Facial expression system on video using widrow hoff

M Jannah¹, M Zarlis² and H Mawengkang³

¹Department of Computer Science, Universitas Sumatera Utara, Medan, 20155.  
²Department of Computer Science, Universitas Sumatera Utara, Medan, 20155.  
³Department of Mathematics, Universitas Sumatera Utara, Medan, 20155.

Email: miftahuljannah0077@students.usu.ac.id, m.zarlis@usu.ac.id, hmawengkang@yahoo.com

Abstract. Facial expressions recognition is one of interesting research. This research contains human feeling to computer application. Such as the interaction between human and computer, data compression, facial animation and facial detection from the video. The purpose of this research is to create facial expression system that captures image from the video camera. The system in this research uses Widrow-Hoff learning method in training and testing image with Adaptive Linear Neuron (ADALINE) approach. The system performance is evaluated by two parameters, detection rate and false positive rate. The system accuracy depends on good technique and face position that trained and tested.

1. Introduction
Facial expression includes rich information regarding individual’s behaviour. In addition, face is an indicator of human reactions and feelings.[6]. The Automatic facial expression has been applied to many practical applications Fields such as human-computer interactions. In recent years, the study of facial expression recognition has become the more important topic to many researchers from different research fields [3]. In the development, image processing as a replacement of human eye while computer vision as a brain that used to process the information. In this case, the computer can detect an image object that moves anywhere (real time) and the computer can decide an expression from an image (happy and sad). In previous research, many researchers find facial expression without real-time image while in this research, the researcher focus on the video image.

2. Theoretical Basis
2.1. Face Detection
The first step of face detection is preprocessing. The step of feature extraction is detecting feature points then rotating to line up then locating and cropping the facial image using a rectangle form. In the process of face area detecting, a single image involves some methods knowledge based that is known by the facial invariant process, template matching and appearance based [2]. Many factors that must be considered in detecting of face such as hairstyles, glasses, this is some variables that vary significantly, so this is not a trivial task in getting face pattern [3]. This is a step of image process Firstly, convert into a binary scale and scan the image for forehead and then find out for a maximum width of continuous white pixels of image until reach eyebrows and last cut the face in such a way to perform face detection image [4].

2.2. Expression Detection
Facial expression recognition is related to the problem of classifying facial images into human expression classes. Expression recognition contains a kind of subjects such as perceptual recognition, affective computing and machine learning[7]. It is very difficult to define the meaning of emotions...
even that almost everyone knows intuitively about the intent of emotion. The emotion as a stimulus that evaluates experiences based on the potential for gain or pleasure[1].

2.3. Video Image
In the object movement detection, there is a matching process called by Frame Difference which reduces one frame to other frame and labeled the larger frame. This process will capture the border of the moving object. First of all the video image will display the image captured by the webcam then converted to 16 grayscale images with the aim to facilitate image processing. When a new image is captured by the webcam, the program will calculate the difference between two images by calculating the average value of all the gray value.

2.4. Adaptive Linear Neuron
Adaline (Adaptive Linear Neuron) is simple linear model two-layer neural network. One layer refers to input and another layer refers to output layer which has a single output neuron. The function of all linear model neurons transfer (y = kx+n) and network exercise the result with Least Mean Squares (LMS) algorithm for learning. The scopes of Adaline network are recognizing patterns, data filtering, or to approximate linear function. Calculate the net input to the output unit “y”

\[ y_{in} = b + \sum_{i=1}^{n} x_i w_i \]  

Where:
- \( y_{in} \): The number of input
- \( b \): bias
- \( w_i \): weight

Update the weights \( w_i \) and bias \( b \) for \( i = 1 \) to \( n \)

\[ w_i \text{ (new)} = w_i \text{ (old)} + \alpha (t - y_{in}) \]  
\[ b \text{ (new)} = b \text{ (old)} + \alpha (t - y_{in}) \] 

Here ‘t’ represents the target value. Apply the activation function over the net input calculate

\[ Y = \begin{cases} 
1 & \text{ if } y_{in} > 0 \\
-1 & \text{ if } y_{in} < 0 
\end{cases} \]  

In Adaline neural network, the learning process held by the adjustment of the weights of the neurons as per the weighted summation of the net inputs. Each neuron in the Adaline neural network accepts more than one input but generates one single output [5]. Adaline network architecture is shown in the following figure.

![Figure 1. The Adaline architecture](image)

2.5. Widrow Hoff
The least mean square error (LMS or Widrow-Hoff) algorithm is an example of supervised training, in which the learning rule is provided with a set of examples of desired network behavior:

\[ \{p_1, t_1\} , \{p_2, t_2\} , \ldots, \{p_Q, t_Q\} \]
As each input is applied to the network, the network output is compared to the target. The difference between the target output and the network output can be calculated to be an error in this method. It is used to double the average number of errors in the network, the target will be compared with the network output.

3. The Scheme Of System

The scheme of System is shown in Figure 2.

![Figure 2. The scheme of facial expression system in video](image)

The step is performed after the system receives video input is gray-scale, convolution, and face pattern recognition test through ADALINE network. In the pre-processing step, the video will be represented in the form of a single channel and ends with edge detection through the convolution process. In the main process, computing using the ADALINE network, the facial pattern vector will be trained to obtain a weight matrix, which then the weight matrix is used as the testing matrix to get the expression present in the face area by inputting the facial data in real-time.

3.1 Gray-Scale

The gray-scale Flowchart processes are developed based on Figure 3:

![Figure 3. Flowchart of the grey-scale process](image)

The grayscale step is required to align the intensity value of a 24-bit video. Consecutive values in channel R, G and B are summed and divided by the number of channels so that it can be calculated by the number of each channel divided by the number of channels. The process of the gray-scale video really helps the acceleration in the next concept.

3.2. Convolution

The convolution Flowchart processes are developed based on Figure 4:
The convolution process is involved in transforming the conditioned intensity values from the previous stage, into the intensity values that represent the edge of the object (face). Each intensity value will be partitioned into a 3x3 matrix and consonant with a Sobel kernel.

3.3. *Widrow Hoff*

Flowchart of Widrow Hoff proces in the system is shown in Figure 5:

In this stage, the network receives a pattern structure input with a sequence of the edges of objects that may contain the face or not the face, and then the network performs an error calculation. After the error calculation is performed, the network performs a correction matrix calculation and ends with a weighted matrix renewal. These steps continue to be repeated until the minimum error or have met the maximum iteration limit.
4. Result and Discussion

4.1 View of the training form
The training form in the system is look like in figure 6:

![Figure 6. View of the example the Training form](image)

In this form, the facial image was captured by webcam and face area located by red shape to lock the pixels of the face area and will save the matrix weigh of the pixel to storage and later use as a reference to the testing process. The training process is carried out in accordance with each expression. Firstly, We can click the auto button to get the facial matrix then click every expression button ((happy, sad, angry and normal) to get the facial expression matrix.

4.2 View of testing form
Testing or result form in the system is look like in figure 7:

![Figure 7. View of the result of facial expression form](image)
In the testing process, the system will read the value of the input video image that will be compared with the value that has been stored during the training process and where the matrix value is close.

5. Conclusions
From the results of research and discussion undertaken conclusions can be drawn as follows:
1. The light that even affects the detection, the intensity is not too bright and not too dark so that the test in the system can work well.
2. For accurate and precise results, the conditions or environments of the image capture must be observed. The conditions in question include the light, place, distance, and position of the object image.
3. Detection of facial expression is very dependent on the existing threshold value.
4. The accuracy of the system depends on the number of samples, the more accurate the sample of faces being trained and tested, the more accurate the expression facial recognition results and the impact on the high percentage.
5. The threshold matrix setting depends on the face when tested, if the face tested shows a higher matrix threshold then approaches the false positive rate, if not too low then approaching the detection rate, the threshold matrix is adjusted from the tested face.
6. Accuracy results in detecting video objects have a range of 85% in detecting happy expression, 80% in detecting sad expression, 75% in detecting normal expression and 85% in detecting angry expression.

References
[1] Singh D 2012 Human Emotion Recognition System Int. J. Image , Graphics and Signal Processing 8 pp 50-56.
[2] Hemalatha G and Sumathi C P 2014 A Study of Techniques for Facial Detection and Expression Classification Int. J. of Comp. Sc. & Engineering Survey (IJCSES) 5 pp 27-37
[3] Su MC, Yang CK, Lin SC, Huang DY, Hsieh YZ and Wang PC 2013 An SOM-based Automatic Facial Expression Recognition System Int. J. on Soft Comp., Artificial Intell. and App. (IJSCAI) 2 pp 45-57.
[4] Chavan P M, Jadhav M C, Mashruwala J B, Nehete A K and Panjari P A 2013 Real Time Emotion Recognition through Facial Expressions for Desktop Devices Int. J. of Emerging Science and Engineering (IJSE) 1 pp 104-108.
[5] Saranya R and Kuppusamy K 2016 Performance analysis of ADALINE and MADALINE Network in Classification of Thyroid Disease Int. J. of Advanced Research Trends in Engineering and Technology (IJARTET) 3 pp 117-119.
[6] Shbib R and Zhou S 2015 Facial Expression Analysis using Active Shape Model Int. J. of Signal Processing, Image Processing and Pattern Recognition 8 pp 9-22.
[7] Ghosh S and Bandyopadhy S K 2015 Method for Human Emotion Recognition System British J. of Math. & Comp. Sc. 11 pp 1-27.