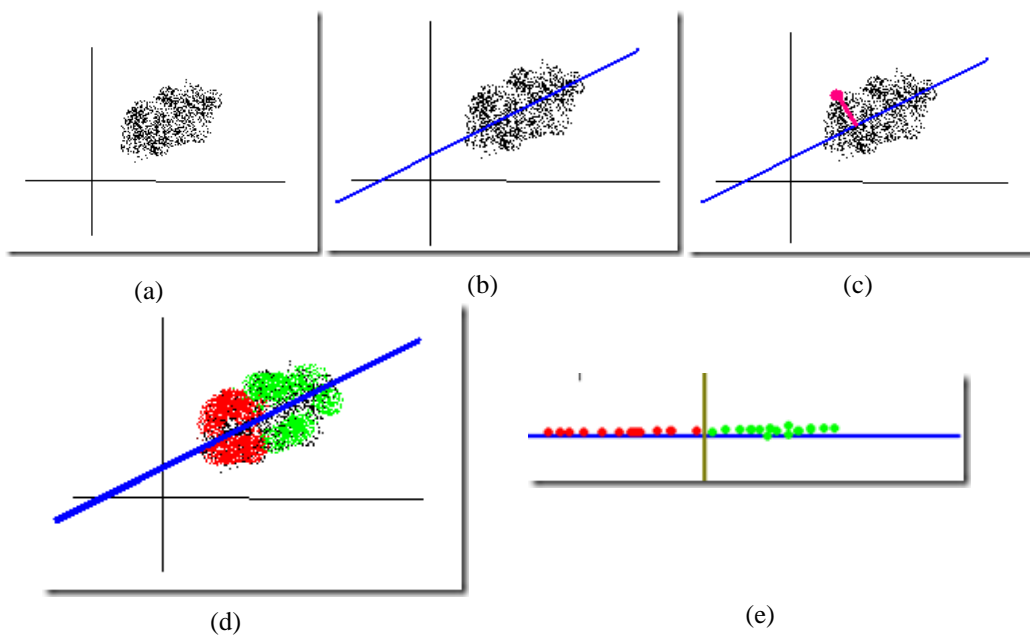


Fig. 2 PCA Process

If two dimensional data are obtained, then plotted in the graph, it will look like the figure 2(a). PCA uses decomposition of Eigen vectors. The result of the decomposition are Eigen vectors along with Eigen values. The Eigen vector that has the largest Eigen value is a vector in the direction of the field with the highest correlation. For the above example, Eigen vector with the largest eigen value as figure 2(b). If the data is projected to the eigen vector field with the largest eigen value, then the data will be one dimensional data as in figure 2(c).

With this one dimensional data will have a great variance. For the case of pattern classification, systems that process one-dimensional data can have better performance than the system processes raw data directly. For example, suppose there are two classes in the above data. Class 1 is represented in red, and class 2 is represented in green, as in figure 2(d). Figure 2(e) is the result of transforming data into one dimension.

3.1 Image Analysis with PCA

PCA analysis can be used to decipher images into a set of major components of orthogonal images also known as eigenfaces. Eigenfaces can be used for image coding, image denoising, or as a feature for image classification.

To obtain an eigenfaces $[E_{ij}]$ for the given image A with the size of $r_0 \times c_0$ pixels, the first image is divided into sub image L (sub-block) A_k with size $r \times c$ (usually 8×8 or 16×16).

The mean of the sub-images is obtained as

$$A_{mean} = \frac{1}{L} \sum_{k=0}^{L-1} A_k \quad (1)$$

The mean image is then removed from each sub-picture as

$$\bar{A}_k = A_k - A_{mean} \quad (2)$$

A $r \times r$ covariance matrix for rows of sub-images is obtained

$$C_r = \frac{1}{L} \sum_{k=0}^{L-1} \bar{A}_k \bar{A}_k^T \quad (3)$$

Similarly, a $c \times c$ covariance matrix for columns of sub-images is obtained

$$C_c = \frac{1}{L} \sum_{k=0}^{L-1} \bar{A}_k^T \bar{A}_k \quad (4)$$

The lines and columns of the covariance matrix then undergo eigenanalysis to produce eigenvectors of r row [r_i ; $1 \leq i \leq r$] and eigenvector vectors [c_j ; $1 \leq j \leq c$].

A $r \times c$ eigenfaces E_{ij} is defined as the product of each eigenvector row with each eigenvector column as

$$E_{ij} = r_i c_j^T \quad 1 \leq i \leq r, 1 \leq j \leq c \quad (5)$$

4 Artificial Neural Network Model Backpropagation

Artificial Neural Network Model Backpropagation is a systematic method of multilayer training of artificial neural networks. This method has a strong, objective mathematical base and this algorithm derives the form of equations and coefficient values in the formula by minimizing the sum of the squares of error error through the developed model (training set).

1. Starting with the input layer, calculate the output of each processing element through the outer layer.
2. Calculate the error on the outer layer which is the difference between the actual data and the target.
3. Transform the error to the appropriate error on the input side of the processing element.
4. Propagation of these errors on the output of each processing element to the error contained in the input. Repeat this process until input is reached.
5. Change the entire weights by using the error on the input side of the element and the output of the connected processing element.

Backpropagation learning method uses the average squared performance index or Mean Square Error. Average squared errors can be obtained from:

- a. Quadratic error divided by number of output components.
- b. The total squared error is divided by the amount of training data.

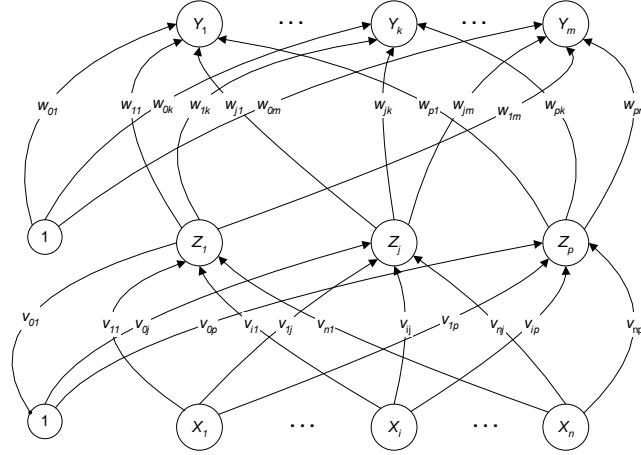


Fig. 2 Artificial Neural Network Backpropagation with 1 hidden layer

In the learning process, convergence will be achieved more quickly if momentum is added to the weight renewal formula. To use momentum, weights (renewal of weights) from one or more previous learning patterns should be stored. For example, the simplest form of Backpropagation with momentum, the new weight for the learning step $(t + 1)$ is based on the weights in the learning steps (t) and $(t - 1)$. The mathematical formulation for Backpropagation with momentum is:

$$w_{jk}(t+1) = w_{jk}(t) + \alpha \delta_k z_j + \mu [w_{jk}(t) - w_{jk}(t-1)] \quad (6)$$

or

$$\Delta w_{jk}(t+1) = \alpha \delta_k z_j + \mu \Delta w_{jk}(t) \quad (7)$$

and

$$v_{ij}(t+1) = v_{ij}(t) + \alpha \delta_j x_i + \mu [v_{ij}(t) - v_{ij}(t-1)] \quad (8)$$

or

$$\Delta v_{ij}(t+1) = \alpha \delta_j x_i + \mu \Delta v_{ij}(t) \quad (9)$$

where the momentum parameter μ is limited to the range 0 to 1.

4.1 Parameters ANN Model Backpropagation

The network architecture used is Backpropagation Neural Network with momentum. The number of hidden layers used only one layer, with the number of processing elements (nodes) in each layer are as follows:

- In the input layer, two processing elements are used, which are used to handle inputs for systems that present a smile and not a smile.
- In the hidden layer, two processing elements are used.
- In the output layer, one processing element is used. In this output element element, outputs that are smaller than 0.5 to 0 will present as a smile, while values between 0.5 and 1 are used to present not a smile.

In the hidden units and the output of this Artificial Neural Network there is bias. Since the network's output value is in the range of 0 and 1, the activation function used in this Artificial Neural Network is a Binary Sigmoid function, which has a range of 0 to 1 values.

In the input layer to the hidden layer we use the weight matrix with the size = number of input nodes x the number of hidden nodes. In the hidden layer to the output layer we use the weight matrix with the size = number of hidden nodes x the number of output nodes. Bias for hidden layer of vector with size = 1 x number of hidden nodes. While the bias for output layer is vector with size = 1 x number of output nodes. At initialization time weights and biases, matrix elements of weights and biased vectors, are randomly assigned between -0.5 to 0.5, with no element of value 0, to avoid the occurrence of the result of the hidden layer of value 0.

5 Research Methods

This research is experimental research with research method as follows:

1. Problem Determination

Determination of the research problem is by using literature study and field study.

2. Determination of computing approach

Computing approach in this study was selected based on a literature study of the state of the art facial recognition expression. In this research the method of Artificial Neural Network Back Propagation model was chosen for modeling expression classification.

As an optimization for input into the ANN algorithm used PCA (Principal Component Analysis) method. PCA is used to reduce the image feature dimension before being processed by ANN.

3. Application of Viola-Jones method

Used to detect face (face detection), also application of Haar-Like Feature Method, to detect facial features, and application of Principal Component Analysis to uniform dimension of feature image as input material of artificial neural network process.

4. Testing by comparing the object image.

Testing the accuracy of the application of Artificial Neural Network Back Propagation model to determine the expression on the face, will be done by comparison between objects that have been known expression.

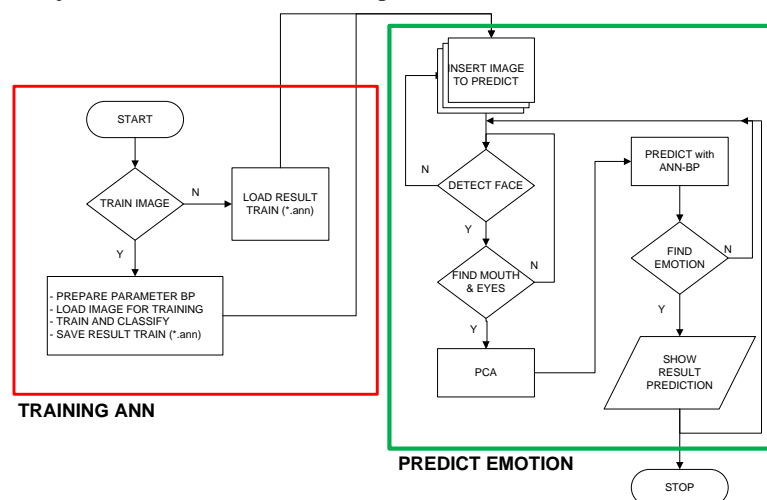


Fig. 3 Flow chart modeling of Artificial Neural Networks to predict expression

6 Testing

At this stage two facial features are taken as input for processing using ANN. These features are eye features and mouth features. Before taking these features, face detection is done first, if the face is detected, then the features of the eyes and mouth will be taken for further processing.

Face detection and features using Haar-Like Feature method or also known as Viola-Jones method. After the mouth and eyes feature is obtained, the next step is to prepare the feature to be input on the ANN, for the expression classification process.

Because of the difference in pixels between the features of the eyes and the features of the mouth, where the eye feature's features are 94x35 pixels, while the mouth feature dimension is 60x36 pixels, so there must be pixel uniformity in order to be further processed on the ANN.

To uniform pixels on these features, the PCA (Principal Component Analysis) method is used, so that the dimensions of the mouth and eye features are made equal, that is 20x20 pixels. The mouth and eye features obtained from the above feature extraction process are used as input for the ANN algorithm.

The proposed ANN algorithm uses the Back Propagation method, which consists of 1 input layer, 2 hidden layers and 1 output layer. The input layer is input with the number of 800 nodes. This number of neurons is derived from the number of matrix results of the reduction of the PCA method. The 800 neuron nodes derived from the eye and mouth features that have been reduced by the PCA, so that each feature is converted into a 20x20 order matrix. From each matrix obtained 400 data matrix with float data type. Because there are 2 pieces of features, so the number of inputs to be processed by the ANN 400 x 2, ie as many as 800 data float, or 800 node neuron.

While on the hidden layer each there are 1000 neuron nodes and 400 neuron nodes, and at the output layer there is 1 node.

6.1 Artificial Neural Network Training

In this module part, described explanations concerning training on artificial neural networks. These stages include:

1. Determine the parameters that must be met for the ANN method
Determination of parameters include determining the size of the epoch, the number of input points (neurons), the number of hidden layers, determine the activation function and determination of learning rate.
The parameters are determined as follows:
 - a. Neuron input point = 800 node
 - b. Hidden Layer = 2, each is 1000 node
 - c. Activation function = symmetric sigmoid
 - d. Learning rate = 0.5
 - e. Epoch = 5000
2. Preparing training data, which is the image that has been done classification, ie data smile or not smile. The picture is prepared as many as 20 pictures, consisting of 10 smile pictures and 10 pictures no smile.
3. Conducting training with parameters specified above, and using image data that has been prepared
Save the results of training data, so that if necessary the training data does not have to do steps 1 - 3 as above.

6.2 Expression prediction

In the last part, described explanations concerning the expression prediction of the images provided by using artificial neural networks. Stages include:

1. Face detection of the given image, then broken down into several features that become input for analysis of expression determination. The feature consists of mouth and eye features.
2. Reduction of feature dimension using PCA method so that feature dimension to be same and can be devoted to next method, that is input method of ANN.
3. The expression prediction of the images given based on the previous training results, then show the predictions.

7 Result

After training so that ANN has a pattern to recognize the object to be observed, the next step is to test against objects that have been recognized expression. Testing done to 60 image smile and 60 image not smile, the result is as follows:

1. During the test to predict smile image, produced as many as 49 mage in accordance with the classification, the expression of a smile. While the rest, 11 images, can not recognize the expression of a smile. So that obtained an accuracy of 81.67% in recognizing the expression of a smile.
2. During the test to predict not smile image, produced as many as 37 mage in accordance with the justification, the expression is not a smile. While the rest, 23 images, can not recognize. So we get an accuracy of 61.67% in observing expression rather than smiles.

Table 1 Results Accuracy of Face Expression

Image	Expression	Predict	Not Predict	Accuracy (%)
60	Smile	49	11	81.67
60	Not Smile	37	23	61.67

8 Conclusion

From the results of research conducted by using artificial neural network model back propagation, and also using the optimization of Principal Component Analysis (PCA), can be drawn the following conclusion:

1. The use of artificial neural network model back propagation and PCA optimization in recognizing expression in this study has not obtained significant results. This can be seen from the accuracy of the expression recognition of a smile and not a smile.
2. Accuracy result obtained to predict smile image is equal to 81.67%, while accuracy to predict image not smile is 61.67%.
3. Some of the things that affect the prediction result of this artificial neural network algorithm are:
 - a. Image quality. Level of image lighting, image taking position.

- b. Determination of artificial neural network parameters. Parameters include, epoch, learning rate, the number of hidden layers, momentum and others.
- c. Position of image capture feature. In the implementation, sometimes get an imperfect mouth feature.
- d. Quality and quantity of image used for training.

References

- [1] R. a Patil, V. Sahula, and a. S. Mandal, "Automatic recognition of facial expressions in image sequences: A review," 2010 5th Int. Conf. Ind. Inf. Syst., pp. 408–413, Jul. (2010).
- [2] M. Pantic, S. Member, and L. J. M. Rothkrantz, "Automatic Analysis of Facial Expressions : The State of the Art," vol. 22, no. 12, pp. 1424–1445, (2000).
- [3] Y. Guo and H. Gao, "Emotion Recognition System in Images Based On Fuzzy Neural Network and HMM," 2006 5th IEEE Int. Conf. Cogn. Informatics, pp. 73–78, Jul. (2006).
- [4] X. Pan, "Research on the Emotion Recognition Based on the Fuzzy Neural Network in the Intelligence Education System," 2011 Second Int. Conf. Digit. Manuf. Autom., no. 3, pp. 1030–1033, Aug. (2011).
- [5] X. Liu, L. Zhang, J. Yadegar, and N. Kamat, "A Robust Multi-Modal Emotion Recognition Framework for Intelligent Tutoring Systems," 2011 IEEE 11th Int. Conf. Adv. Learn. Technol., pp. 63–65, Jul. (2011).
- [6] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion Recognition: a review," 2011 IEEE 7th Int. Colloq. Signal Process. its Appl., pp. 410–415, Mar. (2011).
- [7] K. Ko and K. Sim, "Development of a Facial Emotion Recognition Method based on combining AAM with DBN," (2010).
- [8] S. Lajevardi and M. Lech, "Facial expression recognition from image sequences using optimized feature selection," *Image Vis. Comput. New ...*, (2008).
- [9] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *Image Process. IEEE Trans.*, vol. 16, no. 1, pp. 172–187, (2007).
- [10] K.-E. Ko and K.-B. Sim, "Development of a Facial Emotion Recognition Method Based on Combining AAM with DBN," 2010 Int. Conf. Cyberworlds, pp. 87–91, Oct. (2010).
- [11] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Comput. Vis.*, no. February, (2001).
- [12] O. Ozgonenel and T. Yalcin, "Principal Component Analysis (PCA) based neural network for motor protection," 10th IET Int. Conf. Dev. Power Syst. Prot. (DPSP 2010). *Manag. Chang.*, pp. 101–101, (2010).