Facial Expression Recognition Using Backpropagation

Indra Adji Sulistijono, Zaqiatud Darojah
Abdurahman Dwijotomo, Dadet Pramadihanto

Electronics Engineering Polytechnic Institute
of Surabaya (EEPIS), Indonesia
{indra, zaqiah, dadet}@eepis-its.edu

Naoyuki Kubota

Graduate School of System Design
Tokyo Metropolitan University, Japan
kubota@tmu.ac.jp

Abstract: This paper proposed facial expression recognition using backpropagation of neural network. The procedures are image capturing, face detection, filtering, facial feature extraction and recognition using neural network. We propose forehead, mid forehead, mouth and cheek for the input of facial feature extraction. We define six output of facial expression, i.e.: anger, disgust, surprise, happiness, sadness and fear. Then, we trained the data of image captured using backpropagation of neural network. The proposed method can recognize the facial expression from image captured well.

1 INTRODUCTION

Object recognition is one of the development of image processing study that has many extensive application areas. One of the application is it can be used to recognize face of a human with its features, hence, it also can be developed to recognize a human facial expression. Facial expression is interesting research concerning with a human behavior recognition. It considered to be one of the most powerful and immediate means for human to communicate their emotions, intentions and opinions to each other and this is the why much efforts have been devoted to their study by cognitive scientists an lately computer vision researchers [1]-[3]. In this perspective, a computer as if can feels what a human feels and it will create an interaction between a computer and a human like a friend.

Many experiments have been reported toward facial expression recognition [1],[4]. In this work, the motion information is used to detect the face in real time video which is obtained from a camera. The facial detection is used to extract the features which is useful to recognize a human expression and a human emotion later. There are many ways to extract the facial features to be input in facial expression recognition. Majority, the object used to be input is difficult to interpret, such as, the change of the eyebrows [1], it is almost no change between expression happy, sad, and normal. The schema that we propose in this work is very simple than any other method before. The facial features which extracted in this work are forehead wrinkle, mid forehead wrinkle, cheek wrinkle and mouth length. These facial features are easy to be recognized when a face does an expression. The method assumes that there are no mustache, no beard, and no glasses on a facial human.

The other facial expression recognition is along with the face detection and facial feature points detection module [8]. Faces are deformed by their facial expression. The ways of face change caused by their expression can be measured from the change in locations of facial feature points. The facial expression and the degree of facial expression change depend on the displacement of the facial expression features points from the neutral face. Modeling is the relationship between the degree facial expression change and the displacement of the facial feature points by the use of a B-spline curve. Then a flexible graph matching for tracking these facial features from an input image sequence was used, while matching the trajectory of the features with the expression change models to determine the category and degree of expression change model in the image sequences.

At first, it is important that the machine can detect and tracking the face [9]. This procedure is used for capturing the moving face. Next procedure is extracting the image captures by camera. The method to recognize the facial expression in this work is backpropagation of feedforward neural network. We define input and target as training pairs. The input are the data from forehead wrinkle, mid forehead wrinkle, cheek wrinkle and mouth length, while six output of facial expressions are defined, i.e. anger, disgust, surprise, happiness, sadness and fear to be recognized. After we train it, the proposed method can recognize the facial expression from image captured well.

2 FACIAL FEATURE EXTRACTION

The face of a human has several features such as, mouth, eyes, nose, eyebrows, and forehead. Each of this features has a unique shape and a unique pattern, hence, many experiments have been reported in extracting facial feature for recognizing facial expression. A. Geetha et. al. (2007) [4] used the locations of eyes as the visual features of face. Yh- yeong Chang et.al. (2001) [1] used eyebrows, eyes, and mouth for facial expression labeling. In this paper, we extract four main features: forehead, mid forehead, cheek and mouth for facial expression labeling. We extract the features toward forehead wrinkle, mid forehead wrinkle, cheek wrinkle, and mouth length as seen in Figure 1.

2.1 Line Face Detection

Facial features extracted using edge detection and morphology technique to obtain the lines on the face. We used Canny edge detection. Before applying Canny edge detection method, at first, the images should be optimized using brightness and contrast tuning as shown in

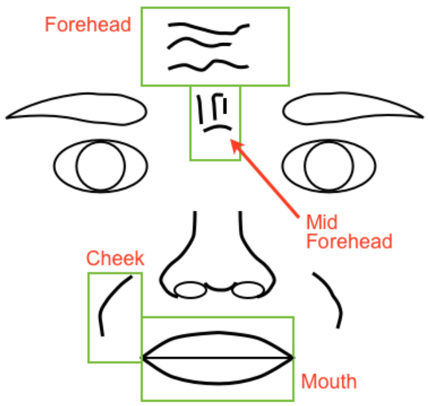


Fig. 1: Parts of Facial Feature Extracted

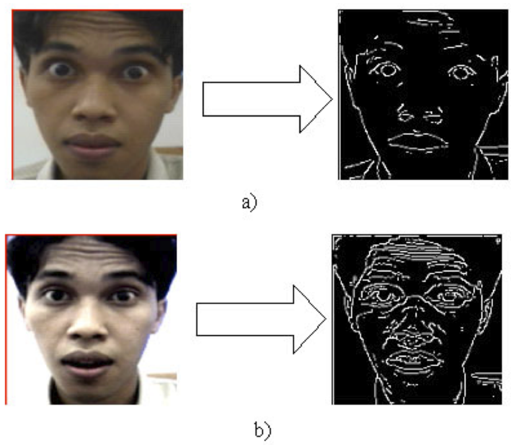


Fig. 2: The image processing (a) Canny edge detection with no contrast and brightness tuning, (b) Canny edge detection with contrast and brightness tuning

Figure 2. The aim of this optimizing is to detect the vague lines such as the face wrinkle.

2.1.1 Canny Edge Detection

The Canny edge detector uses a filter based on the first derivative of a Gaussian, because it is susceptible to noise present on raw unprocessed image data, so to begin with, the raw image is convolved with a Gaussian filter. The result is a slightly blurred version of the original which is not affected by a single noisy pixel to any significant degree [7].

The next is finding the intensity gradient of the image. An edge of an image may point in a variety of directions, so the Canny algorithm uses four filters to detect horizontal, vertical and diagonal edges in the blurred image. The edge detection operator (Roberts, Prewitt, Sobel for example) returns a value for the first derivative in the horizontal direction (G_y) and the vertical direction (G_x). From this the edge gradient and direction can be determined:

$$|G| = |G_x| + |G_y| \tag{1}$$

next is finding the edge direction. The formula for finding the edge direction is:

$$\theta = inv\ tan(G_y/G_x) \tag{2}$$

The edge direction angle (θ) is rounded to one of four angles representing vertical, horizontal and the two diagonals ($0^o, 45^o, 90^o$ and 135^o for example). Once the edge direction is obtained, the next step is related the edge direction to a direction that can be traced in an image. Finally, hysteresis is used as a means of eliminating streaking. Hysteresis uses two thresholds, a high and a low. Any pixel in the image that has a value greater than T_1 is presumed to be an edge pixel, and is marked as such immediately. Then, any pixels that are connected to this edge pixel and that have a value greater than T_2 are also selected as edge pixels.

2.1.2 Morphology Technique

For morphology, we use two operations, dilation and erosion for getting a potential area for edge intensity. The dilation and erosion operators for grayscale images are defined conventionally [6]. We define the edge intensity ϕ_{edge} as

$$\phi_{edge} = dilation - erosion \tag{3}$$

The data from the lines on face feature, such as mouth, are obtained using morphology technique because using edge detection can cause many of noises. While the data from the face wrinkle easy to be applied by usual edge detection, such as Canny edge or Sobel edge detection. Once the lines is obtained, the next step is counting the number of the wrinkle and the result is as percentage how serious a face did an expression.

2.2 Forehead Extraction

Forehead has a wide area on face, the location of it is one third upper part of face. By this condition, we can put a sign into the location of forehead easy as a rectangular box. The size of this rectangular has been identified and it can not change. The location of this rectangular box which is used to narrow and simplify the forehead wrinkle processing from a whole face. The image is simplified to be grayscale first, and then the data from the forehead wrinkle can be obtained by Canny edge detection. Figure 3 describes how the procedures of forehead extraction given.

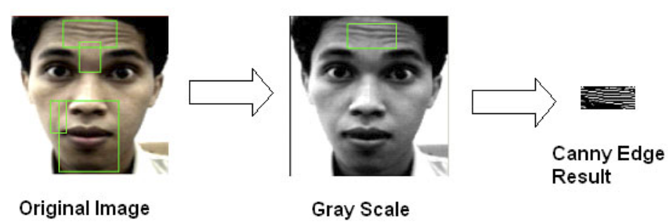


Fig. 3: The procedures of forehead extraction

Once the forehead wrinkle, represented in columns and rows pixel, is obtained using edge detection, the number of the wrinkle is counted. The shape of forehead wrinkle is always in horizontal, hence, other lines are classified as noise. To reduce the noise, all columns from the image of forehead can be simplified to be some columns. Each column has one search line which used to find the wrinkle vertically as describing in Figure 4. The wrinkle which detected is signed by red dot. Later, these red dots are

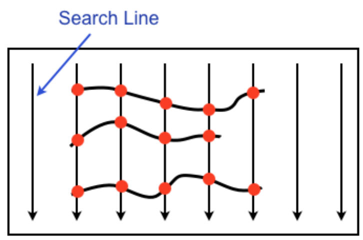


Fig. 4: The procedures of counting the Forehead wrinkle

counted and scaled value to 0-1 range for neural network input.

The search line that used in this work is 17 lines and is set in 17 columns. From the experiment, the maximum of red dot ($RDmaxInForeHead$) is about 30. The value of Forehead Wrinkles (FHW) which has 0-1 range is defined by

$$FHW = \frac{TotalRedDotInForeHead}{RDmaxInForeHead} \tag{4}$$

The result of this computation is used to be input of neural network.

2.3 Mid Forehead Extraction

The step to obtain the data from mid forehead wrinkle is almost same with the step to obtain the data from forehead contraction. First, we should find the location of the mid forehead. The location of mid forehead is in between two eyebrow. It is under forehead location in the middle area. Then, this location is signed by rectangular box which has size 21 x 29 pixels and is fixed. The mid forehead wrinkle is more vague than the forehead wrinkle, hence using such Sobel edge detection method is difficult to apply in this case. We use Canny edge detection to obtain the data from mid forehead wrinkle as describing in Figure 5.

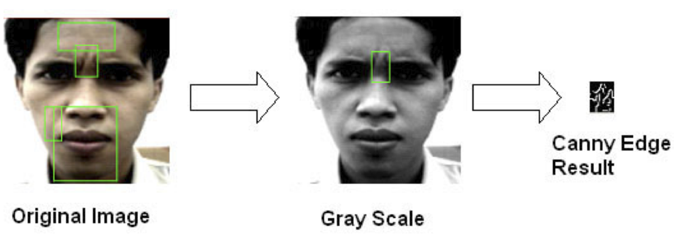


Fig. 5: The procedures of mid forehead extraction

The step to extract mid forehead wrinkle is also almost same with the step to extract forehead wrinkle before. Figure 6 describes how to count the mid forehead wrinkle. The shape of mid forehead wrinkle is always in vertical, hence, other lines are classified as noise. To reduce the noise, all rows of the image of mid forehead can be simplified to be some rows. Each row has also one search line which used to find the wrinkle horizontally. The wrinkle which detected is signed by red dot. Later, these red dots are counted and scaled value to 0-1 range for neural network input.

The search line that used in finding mid forehead wrinkle is 7 lines and is set in 17 columns. From the experiment, the maximum of red dot (*RDmaxInMidForeHead*) is about 18. The value of Mid Forehead Wrinkles

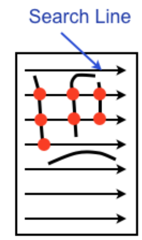


Fig. 6: The procedures of counting the Mid Forehead wrinkle

($MFHW$) which has 0-1 range is defined by

$$MFHW = \frac{TotalRedDotInMidForeHead}{RDmaxInMidForeHead} \tag{5}$$

The result of this computing is used to be input of neural network.

2.4 Mouth Extraction

The first step to extract the data from mouth is determining the location of the mouth. The methods have been proposed in the literature for determining the location of the eyes. Location-based approach is commonly used for the locations the eyes. It's location is one-third of lower facial width and it is under the hole of nose. The work area of mouth is various because it can become wide and also become long. After we obtained the exact location of the mouth, a rectangular box 57 x 66 pixel is signed which can become the region of interest. Using edge detection either Canny edge or Sobel edge is difficult to apply, because the color of the edge of mouth is vague and it is near to the skin of face. From the experience of some technique, morphology technique is the best to extract the lines in the edge of mouth, although there are still small noises surrounding the mouth. To solve this problem, threshold color and Gaussian smoothing are used. Figure 7 illustrates the procedures of mouth extraction.

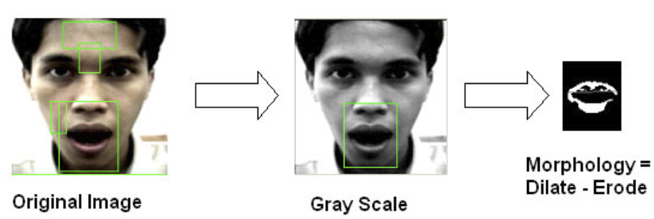


Fig. 7: The procedures of mouth extraction

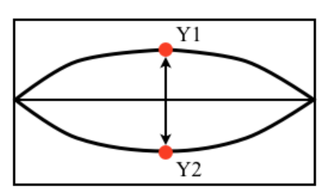


Fig. 8: The procedures of counting the mouth length

The length of mouth is obtained by finding upper dot (Y1) and lower dot (Y2) of mouth. The distance of the two dots is obtained from the gaping mouth as shown in Figure 8. The mouth length is then scaled into 0-1 range for the input of neural network. From our experiments, the maximum value of mouth length ($VMLg$) is 20. The value of $VMLg$ is defined by

$$VMLg = \frac{(Y2 - Y1)}{VMLg} \tag{6}$$

2.5 Cheek Extraction

The location of cheek is inside of mouth. After detecting the location of the cheek, the image of cheek is extracted into grayscale form. This location is then signed by rectangular box which has size 15 x 30 pixel. Because of there are much of noises around of cheek, we use Gaussian smoothing technique to obtain cheek wrinkle. Figure 9 describing the procedures of cheek extraction.

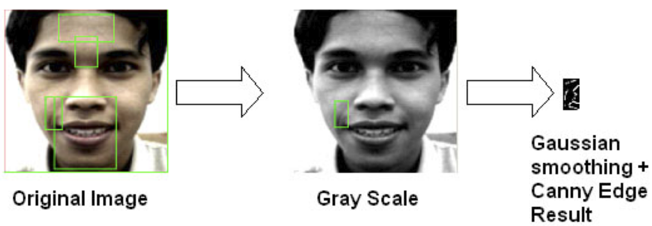


Fig. 9: The procedures of cheek extraction

The step to obtain the cheek wrinkle is the same with the steps before. Each search line find the wrinkle which vertical. Each wrinkle detected is then signed by red dot. All of these dots is counted and scaled to 0-1 range for neural network input. Figure 10 shows the procedures of counting the cheek wrinkle.

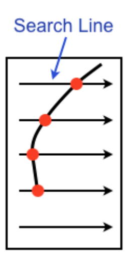


Fig. 10: The procedures of counting the Cheek Wrinkle

The search line that used for finding cheek wrinkle is 7 lines and is set in 7 rows. From the experiment, the maximum of red dot ($RDmaxInCheek$) is about 18. The value of Cheek Wrinkles (CkW) which has 0-1 range is defined by

$$CkW = \frac{TotalRedDotInCheek}{RDmaxInCheek} \tag{7}$$

The result of this computing is used to be input of neural network.

3 FACIAL EXPRESSION RECOGNITION

We use backpropagation algorithm to recognize of facial expression with feedforward architecture. Backpropagation neural networks are the most widely used network and are considered the work horse of artificial neural network [5],[6]. It can be used to model complex relationships between inputs and outputs or to find patterns in data.

The backpropagation of feedforward architecture is designed based on facial features extracted as illustrated in Figure 11. It consists of (1) an input layer containing four neurons representing input variable to the problem, that is extracted data from the forehead wrinkle, the mid forehead wrinkle, the cheek wrinkle, and the mouth length; (2) one hidden layers containing one or more neurons to help capture the nonlinearity in the data; and (3) an output layer containing six nodes representing output variable to the problem, that is facial expressions : anger, disgust, surprise, happiness, sadness and fear. The neurons between layers are fully interconnected with weight v_{ij} and w_{ij}.

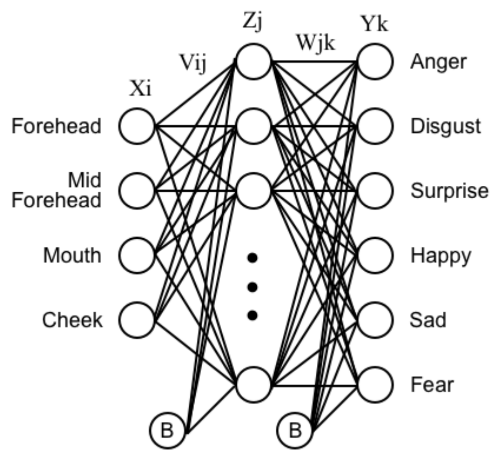


Fig. 11: Architecture of feedforward backpropagation neural network for facial expression recognition

The training of a network by backpropagation neural network involve three stages: the feedforward of the input training pattern, the calculation and backpropagation of associated error, and the adjustment of the weights [6]. The data are fed forward from the input layer, through hidden layer, to output layer without feedback. Then, based on the feedforward error-backpropagation learning algorithm, backpropagation will search the error surface using gradient descent for point(s). Based on the error, the portion of error correction is computed, and then the weights for all layers are adjusted simultaneously.

In many neural network applications, the data (input or target patterns) have the same range of values [6]. We use the binary sigmoid function, which has range of (0,1) and is defined as $f(x) = 1/(1 + exp(-x))$, that's why the data is also represented in binary form or has range of 0-1. The representation data of input is explained in the section before. Table 1 shows the data of training pairs (input and target patterns) in backpropagation of neural network. We use two pairs of training input data for each of six output expressions. The first row is for neutral expression.

Table 1: The data of training pairs in backpropagation of neural network

Input				Output					
ForeHead	Mid ForeHead	Mouth	Cheek	Anger	Disgust	Surprise	Happy	Sadness	Fear
0	0	0	0	0	0	0	0	0	0
0	1	0	0	1	0	0	0	0	0
0	1	1	0	0	1	0	0	0	0
1	0	1	0	0	0	1	0	0	0
0	0	0	1	0	0	0	1	0	0
1	1	0	0	0	0	0	0	1	0
1	1	1	0	0	0	0	0	0	1
0	0.5	0	0	0.5	0	0	0	0	0
0	0.6	0.2	0	0	0.5	0	0	0	0
1	0.15	0	0	0	0	0.5	0	0	0
0	0	0	0.3	0	0	0	0.5	0	0
0.8	0.2	0	0	0	0	0	0	0.5	0
1	0.5	0.2	0	0	0	0	0	0	0.5

4 EXPERIMENTAL RESULTS

The proposed method was implemented on a personal computer with Pentium dual core 3 GHz CPU and 2 GB RAM. First, we do the training for the all training pair . The aim is to train the net to achieve a balance between the ability to respond correctly to the input patterns that are used for training (memorization) and the ability to give good responses to the input that is similar, to the used in training. After training, the backpropagation neural net is applied by using only the feedforward phase of the training algorithm.

Table 2 shows a snapshot of the real time facial expressions as input with the output result. There are 2 persons of pair data, the first person is with 6 facial expressions and the second person is with 5 facial expressions. The input data range is 0-1 and the output data is percentage. The lowest percentage of expression is sadness (32.4%) for the first person and disgust for the second person (45.5%). The expression percentage of disgust and anger showed that those expression have high similarity for the first person (the 2^{nd} row from top), while the percentage expression of happy and disgust have high similarity for the second person (the 2^{nd} row from bottom). The highest percentage of expression is happy (83.6%) for the first person and sadness for the second person (95.8%)

It shows that the six expressions identified can be recognized well based on the memory of the weight which learned from face features with higher percentage of the appropriate expression.

5 CONCLUSIONS

In this paper, a simple method for facial expression recognition using backpropagation was proposed . The experimental results show that backpropagation algorithm with the facial features extracted method can recognize well the appropriate facial expressions with the higher percentage than another facial expressions. The expression of sadness and disgust are more difficult than the others to recognize.

Generally speaking, online and spontaneous expression recognition is a difficult task. We focus to tackle the recognition of subtle spontaneous facial expressions. Furthermore, we would like to apply unsupervised learning with an online clustering technique, and to estimate the intensity of facial expressions.

References

[1] Chang Jyh-Yeong, Chen Jia Lin, Automated Facial Expression Recognition System Using Neural Networks, Journal of the Chinese Institute of Engineers, Vol. 24, No.3, pp.345-356, 2001

[2] P. Ekman, W.V. Friesen, The Facial Action Coding System: A Technique for measurement of facial movement, Consulting Psychologists Press, Palo Alto, CA, 1978

[3] M. Pantic, L, Rothkrantz, Automatic Analysis of Facial Expression: the state of the art, IEEE Transaction on Pattern Analysisi and Machine Intelligence 22 (12)(2000), pp.1424-1445, 2000

[4] A. Geetha, V, Ramalingam, S. Palanivel, B. Palaniappan, Facial Expression Recognition-A real time approach, Elsevier:Expert System with Applications 36 (2009), 303-308, 2009

[5] Te-Hsiu Sun, Fang-Chih Tien, Using Backpropagation Neural Network for Face Recognition with 2D + 3D Hybrid Information, Elsevier:Expert System with Applications 35 (2008), pp.361-372, 2008

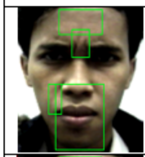
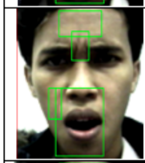
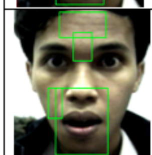
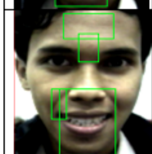
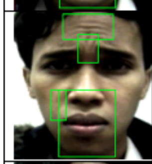
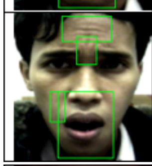
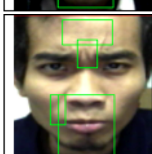
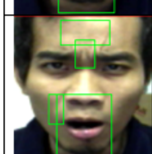
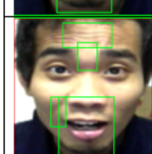
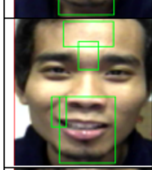
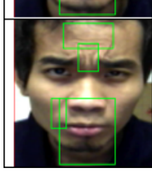
[6] Fausett Laurene, Fundamentals of Neural Networks: Architectures, Algorithms, and Applications, Prentice Hall, 1994

[7] Gonzales Rafael C., Woods Richard E., Digital Image Processing, Addison-Wesley Publishing Company, Inc, New York, 1993

[8] D.Pramadihanto, Y.Iwai and M.Yachida, Integrated face identification and facial expression recognition, IEICE Transaction on Information and Systems, vol.E84-D, no.7, pp.856-866, 2001

[9] Indra Adji Sulistijono, Naoyuki Kubota, Human Head Tracking Based on Particle Swarm Optimization and Genetic Algorithm, Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), Vol. 11, No. 6, Fuji Technology Press Ltd., July, 2007, pp.681-687.

Table 2: The experimental result of facial expression using forehead, mid forehead and mouth as input with six facial expressions as output (%)

Input 0~1					Output					
Original Image	Fore Head	Mid Fore Head	Mouth	Cheek	Anger %	Disgust %	Surprise %	Happy %	Sadness %	Fear %
	Val= 0	Val= 0.55	Val= 0	Val= 0.15	73	0.4	0	2.0	0.7	0.2
	Val= 0	Val= 0.75	Val = 0.3	Val = 0	18.9	35.7	0	0.2	0.2	2.0
	Val= 0.48	Val= 0	Val= 0.05	Val= 0	0.0	0.0	42.4	0.2	0.1	0.8
	Val= 0.04	Val= 0.3	Val= 0.1	Val= 0.6	1.6	1.5	0.1	83.6	0.0	0.0
	Val= 1	Val= 0.45	Val= 0	Val= 0.2	0.0	0.0	7.9	0.1	32.4	1.0
	Val= 1	Val= 0.7	Val= 0.2	Val= 0	0.0	0.0	0.2	0.0	20.0	52.7
	Val= 0.04	Val= 0.5	Val= 0	Val= 0.01	46.9	1.5	0.0	0.5	0.7	0.0
	Val= 0.08	Val= 0.6	Val= 0.25	Val= 0	0.7	45.5	0.0	0.0	0.3	0.1
	Val= 1	Val= 0.2	Val= 0.25	Val= 0.3	0.0	0.0	91.7	0.0	0.0	0.4
	Val= 0	Val= 0.2	Val= 0.3	Val= 0.7	0.1	41.9	0.3	76.8	0.0	0.0
	Val= 1	Val= 0.8	Val= 0	Val= 0.2	14.8	0.0	0.0	0.0	95.8	0.2