Int. J. Advance Soft Compu. Appl, Vol. 7, No. 3, November 2015
ISSN 2074-8523

Neural Network Model for Prediction of Facial Caricature Landmark Configuration using Modified Procrustes Superimposition Method

Suriati Sadimon[1], **and Habibollah Haron**[1]

[1]Department of Computer Science, Faculty of Computing,
Universiti Teknologi Malaysia, 81300, Johor Bahru, Johor, Malaysia.
e-mail: suriati@utm.my, habib@utm.my

Abstract

Artificial Neural Network which possessing self-learning ability has shown great promise in addressing problem of learning an artist style in generating facial caricatures. This paper presents Artificial Neural Network model to imitate a particular artist style and to predict a facial caricature configuration for a given original face image (photo). This paper also describes the data preparing process that proposes a modified procrustes superimposition method in deriving the datasets for the neural network model. The experiment is carried out to compare the modified procrustes superimposition method with the original one and to find the appropriate neural network structure that would yield the most accurate prediction results. Different datasets (N1, N2, and N3) derived from the same raw data but using the different method in preparing the dataset and different numbers of hidden nodes (6, 12, 18 and 24) are tested. The experimental result and its detail analysis are given and discussed. It proves that neural network has an ability to predict how the artist exaggerates the original facial feature point. Dataset N2 which use Modified Procrustes Superimposition method and the simple structure of a single hidden layer neural network, in which 6 is the number of hidden node, give the best accuracy of the prediction.

Keywords: *Neural Network, Caricature, Face Image, Procrustes Superimposition, Artist Style, Exaggeration.*

1 Introduction

Facial caricature is a representation of the essence of a person face by exaggerating their distinctive facial features. It is commonly used in wide variety ways in our daily life such as in the mobile and internet application. The peoples can protect their identity and real image for security purposes from being manipulated by other peoples but at the same time the basic facial gestures still can be recognized when their caricature is used for social communication. In order to produce a stylistic caricature that is very similar to the real artist's work, the style of a particular artist and his works need to be learnt and considered in the process of generating facial caricature. However, the artist style, in how they exaggerate the facial features, is hard to be formulated and algorithmically coded. Learning based method and machine learning technique provide a promising way to address this problem. Some related works have been done by [1] [2-5]. Liang et al. [1] proposed example based caricature system which use partial least square to estimate the distinctive facial features and the rate of exaggeration. Junfa et al. [2] employed Principle Component Analysis (PCA) and Support Vector Regression (SVR) to predict the caricature for the input face image. Liu et al. [4] further came up with non-linear mapping model which employed semi-supervised manifold regularization learning. Yang [5] proposed a learning based system which use locality preserving hallucination algorithm based on radial basis function regression to learn the relationships between the shape of the original photo and the caricature. Lai et al. [3] adopted Neural Network to generate caricature but has no numerical analysis on the accuracy of the result. The problems in using learning based method are limited data collection of face image-caricature pairs and inconsistency of the artist styles. However, neural network is capable to predict on the small datasets but requires some precaution to avoid any potential problem such as overfitting and also the data preparation process must provide significant information and good quality data for the neural network input since the quality of the input data into neural network models can affect the obtain results [6]. The different ways in preparing the dataset which is used as input and target output of the neural network model for this domain problem also may yield the different results. Therefore, the strong interest to find a way to produce high accuracy of prediction result and also to complete inadequacy of the work done by Lai et al. [3] have motivated this research.

This paper purposes the best neural network structure in term of the hidden nodes number and the best way in preparing the dataset that will be used as input and target output of the neural network model in order to produce the best prediction accuracy result. The scope is only in front view facial contour.

This paper is organized as follows: Computer Generated Caricature and Artificial Neural Network are briefly explained in section 2 and section 3 respectively. Section 4 presents the methodology of the experiment. Data preparation is described in section 5 and the neural network model is explained in section 6.

Section 7 gives experimental result, detailed analysis and discussion. Lastly, section 8 concludes the result with suggestion for further research.

2 Computer Generated Caricature

Facial caricature emphasizes the prominent features of a person by exaggerating those features in an appropriate way and simplifying others features to make it easily recognized and memorized for variety of purposes such as humorous, insulting or offensive. The basic things must presences in a caricature are exaggeration and likeness. The artists need to observe the prominent features of a person, exaggerate it with their own style and at the same time, maintain the likeness of the person. Not all people have a talent of drawing caricature and the underlying process is hard to be explained accurately. Therefore, computer generated caricature has become a challenging research topic. Computer generated caricature is studied to produce facial caricature automatically or semi-automatically using computer graphics and image processing techniques in order to assist not only the skilled user like artist but also those who do not have any ability in drawing caricature. It attempts to imitate how the caricature is drawn by the artist by converting the process of drawing caricature into the formula and algorithm that will be executed by computer. The process of generating caricature by computer involves four steps. The first step is the facial feature points are extracted from the original face image (photo). Next, the prominent features are determined and then, exaggerate those features points to the new ones. Lastly, the original face image or sketch face is warped according to the new points or the new points are connected by line or curve to produce facial caricature [7]. In previous works, there are four approaches for generating facial caricature from the original face image (photo): interactive approach, regularity-based approach, learning-based approach and predefined database of caricature illustration approach [8].

In regularity based approach, the basic rule in drawing caricature is exaggerating the difference from the reference face. The reference face can be a standard face model [9,10] [11] or an average face. Average face is widely agreed among the caricaturists and the researchers to be used as a reference face. [12-16] employ an average face derived from a collection of face images in database as a reference to determine the distinctive facial features. It is based on the psychologies [17] that human beings have a “mean face” recorded in their brain, which is an average of faces they encounter in life. The artist determine the prominent features by comparing the facial features of a person to the mean face in their mind and found which feature is larger, rounder or smaller that make the person unique. The difference between original face image and the average face will be exaggerated to produce a facial caricature. It can be defined in Equation (1) below [18] [19].

$$Q = P + b * (P - S) \tag{1}$$

where Q is generated caricature, P is original face images, S is the average face, b is the rate of exaggeration and (P-S) is the difference between original face image

and the average face. The average of face from the database of face images can be calculated as shown in Equation (2) [19].

$$x_i^{(s)} = \frac{1}{M}\sum_{j=1}^{M} x_i^{(Pj)} \quad y_i^{(s)} = \frac{1}{M}\sum_{j=1}^{M} y_i^{(Pj)}$$

$$i = 1,2,...,N \tag{2}$$

where $x_i^{(Pj)}$ and $y_i^{(Pj)}$ are the *x* and *y* coordinates for the i-th feature point of the j-th normalized face data respectively. *M* is the number of face images in the database. *N* is the number of feature points in each face.

3 Artificial Neural Network for Face Caricature

Artificial Neural Network has been applied successfully in many problems related to the face such as in face ageing [20,21] , face recognition [22-24], face detection [25,26] and face reconstruction [27]. It is a robust approach and among the most effective learning methods to approximate extremely complex functions. The key features of Neural Networks are fault-tolerance, self learning ability, and can cope with complex non linear problem. It is applicable to any problem in which a relationship between input variables (predictor) and output variables (predicted) exist, even the exact nature of the relationship could not be known, very complex, not easy to interpret, and the data corresponds to noisy. These features are in line with the domain problem which attempt to learn the artist style in producing caricature. It is because the relationship between original face image and its correspondence caricature is not exactly known and how a particular artist exaggerates the distinctive features is hard to be quantified. Some of the previous works assume the relationship is linear [1,2], whereas others [3,4] believe the relationship must be non-linear. Very few works [3][28] use artificial neural network for generating facial caricature.

Artificial Neural Network or Neural Network has been inspired by the biological learning system and it is a very much simplified version of biological neural systems [29]. It is built out of a densely interconnected set of nodes where each node has inputs, outputs and performs a simple computation by its node function. Each connection between nodes has its weight value. The simple and most popular network architecture is multilayer perceptrons [30], which nodes are arranged in a feedforward topology. There are three layers: input layer, hidden layer and output layer with n, k and m nodes respectively as shown in Fig. 1. *Xi* represents the input variable, *w* and *l* stand for the weights, *f* and *p* represent the activation value and *Z* and *Y* stand for the output of the nodes. The input layer nodes contain the values of the input variables. Each node in hidden layer and output layer is connected to all of the nodes in the preceding layer. This node performs computation to get activation value, f_k by summing all (*n*) outputs of the nodes in the preceding layer, X_i which has been multiplied by its weight *w*, and adding the threshold, *bias* as shown in Equation (3).

$$f_k = w_{1k}X_1 + w_{2k}X_2 + ... + w_{nk}X_n + bias$$
$$= \sum_{i=1}^{n} w_{ik}X_i + bias \quad (3)$$

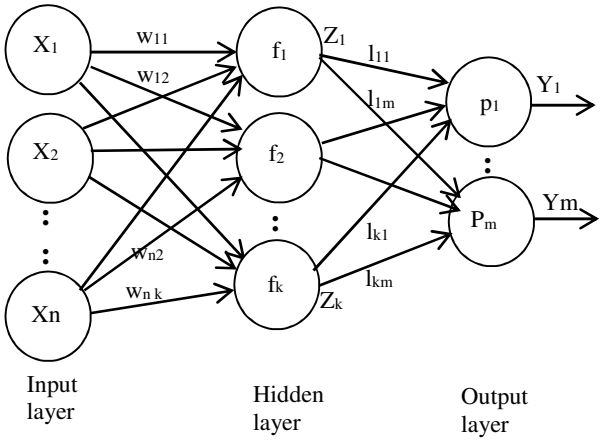


Fig 1: Multiple Layer Perceptrons

All inputs to one node come in at the same time and remain activated until the output is produced. To produce an output of this node, the activation value needs to pass through the activation function or transfer function. The most common activation function is logistic function or sigmoid function as shown in Equation (4).

$$Z_k = sigmoid(f_k) = \frac{1}{1+\exp(-f_k)} \quad (4)$$

where Z_k is the output of node k and f_k is the activation value of node k.

The objective of learning algorithm is to determine the weight values of the network that best fit to the data set and to minimize the prediction error made by the network so that the prediction result will be close to the target output. The prediction error can be known by comparing the output of the training result, Y_k with the target output, T_k from the observation. The prediction errors of all the data training are combined together by an error function to produce the network error. The typical error function, E used for regression problem is mean squared error as shown in Equation (5).

$$E = \frac{1}{m}\sum_{k=1}^{m}(T_k - Y_k)^2 \quad (5)$$

where m is the number of output layer nodes, T_k is the target output, Y_k is the output of the training result. Back propagation is the best known example and the easiest algorithm to understand. In backpropagation algorithm, the global minimum error

is sought iteratively through a number of epochs. In each epoch, the weights of the network are adjusted based on the network error and then the process of producing output repeats. The iteration is stopped when a given number of epochs have passed or when the acceptable level of error has reached or when no improvement in the error. There are many training algorithm used to find the minimum error such as gradient descent, conjugate gradient descent, quasi-Newton and Levenberg Marquardt. The Levenberg Marquardt is the fastest convergence for networks that contain up to a few hundred weights especially if very accurate training is required and is able to obtain lower mean square error than other algorithm. In addition, the Levenberg Marquardt is the most efficient algorithm for small- and medium-sized networks [31]. However, it requires a larger storage space. In this research, feed forward back propagation neural network with one hidden layer is employed to learn the relationship between the original face image and its corresponding caricature. The Levenberg Marquardt is chosen as the training algorithm due to its described advantages.

4 Methodology

The methodology of the experiment can be summarized as shown in Fig. 2. Some pairs of original face image (photo) and its corresponding caricature are collected and extracted into facial features points based on the landmarks that have been predetermined and then, all the landmark points are normalized using original or modified procrustes superimposition method. Datasets N1, N2, and N3 are generated from the different methods in data normalization. Dataset N1 is generated by Procrustes superimposition method, whereas dataset N2 and N3 are generated by the Modified Procrustes superimposition method. Dataset N2 differs to N3 only in the first step of the method. For each datasets, calculate the average face by averaging all the original face feature points. Then, calculate the difference between the original face to the average face and also the original face to its caricature that will be used as input and target output, respectively to the neural network model. The training process of the neural network with specific architecture and parameters is performed to get the value of weights and biases of the network. The neural network output is compared to the target output to find the accuracy of the output result. The training and testing process are done repeatedly for the different datasets and different number of hidden nodes to find the neural network structure and the dataset that would produce the most accurate prediction results. Lastly, the predicted facial caricature for the new image is acquired by adding the predicted output from the neural network to the original face feature points.

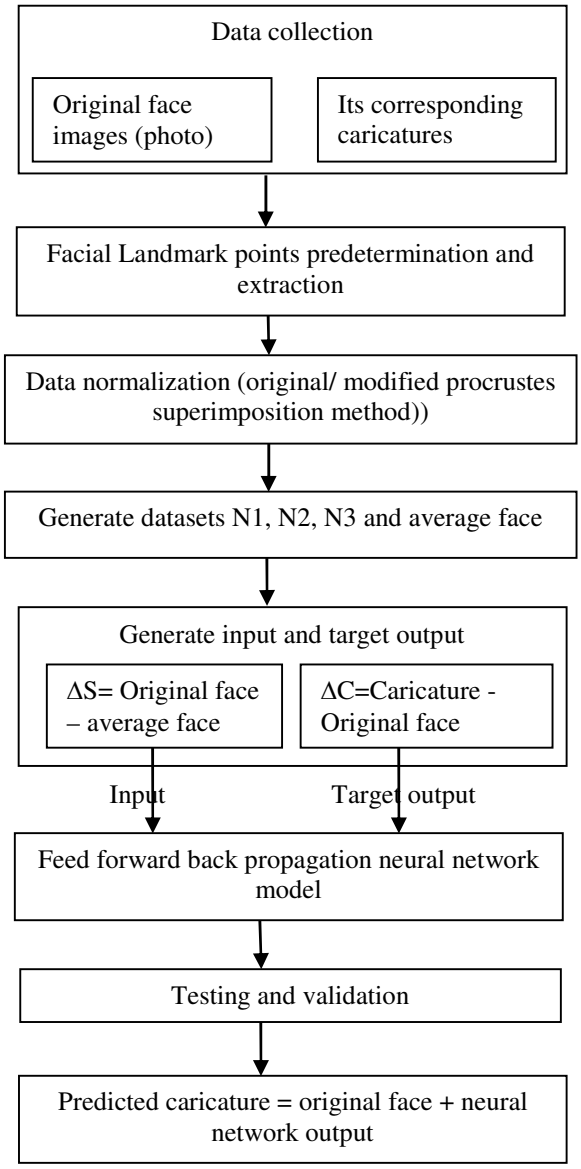


Fig. 2: Methodology of the Experiment

5 Data Preparation

5.1 Facial Landmarks Points Predetermination and Extraction

There are thirty two pairs of original face images (photo) and its correspondence hand-drawn caricatures have been collected from internet. The caricatures are drawn by artist Pritchett [32]. The images need to be preprocessed to make it clear and sharp. Only frontal view face images and face contour or shape are considered in this research. Facial feature points are extracted from the images basically based

on landmark-based geometric morphometric. Geometric morphometric is a quantitative approach to analyze shape and to study the variations in the shape geometry among various samples based on the Cartesian landmark coordinates [33]. Landmarks are a finite set of homologous points on the surface of an object that accurately describe the shape and correspond to each object that matches within all measured objects[34]. Landmark can be categorized as anatomical, mathematical and pseudo landmark[35].

The number of anatomical landmarks for face contour or face shape is about 8 points [36]. It is not enough to represent the face shape. Thus, Added landmark should be used to accurately describe the face shape. In this work, 24 landmark points are proposed to be used to represent the face contour as shown in Fig 3. The number of points used in the previous work is varies according to the purpose of their use [8] and usually ranges from 9 points [37] to 39 points [38]. In fact, there is no fixed rule or specific calculation to determine the exact number of the points that must be used in representing the face contour for all purposes.

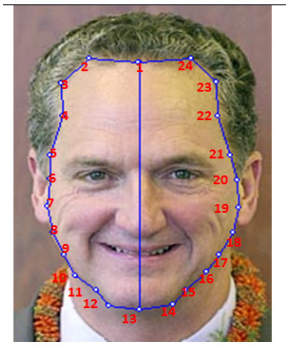


Fig. 3: Landmark Points on the Original Face Image

High number of the points can represent the face features in more detail and can generate variety form of exaggeration but need high computation and storage complexity [8] and vice versa for the lower number of points. The number of points used in this research is moderate and within the range of the number of facial feature points used by the pervious works. The same number of points as in [39] used in this research. However, the points used in [39] are just located equidistant along the face contour. On the other hand, the facial feature points in this research based on the region on the face and has significant biological meaning which each three consecutive points represent the boundary of part of the face for example point 6, 7 and 8 represent boundary of the left cheek and point 9, 10 and 11 represent boundary of the left jaw as shown in Table 1.

In addition, some of the points are referred to the inner part of the face for example point 6 and 20 refer to the corner of the eyes, point 8 and 17 refer to the mouth. The points on the face caricature that correspond to the points on the original face image are selected by observing the similarity of the gradient or the curvature of the face contour in both images (Fig. 4).

Table 1: Face landmark Points Description

Points	Description
P24, P1, P2	Top forehead
P3, P4, P5	Left forehead
P6, P7, P8	Left cheek
P9, P10, P11	Left jaw
P12, P13, P14	chin
P15, P16, P17	Right jaw
P18, P19, P20	Right cheek
P21, P22, P23	Right forehead

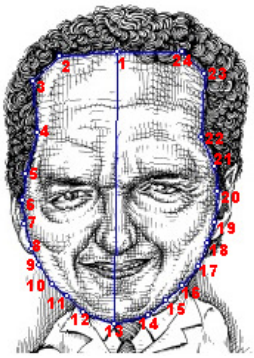


Fig. 4: Correspondence Points on the Face Caricature

If there are similar corners, similar curvatures, or inflection points in both images, choose those locations as the correspondence points. Other parts of the face caricature are also used to ensure the selected point is on the right place. There are 24 points marked manually on each image using Vextractor 9 software in order to ensure the accuracy of the data obtained. These points are numbered in anticlockwise direction starting with the point on the top of the forehead. The origin point (0, 0) is located at the bottom-left. Fig. 5 shows landmark points for several samples of the original face image.

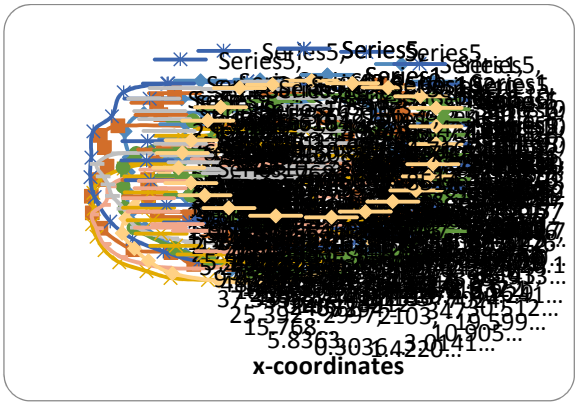


Fig. 5: Landmarks points of several samples of the original face image (photo)

Furthermore, the inner and outer corners of both eyes and the midpoint between eyes are marked as well. Those points will be used in normalization process and in generating datasets. Each features point has two coordinate values, x and y.

5.2 Data Normalization

Since the images of original face and facial caricature are in a variety of scales, or sizes, normalization is required so that all the landmark points are in the same scale. In this work, the landmark points are normalized using Procrustes Superimposition method to generate dataset N1, and using Modified Procrustes Superimposition Method to generate dataset N2 and N3. These different methods will be experimented to get the best dataset of the landmark points that lead to a better prediction accuracy.

5.2.1 Procrustes Superimposition Method

Procrustes Superimposition is a method to standardize the position, orientation and size of all the landmark configurations by translating them to the same centroid, scaling them to the same centroid size, and iteratively rotating until it shows a minimum sum of squared distances between the landmarks and their corresponding reference landmarks [33]. This method involves three steps: translation, isomorphic scaling and rotation.

5.2.1.1 Translation

All the facial landmark configurations are translated to the common center position. The origin (0, 0) is the most likely to be chosen as the common center position. The centroid of each landmark configuration is a mean value of x and y coordinates for all k landmarks in each configuration. It is calculated from the sums of the x and y coordinates divided by the number of landmark, k as shown in Equation (6).

$$X_c = (\bar{x}, \bar{y}) = \left(\frac{1}{k} \sum_{i=1}^{k} x_i \ , \ \frac{1}{k} \sum_{i=1}^{k} y_i \right) \quad (6)$$

This centroid is subtracted from each facial landmark coordinates in the configuration to center it at the origin as shown below.

$$X_{new} = X - X_c \tag{7}$$

Where, X is the matrix k x m of the coordinates of the k facial landmarks in m dimension. X_{new} is the new coordinates of X centered at the origin. X_c is the centroid coordinates. Fig. 6 shows the facial landmark configuration after translation.

5.2.1.2 Isomorphic scaling

All the facial landmark configurations are scaled to the common centroid size of 1 (|x|=1). Isomorphic scaling is a linear transformation that changes the size of an object smaller or larger by a scale factor that is the same in all direction. The ratio of the object proportions is maintained.

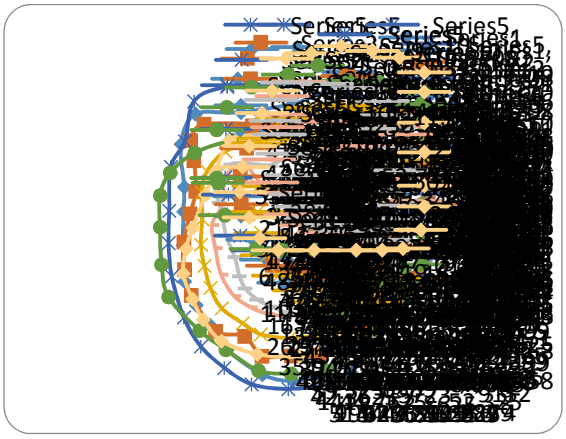


Fig. 6: Facial landmarks configurations after translation

Centroid size is the square root of the sum of the squared distances between all landmarks and their centroid as shown in Equation (8).

$$S(x) = \sqrt{\sum_{i=1}^{k}[(x_i - \bar{x})^2 + (y_i - \bar{y})^2]} \tag{8}$$

Where, x_i, y_i is the coordinates of facial landmarks. $(\bar{x}, \bar{y})$ is the centroid coordinates. The centroid size of each facial landmarks configuration that will be used as a scale factor is calculated using the equation above. In order to get the centroid size of 1 for all the facial landmark configurations, the facial landmarks coordinates are divided by the scale factor, *S(x)* as shown in equation below

$$X_n = X\left(\frac{1}{S(x)}\right) \tag{9}$$

Where, X is the matrix of the coordinates of the facial landmarks centered at the origin, X_n is the matrix of new coordinates of X and $S(x)$ is a scale factor. Fig. 7 shows the facial landmarks configurations after translation and scaling.

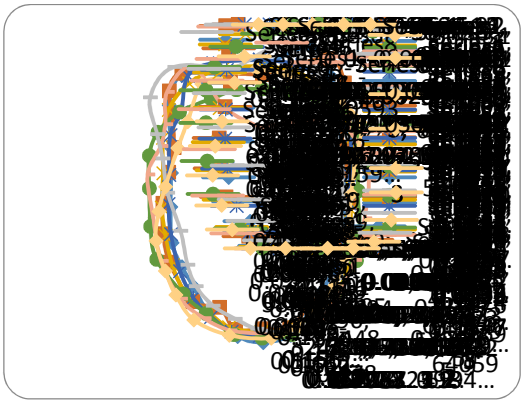


Fig. 7: Facial landmarks configurations after translation and scaling

5.2.1.3 Rotation

All the landmarks configurations are iteratively rotated to eliminate the variation due to the difference of orientation. Firstly, the mean of all the landmarks configurations is calculated using the equation below.

$$\bar{X} = \frac{1}{N}\sum_{j=1}^{N} X_j \tag{10}$$

Where, $\bar{X}$ is the mean of the landmarks configurations, N is the number of configurations, X_j is the coordinates of the landmarks of configuration j. Then, each facial landmarks configuration is aligned to the mean configuration of landmarks that is used as the reference configuration. The sum of square distances between all the landmarks, k and its corresponding reference landmarks is calculated by the equation below

$$D(x) = \sum_{i=1}^{k}[(x_i - \bar{x}_i)^2 + (y_i - \bar{y}_i)^2] \tag{1}$$

Where, (x_i, y_i) is the coordinates of the landmark i and $(\bar{x}_i, \bar{y}_i)$ is the coordinates of the corresponding landmark i in the reference configurations.

The coordinates of the facial landmarks after rotation by angle θ around the origin, X_n are calculated as in the equation below.

$$X_n = X\begin{bmatrix} cos\ \theta & sin\ \theta \\ -sin\ \theta & cos\ \theta \end{bmatrix} \tag{22}$$

Where, X is the matrix k x 2 of the coordinates of the facial landmarks centered at the origin and has been scaled. The facial landmarks configuration is iteratively rotated by 0.2 degree in anti clockwise direction. If the value of D(x) is increased in the first cycle, the rotational direction is changed to clockwise. The rotation by 0.2 degree is iterated for several cycles until the value of D(x) is a minimum that is less than 1x10e-8 or there is no downward change in the value of D(x) or the change is less than 1x10e-8. Fig. 8 shows the facial landmarks configurations after translation, scaling and rotation.

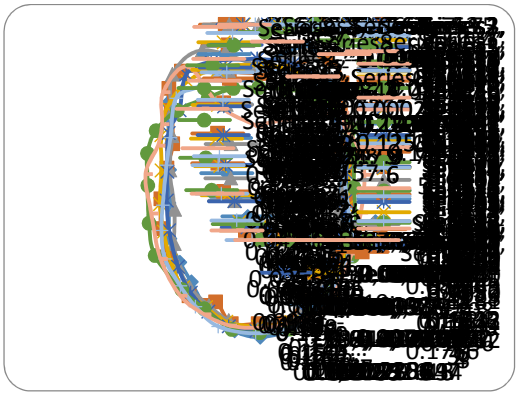


Fig. 8: Facial landmarks configurations after translation, scaling and rotation.

5.2.2 Modified Procrustes Superimposition Method

In this work, modified procrustes superimpose method is proposed to get better facial landmarks coordinates data that leads to a better prediction accuracy. This modified method differs in how the differences in size and orientation are removed. This method also consists of three steps: translation, isomorphic scaling and rotation. The modification is done in scaling and rotation step. The translation step is the same as described in the earlier method (section 5.2.1.1) which all the landmarks configurations are centered at the origin by subtracting the landmarks coordinates with the centroid coordinates that is obtained from the average of all k-landmarks in the configuration.

5.2.2.1 Isomorphic scaling

In this work, the distance between the two eyes center, instead of the centroid size, is used as a scale factor. The distance between eyes also has been used for normalization in generating facial caricature [3][12][40][41]. However, [41]) applied the distance between eyes center as a scale factor only to x direction, [12][41] use point between eyes as an origin and others did not mention about translation [3][41] and rotation [3]. The center point of an eye is determined from the two eye corner points that have been accurately extracted from the face image. The center of an eye, C^i is calculated by subtracting the right corner point, rc^i with the left corner point, lc^i , divided by 2 and then add it to the left corner point, lc^i as shown below

$$C^i = \frac{(rc^i - lc^i)}{2} + lc^i \tag{3}$$

Where i is left eye, l or right eye, r. The distance between the two eyes center is determined as below

$$S(x) = \sqrt{\left[\left(C_x^r - C_x^l\right)^2 + \left(C_y^r - C_y^l\right)^2\right]} \tag{4}$$

where, (C_x^r, C_y^r) is the coordinates of the center point of the right eye and (C_x^l, C_y^l) is the coordinates of the center point of the left eye. The new coordinates of the landmarks configuration are obtained by dividing the landmarks coordinates by the distance between the two eyes center, $S(x)$ as shown in Equation (9) in section 5.2.1.2.

5.2.2.2 Rotation

In the earlier method, rotation is based on all landmarks in a configuration and its corresponding reference landmarks but in this modified method, the rotation is only based on two landmarks, P1 which is located at the top of the forehead and P13 which is located at the bottom of the chin. The facial landmarks are iteratively rotated until the distance in x direction between the landmark P1 and landmark P13 is a minimum. The rotation process is done in such a way because the nature of the human face is vertical which the line connecting the top of the forehead and the bottom of the chin is in a vertical position. The horizontal distance between landmark P1 and P13, $D(x)$ is calculated as shown below

$$D(x) = |x_{13} - x_1| \tag{5}$$

Where, x_1 is coordinate x of the landmark, P1 and x_{13} is coordinate x of the landmark, P13. The facial landmark coordinates is rotated iteratively by the angle of 0.2 degree as describe in section 4.2.1.3. The new coordinates of the facial landmarks configuration are determined as in the Equation (12).

5.3 Generating Different Datasets, Average Face, Input and Target Output

In this research, there are three different datasets generated by different method in normalization process. Dataset N1 is generated by using procrustes superimposition method as describe in section 5.2.1. Dataset N2 is generated by using Modified superimposition method in normalization process as describe in section 5.2.2. Lastly, dataset N3 is generated also using Modified superimposition method but in the translation step, the midpoint between eyes is used as an origin and all the landmark configurations are aligned at this point instead of the centroid point. These three different datasets are being tested in order to find the best dataset for prediction. For each dataset, average face feature points are calculated from the database of the original face images using the Equation (1). Then, this average face will be compared with the original face image to find the differences between them. The differences between original face and average face, ΔS is obtained by subtracting the average face feature points from the original face feature points and can be defined by Equation (16).

$$\Delta S = original\ face - average\ face \tag{16}$$

It also represents how distinct the original face is compared to the average face.

The difference between original face and its caricature can be obtained by subtracting the original face feature points from its corresponding caricature feature points as shown in Equation (17).

$$\Delta C = caricature\ face - original\ face \quad (17)$$

This difference shows how the original face is changed or exaggerated by the artist. The difference between original face and average face (ΔS) and the difference between original face and its corresponding caricature (ΔC) are used as an input and a target output, respectively to the neural network model. ΔS and ΔC are calculated for each datasets.

6 Neural Network Model

6.1 Parameter setting

The number of input variables is equal to the number of facial feature points which is twenty four. Since the number of sample data is not too far from the number of input variables, early stopping and k-fold cross validation are employed [42] to prevent overfitting during the training. These 32 data samples are divided into 8 folds. Each fold contains four samples. Six folds are selected for training, one fold for validation and another one fold for testing. Different number of hidden node (6, 12, 18 and 24) is experimented to find the best neural networks structure and different datasets (N1, N2, and N3) is used to find which of the dataset provide the best result. The experiment is carried out by using MATLAB 7. Feed forward back propagation network with only one hidden layer is used as the architecture. The neural network is trained using Levenberg-Marquardt. A summary of the neural network architecture and parameters is shown in Table 2.

Table 2: The Neural network architecture

Neural network type	**Feed forward back propagation**
Number of nodes in input layer	24
Number of nodes in output layer	24
Number of nodes in hidden layer	24, 18,12, 6
Training function	Levenberg Marquardt
Performance function	Mean squared error
Training performance goal	0
Number of hidden layer	1
Hidden layer transfer function	Tan sigmoid
Output layer transfer function	Pure linear

6.2 Execution and Creating Predicted Facial Caricature

X and Y coordinates of landmark points in each datasets are trained separately. Early stopping is employed to avoid overfitting by terminating the training iteration when the validation error increase for more than 5 times since the last decrease. Twenty training runs are made per dataset to allow a wide range of initial weights and biases to be explored. In each run, different folds are chosen so that all the sample data can be used for training, validation and testing. It can avoid poor division of the data set. Mean Square Error (MSE) is used as a performance measurement. It is a way to quantify the difference between the predicted output from neural network and the target output from observation. It can be calculated as shown in Equation (18).

$$MSE = \frac{\sum_{k=1}^{N} (Po_k - To_k)^2}{N} \tag{18}$$

where *Po* is the predicted output, *To* is the target output and *N* is the total number of facial feature points in training or testing data. Mean Square Errors (MSE) of the twenty training runs are then averaged. Mean squared error (MSE) on the training data is the minimum error in training set which is used to find the network weights and biases, whereas MSE on the testing data is used as performance measurement to compare the different model of neural network and to verify the network design. If the mean square error (MSE) on the testing data is relatively low, the accuracy of the neural network model is high. Root mean square error (RMSE) and mean absolute error (MAE) are also used in this work as a performance measurement as shown in equation (19) and equation (20).

$$RMSE = \sqrt{MSE} \tag{19}$$

$$MAE = \frac{\sum_{k=1}^{n} |Po_k - To_k|}{n} \tag{20}$$

where *Po* is the predicted output, *To* is the target output and *N* is the total number of facial feature points in training or testing data. To accurately compare the result of different datasets, the normalized mean square error (NMSE) is used [43][44] as shown in Equation (21), since the distribution of the target output of different datasets are different.

$$NMSE = \frac{MSE}{\frac{1}{N}\sum_{k=1}^{N} (To_k - MTo)^2} \tag{21}$$

Where *To* is the target output, *MTo* is mean of the target output, *N* is the number of feature points on the testing data.

The predicted facial caricature features points, *PC* are derived by adding the predicted output, *Po* which obtained from the neural network to the original face features points as follow.

$$PC = original\ \ face + Po \tag{22}$$

All the predicted facial caricature features points are connected by line to create predicted facial caricature shape.

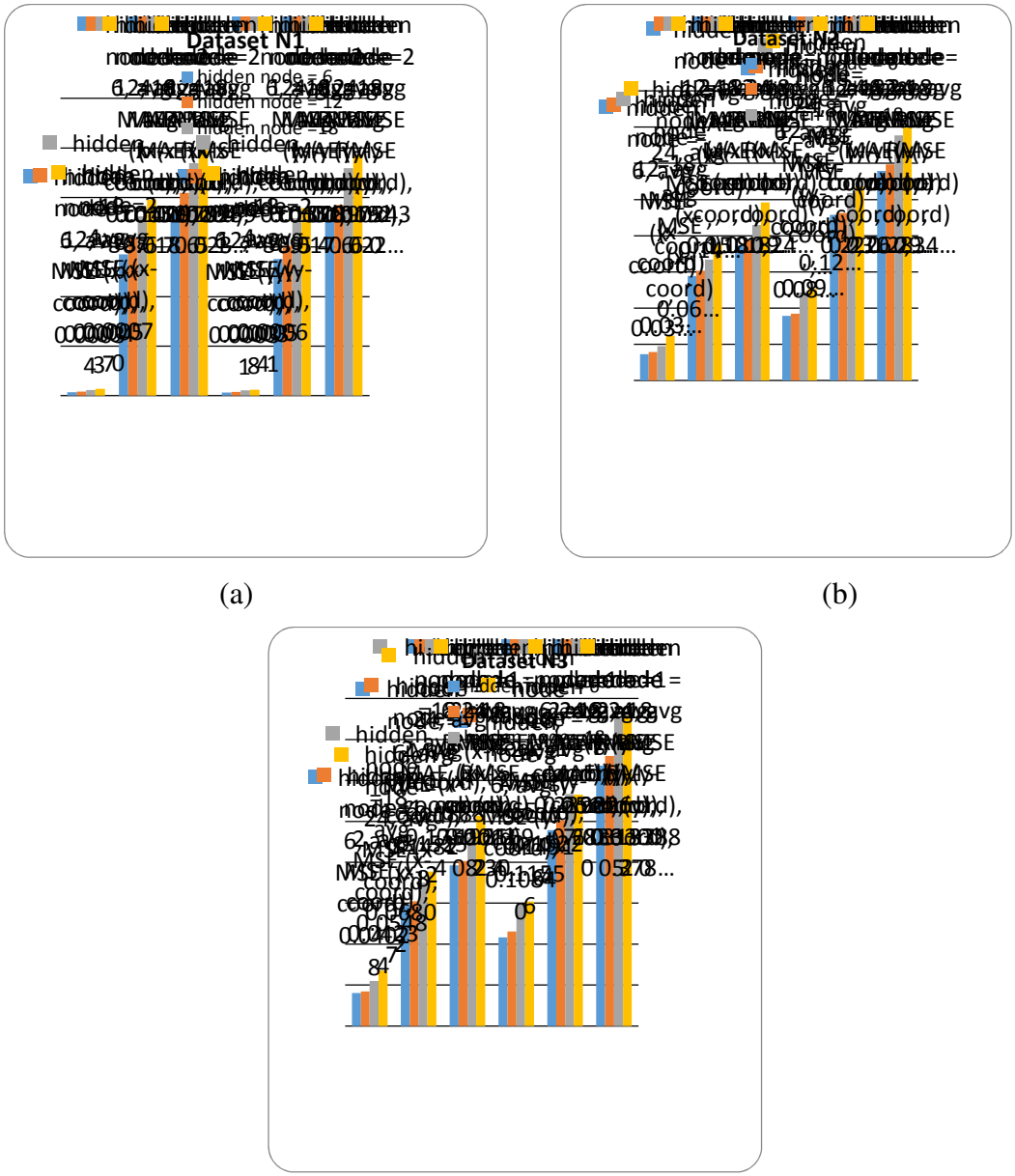
7 Result and Discussion

The primary concern in this experiment is the ability of generalization which could accurately predict the facial caricature points when presented with the new data. The analysis of the result is conducted to find the best structure of neural network and type of dataset that provide the most accurate results. Table 3 presents the overall result containing average MSE, average RMSE, average MAE and the average NMSE on the testing data for variety of hidden node numbers (6, 12, 18 and 24) and variety of datasets (N1, N2, and N3) for coordinate x and y. Error of the testing data is decreased when the number of hidden nodes was reduced from 24 nodes to 6 nodes.

Table 3: Experimental Results

	x-coordinate				y-coordinate			
	hidden nodes = 6	hidden nodes = 12	hidden nodes = 18	hidden nodes = 24	hidden nodes = 6	hidden nodes = 12	hidden nodes = 18	hidden nodes = 24
Dataset N1								
Avg MSE	0.00034	0.00043	0.00057	0.00070	0.00031	0.00038	0.00054	0.00061
Avg NMSE	1.57958	1.98009	2.71367	3.36552	1.59721	1.94662	2.77758	3.09784
Avg MAE	0.01427	0.01596	0.01808	0.01977	0.01379	0.01515	0.01791	0.01894
Avg RMSE	0.01836	0.02045	0.02346	0.02591	0.01756	0.01926	0.02294	0.02430
Dataset N2								
Avg MSE	0.03634	0.03939	0.04765	0.06211	0.08935	0.09224	0.12185	0.12659
Avg NMSE	1.14401	1.24619	1.47967	2.08989	0.97969	1.46797	1.73885	2.24345
Avg MAE	0.14448	0.15150	0.16706	0.18801	0.22902	0.23126	0.25888	0.26779
Avg RMSE	0.18653	0.19443	0.21487	0.24604	0.28956	0.29840	0.33865	0.34933
Dataset N3								
Avg MSE	0.04028	0.04234	0.05487	0.06802	0.10840	0.11556	0.14916	0.15224
Avg NMSE	1.19760	1.27107	1.67104	2.16543	1.02447	1.41520	1.70378	2.17683
Avg MAE	0.14828	0.15252	0.17504	0.18881	0.23927	0.25146	0.28330	0.28260
Avg RMSE	0.19618	0.20132	0.23079	0.25694	0.31835	0.33002	0.37836	0.37880

For all the dataset N1, N2, and N3 (x coordinate value and y coordinate value), a neural network structure with 6 nodes in hidden layer reveals the most accurate result, followed by hidden nodes equal to 12, then 18 and lastly, hidden node equal to 24 as shown in Fig. 9.



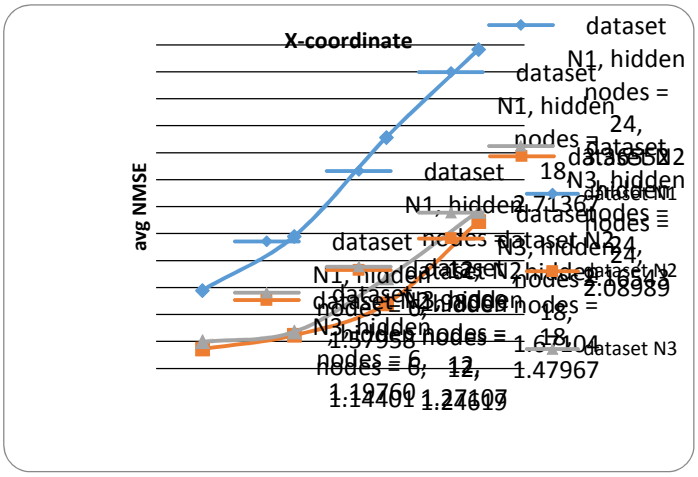
(a) (b)

(c)

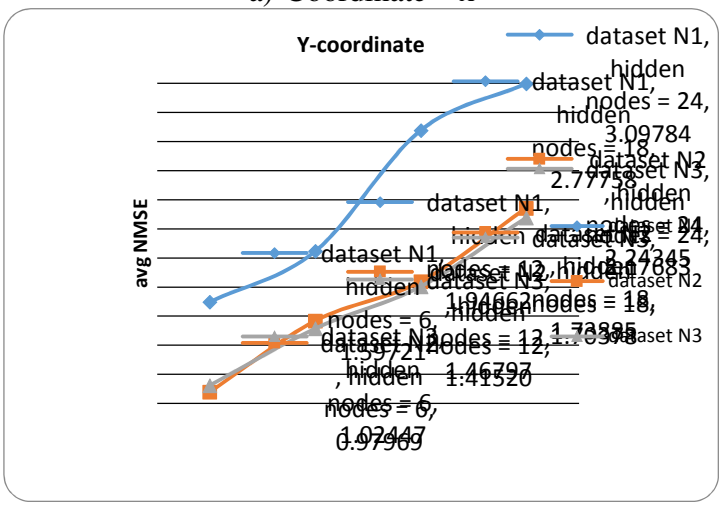
Fig. 9: Experimental result (a) dataset N1 (b) dataset N2 (c) dataset N3

Generally, dataset N2 and N3 give a better result than dataset N1 for coordinate x and y (Fig. 10). For coordinate-x, the most accurate result is given by dataset N2 for all number of hidden nodes, followed by dataset N3 and lastly dataset N1 (Figure 10a). Among all the different type of datasets and different number of hidden nodes, the lowest average NMSE is shown by dataset N2 that is 1.14401 for coordinate x and 0.97969 for coordinate y with the number of hidden nodes equal to 6 (Table 3). In addition, coordinate x and y need to be combined to generate facial caricature contour. Thus, the average NMSE of both coordinates are

combined and averaged as shown in Table 4. It is also found that dataset N2 and the number of hidden nodes equal to 6 provides the most accurate prediction results which its average NMSE is 1.06185.



a) Coordinate – x



b) Coordinate-y

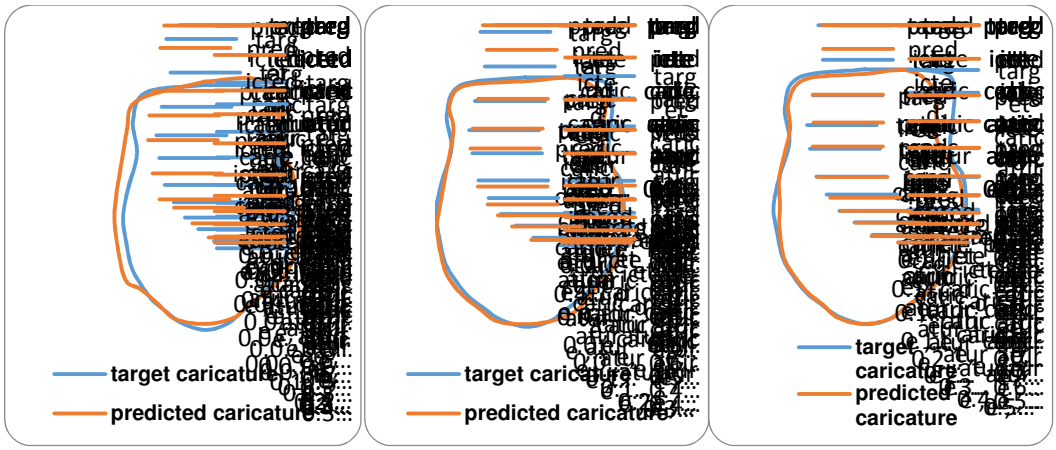
Fig. 10: Graph of Number of hidden nodes versus average NMSE

Table 4: Average NMSE for the combination of coordinate x and y

hidden nodes	Average NMSE		
	Dataset N1	Dataset N2	Dataset N3
6	1.58839	1.06185	1.11103
12	1.96336	1.35708	1.34314
18	2.74563	1.60926	1.68741
24	3.23168	2.16667	2.17113

If compared to Lai et al.[3], the dataset of face image-caricature pairs used in this work are different from that are used by them. They employ twelve pairs that consist of male face only and the caricatures are drawn by different artists, whereas thirty two pairs including male and female face are used in this work and the caricatures are drawn by one artist. The representation of the face contour and the data preparation process in this work are also not the same as them. The neural network model in Lai et al [3] is not tested with variety of parameter values and there is no numerical or statistical analysis on the accuracy of the result. However, both of us use backpropagation neural network and Levenberg Marquardt. The neural network structure with 24 hidden nodes in this experiment is similar to the neural network structure used by Lai et al. [3], which the number of hidden nodes is equal to the number of input variable. The obtained result shows that this number of hidden node could not provide a better prediction result compared to the network with the smaller number of hidden nodes. This experiment offers an increasing of the accuracy of the prediction by reducing the number of the hidden nodes from 24 to 6 because the neural network structure is smaller and simpler. The small neural network is better than a bigger one for the small amount of dataset as stated in [45]. The experimental result also shows that the different datasets provided by data preparation process can affect the obtained result. It is because the quality of the input data strongly influence the results and if important input data are missing or distorted, the neural network's performance can be affected [6]. Dataset N2 and N3 give better result compared to dataset N1. It reflects that modified Procrustes Superimposition method is better than the original one. It might because rotation in original procrustes method not respects to the orientation of biologically relevant axes. The variance is assumed to be equal and circularly distributed for each landmark points [46]. This assumption is not true for a human face. Landmark point at the top, P1 and at the bottom, P13 of the face has smaller variance than other landmark points due to the nature of the human face is in vertical position. In addition, the original procrustes superimposition method also scales the face shape into the common centroid size of 1. So, the value of the differences between the original face and its corresponding caricature and also between the original face and the average face which is used as input and target output are very small that is in between -0.05 and 0.05. Thus, the important and significant information might be missing and affect to the result. On the other hand, use of the distance between two eyes center as a scaling factor gives a better result. It is because it is compatible with the caricature drawing principle that the inter-iris distance needs to be maintained while drawing the facial caricature since it plays an important role in identifying person [47]. Moreover, in this work, the facial caricature is compared to the original face image which both refers to the same person and most certainly has the same distance between eyes center. Dataset N2 shows a better prediction result compare to N3. It reflect that aligning the facial landmark configuration at the center point is a better way because it could distribute the difference values caused by the different location well to all the feature points and it also can avoid the error that may happens during extracting the midpoint between eyes.

The neural network output is the prediction of the difference between original face feature points and its caricature. When this output is added to the original face feature points using Equation (22), a predicted facial caricature points can be obtained. Figure 11 shows target facial caricature contour and predicted facial caricature contour for the minimum MSE for the different datasets produced by the neural network structure with six hidden nodes. Generally, the predicted caricature generated by dataset N2 (Figure 11b) shows the most similar shape to the target caricature. The difference between the predicted and the target caricature is not too far for all part of the face contour. For the dataset N3 (Figure 11c), the upper part (forehead) of the predicted caricature contour is narrower than the target caricature. For the dataset N1 (Figure 11a), the predicted caricature contour is wider than the target caricature.



(a) Dataset N1 (b) Dataset N2 (c). Dataset N3

Fig. 11: Target caricature and predicted caricature

8 Conclusion

This paper presents the neural network model for predicting a facial caricature configuration together with the analysis of the accuracy result. From the experimental result, the best performance is given by the dataset N2 and the neural network structure with the number of hidden nodes equal to 6. It shows that the modified procrustes superimposition method is better than the original one. It reflects that the best way for preparing the dataset for the neural network is first, all the facial landmarks configurations are aligned at the centroid point, scale it by using the distance between two eyes center as a scale factor and lastly, rotate iteratively until the distance in x direction between the landmark point, P1 where located at the top of the face contour and the landmark point, P13 where is located at the bottom of the face contour is a minimum. Reduction of the number of hidden nodes from 24 to 6 also can increase the prediction accuracy for this neural network

model which has the same number of input and output nodes. It means that less complex neural network model can provide faster training and higher accuracy of prediction for this problem. The result also shows obviously that the neural network has the ability to predict how the original face image would be exaggerated. However, the accuracy of the prediction still can be improved for further research by using different input variables or different type of neural network such as generalized regression or by hybridizing the neural network model with other artificial intelligent techniques.

ACKNOWLEDGEMENTS
This work is partially supported by the Fundamental Research Grant Scheme (R.J130000.7828.4F497)

References

[1] Liang L, Chen H, Xu Y-Q, Shum HY. 2002. Example-based Caricature Generation with Exaggeration. In: *Proceeding of 10th pacific conference on computer graphics and applications.* pp 386-393

[2] Junfa L, Chen Y, Gao W. 2006. Mapping Learning in Eigenspace for Harmonious Caricature Generation. In: *14th ACM International Conference on Multimedia*, Santa Barbara, USA, October 22-27 pp 683–686

[3] Lai KH, Chung PWH, Edirisinghe EA. 2006. Novel approach to neural network based caricature generation. In: *IET International Conference on Visual Information Engineering (VIE 2006)* Bangalore, India, 26-28 Sept. pp 88-93

[4] Liu J, Chen Y, Xie J, Gao X, Gao W. 2009. Semi-supervised Learning of Caricature Pattern from Manifold Regularization. In: *Proceedings of the 15th International Multimedia Modeling.* pp 413–424

[5] Yang T-T, Lai S-H. 2010. A Learning-based System For Generating Exaggerative Caricature From Face Images With Expression. In: *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP),* 14-19 March. pp 2138-2141

[6] Yu L, Wang S, Lai KK. 2006. An integrated data preparation scheme for neural network data analysis. IEEE Transactions on Knowledge and Data Engineering vol 18:p 217-230

[7] Sadimon SB, Sunar MS, Mohamad D, Haron H. 2010. Computer generated caricature: A survey In: Proceedings *International Conference on Cyberworlds*, Singapore, 2010. pp 383-390

[8] Sadimon SB, Sunar MS, Haron H. 2011. A Review of Facial Caricature Generator. *Journal of computing* 3 (5):20-33

[9] Boyer V. 2005. An Artistic Portrait Caricature Model. In: Advances in Visual Computing. pp 595-600

[10]Ni F, Fu Z, Cao Q, Zhao Y. 2008. A new method for facial features quantification of caricature based on self-reference model. *International Journal of Pattern Recognition and Artificial Intelligence* 22 (8):1647-1668

[11]Le NKH, Why YP, Ashraf G. 2011. Shape Stylized Face Caricatures. In: Advances in Multimedia Modeling, vol 6523. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp 536-547.
[12]Brennan SE. 2007. Caricature Generator: The Dynamic Exaggeration of Faces by Computer. Leonardo 40 (4):392-400
[13]Benhidour H, Onisawa T (2008) Interactive face generation from verbal description using conceptual fuzzy sets. *Journal of Multimedia* 3 (2):52-59
[14]Chiang PY, Liao W-H, Li T-Y. 2004. Automatic caricature generation by analyzing facial feature. In: *Proceeding of Asian Conference on computer vision*, Jeju Island, Korea, Jan 27-30.
[15]Wenjuan C, Hongchuan Y, Minyong S, Qingjie S. 2009. Regularity-Based Caricature Synthesis. In: *International Conference on Management and Service Science,* pp 1-5
[16]Yu H, Zhang J, 2011. Mean value coordinates based caricature and expression synthesis. Signal, Image and Video Processing 7 (5):899-910.
[17]Rhodes G, Byatt G, Tremewan T, Kennedy A. 1997. Facial distinctiveness and the power of caricatures. Perception 26 (2):207-223,
[18]Murakami K, Tominaga M, Hoshimizu H. 2000. An interactive facial caricaturing system based on the gaze direction of gallery. In: *Proceedings 15th International Conference on Pattern Recognition*. pp 710-713
[19]Fujiwara T, Tominaga M, Murakami K, Koshimizu H. 2000. Web-PICASSO: Internet implementation of facial caricature system PICASSO. In: Proc. of 3rd *International Conference on Advances in Multimodal Interfaces,* Springer-verlag, Berlin. pp 151-159
[20]Thakur S, Verma L. 2012. Identification of Face Age range Group using Neural Network *International Journal of Emerging Technology and Advanced Engineering* 2 (5):250-254
[21]Sumathi G, Raju R, 2012. Software Aging Analysis of Web Server using Neural Networks. *International Journal of Artificial Intelligence & Applications* 3 (3):11-21
[22]Daramola SA, Odeghe OS, 2012, Efficient Face Recognition System using Artificial Neural Network. *International Journal of Computer Applications* 41 (21):12-15
[23]Agarwal P, Prakash N, 2013. An Efficient Back Propagation Neural Network Based Face Recognition System Using Haar Wavelet Transform and PCA *International Journal of Computer Science and Mobile Computing* 2 (5):386-395
[24]El-said SA. 2013. Reliable Face Recognition Using Artificial Neural Network. International Journal of system dynamics applications 2 (2):14-42
[25]Bashier HK, Abusham EA, Khalid F. 2012. Face Detection Based on Graph Structure and Neural Networks. Trends in Applied Sciences Research *(7):683-691*

[26]Mall A, Ghosh S, 2012. Neural Network training Based Face Detection and Recognition. *International Journal of Computer Science and Management Research 1* (2):103-109

[27]Sujatha.C, Sathiya.S SB, 2013. Three-Dimensional Face Reconstruction from a Single Image by Neural Network. *International Journal of Advanced Information Science and Technology (IJAIST) 14* (14):80-84

[28]Rupesh .N.S, Ka H.L, Eran E.A et al., 2005. Use of Neural Networks in Automatic Caricature Generation: An Approach Based on Drawing Style Capture. *IEE International Conference on Visual Information Engineering, Scotland*, 3523/2005, pp.343–351

[29]Alpaydin E .2010. Introduction to Machine Learning. The MIT Press, London, England

[30]Bishop C ,1995. Neural Networks for Pattern Recognition. Oxford Universiti Press, Walton Street, Oxford

[31]Yu H, Wilamowski BM ,2010. Levenberg Marquardt Training. In: Industrial Electronics Handbook, vol 5 - Intelligent Systems. 2nd edn. CRC Press, pp 12-11 to 12-16

[32]Pritchett JS, 2010.Caricature. http://www.pritchettcartoons.com/caricature.htm. Accessed January 2013

[33]Mitteroecker P. Gunz, P, Windhager, S and Schaefer, K,(2013). A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. *Hystrix, the Italian Journal of Mammalogy*, 24(1), pp.59–66

[34]Salmaso, L. & Brombin, C. 2013. Permutation Tests in Shape Analysis. In SpringerBriefs in Statistics. New York, NY: Springer New York, pp. 1–16.

[35]Shi, J., Samal, a. & Marx, D., 2006. How effective are landmarks and their geometry for face recognition? *Computer Vision and Image Understanding*, 102(2), pp.117–133

[36]Anies, O.S. et al., 2013. Landmark-Based Geometric Morphometrics in Describing Facial Shape of the Sama-Banguingui Tribe from the Philippines. *Journal of Medical and Bioengineering*, 2(2), pp.131–136.

[37]Sato M, Saigo Y, Hashima K, Kasuga M, 2003. An Automatic Facial Caricaturing Method for 2D Realistic Portraits Using Characteristic Points. *Journal of the 6th Asian Design International Conference*,Tsukuba, Japan vol 1 (E-40)

[38]Xu G, Kaneko M, Kurematsu A (2005) Synthesis of facial caricature using eigenspaces. Electronics and Communications in Japan, Part III: Fundamental Electronic Science (English translation of Denshi Tsushin Gakkai Ronbunshi) 87 (8):p 43-54

[39]Xu YQ, Shum HY, Cohen M, Liang L, Zhong H, 2005. Caricature exaggeration. United States Patents Application Publication, US 2005/0212821 A1

[40]Pujol, A., Villanueva, J.J. & Wechsler, H., 2000. Automatic view based caricaturing. In *International Conference on Pattern Recognition*. pp. 1072–1075.

[41]Kaneko, M. & Meguro, M., 2002. Synthesis of Facial Caricatures Using Eigenspaces and Its Applications to Humanlike Animated Agents, *7th Pacific Rim International Conference on Artificial Intelligence*. pp 1-6

[42]Landsiedel C, Edlund J, Eyben F, Neiberg D, Schuller B,2011, Syllabification of conversational speech using bidirectional long-short-term memory neural networks. In: *IEEE International Conference on Acoustics Speech and Signal Processing*, Prague, Czech Republic, 2011. pp 5256-5259

[43]Kanamori T A New Sequential Algorithm for Regression Problems by Using Mixture Distribution. In: R.Dorronsor J (ed) International Conference on Artificial Neural Network (ICANN), 2002. Lecture Notes in Computer Science. Springer vol 2415, pp 535-540

[44]Mao GQ, Liu HB (2007) Real Time Variable Bit Rate Video Traffic Prediction. *International journal of Communication system* 20 (4):491-505

[45]Mekid S, Ogedengbe T, 2010. A review of machine tool accuracy enhancement through error compensation in serial and parallel kinematic machines.*International Journal Precision Technology* vol 1(3/4):pp 251-286

[46]Webster, M. & Sheets, H., 2010. A practical introduction to landmark-based geometric morphometrics. *Quantitative Methods in Paleobiology* vol 16. pp 163-188

[47]Dal, U., Abraham, S. & Dal, D., 2011. A Facial Caricature Generation system using Adaptive Thresholding. In *2011 World Congress on Information and Communication Technologies*. IEEE, pp. 682–687