NOVEL APPROACH TO NEURAL NETWORK BASED CARICATURE GENERATION

K.H. Lai, P.W.H. Chung, E.A. Edirisinghe

Department of Computer Science, Loughborough University, UK
K.H.Lai@lboro.ac.uk Fax: +44 (0)1509 635722

Keywords: Caricature Generation, Drawing Style, Artificial Intelligence, Neural network, Pattern Recognition.

Abstract

A caricature is defined as a funny drawing of someone that makes some of his/her distinct features appear exaggerated or more amusing. However the caricatures of the same person created by different artists can be very different, since the artists drawing styles play an important role [1]. Therefore learning the drawing style of an artist provides the key to the computer based automatic generation of professional caricature. Unfortunately, no caricature generation system in the past has attempted to address this issue with the aid of artificial intelligence technologies. In this paper, we propose an example-based caricature generation system with experimental results and detailed analysis to prove that neural networks can be used for capturing the drawing style of an artist. This work is the first system to use neural networks in generating caricature.

1 Introduction

Caricatures that provide humor and entertainment are common in our daily lives, with frequent appearances in magazines and newspapers (see figure 1).

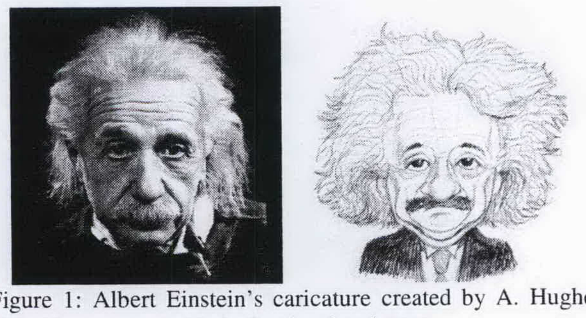
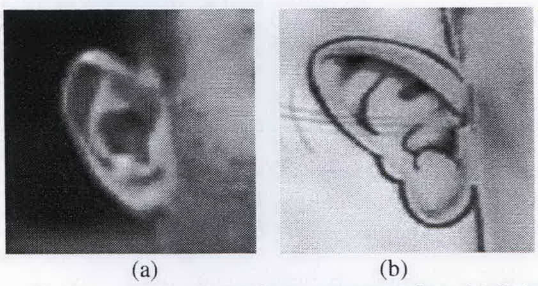


Figure 1: Albert Einstein's caricature created by A. Hughes [1] with exaggerated hair, forehead and nose.

The caricaturing process is based on deforming the features of a face selectively. A caricaturist captures the essence of the subject, exaggerates the features and distorts the less important parts (or leaves them unchanged). These change the ratios among the subject's facial features and give a deeper impression to the viewers.

Unfortunately the talent of drawing caricature only exists in few people. This inborn talent, embedded in their subconscious mind, often makes it difficult for them to explain the craft of caricature generation to others. Therefore, caricature generation by computer has become a challenging research topic [2-7] especially as the underlying processes involved cannot be accurately explained. None of the existing computer based approaches have attempted to learn the drawing style of an artist by using artificial intelligence technologies. Nevertheless this is essential if caricatures of realistic expression are to be automatically generated.

On top of this, all of the existing state-of-the-art automatic caricature generation systems only provide linear exaggeration of facial components by scaling with a factor. However, non-linear exaggerations are an unavoidable key factor in professional caricature drawings. As illustrated in figure 2, a caricatured ear is not just exaggerated in size, but also non-linearly changed in shape. Inadequate handling of non-linear changes will lead to a decrease in reality. In our previous work [8], by using simple geometric shapes, we proved that a neural network can be used to capture the non-linear differences between two drawing objects. In this paper, we further extend this work and propose a novel algorithm to automatically generate photorealistic caricature.



(a) (b)

Figure 2: An example of non-linear exaggeration. (a) Original image. (b) Corresponding caricature.

For clarity of presentation the paper is organised as follows: Section-2 presents the proposed methodology. Section-3 presents experimental results and detailed analysis proving the validity of the proposed algorithm's ability in capturing the drawing style of a particular caricaturist. Section-4 discusses the constraints and limitations of the project. Finally section-5 concludes with an insight into further research that is currently being considered as a result of it.

2 Proposed methodology

The proposed drawing style capture algorithm is summarised as a block diagram in figure 3.

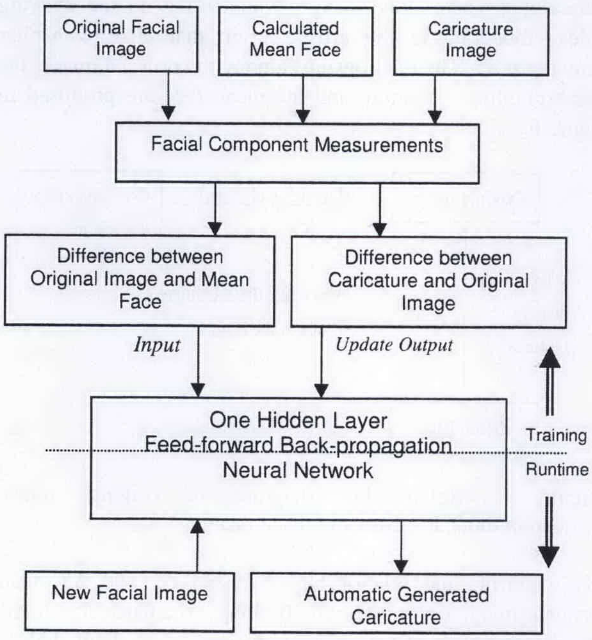


Figure 3: Proposed drawing style capture algorithm.

2.1 Face selection and normalisation

We have used the AR face database [9] to provide facial images for testing. To maintain simplicity, only natural faces are chosen, any facial images with accessories such as glasses or hats have not been considered within the context of this work. Further, all chosen images are male, with short hair.

Twelve facial images were chosen and two professional caricaturists were invited to draw caricatures. Each artist drew a set of caricatures that consisted of a drawing for each original image. An original image-caricature set drawn by one of our artists is shown in figure 4. In order to fully explore the drawing style of our artists, there were no drawing constraints to caricaturists, i.e., they were allowed to exaggerate or distort any facial component according to their styles, in contrary to previous methods. All original image-caricature pairs were subsequently normalised by using the distance between irises as the scaling factor. This is a widely accepted practice in face normalization in literature [2-7].

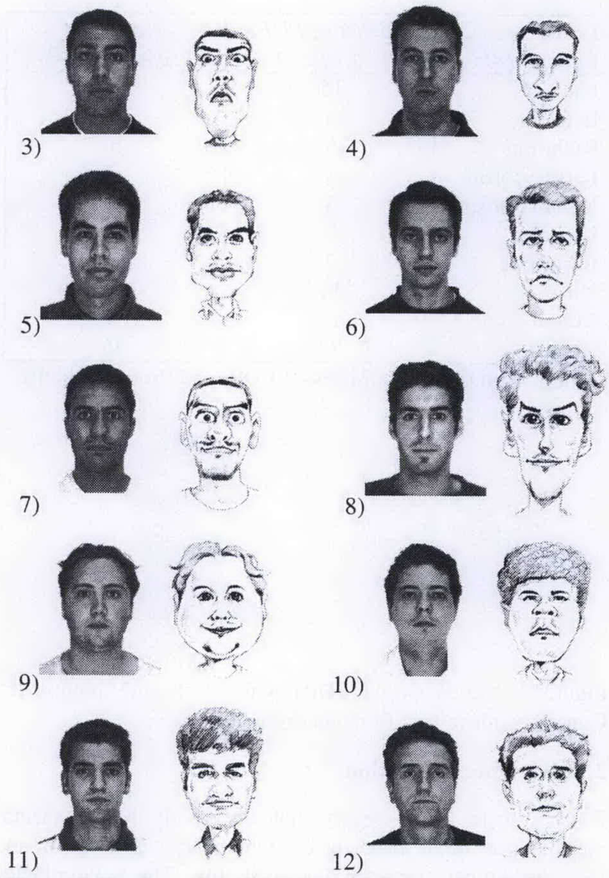


Figure 4: Twelve original images and their corresponding caricatures drawn by one of our artists.

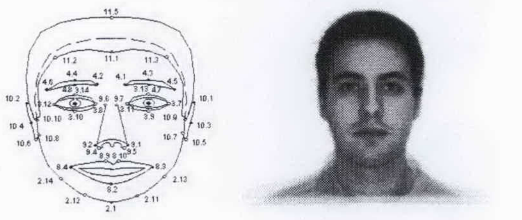
2.2 Face modeling framework

To represent and control each facial feature, a geometric face model is required. The Facial Definition Parameters (FDPs) [10] of MPEG-4 standard has been chosen as the basis for the proposed work.

As the original design of FDPs in MPEG-4 is mainly used in facial animation, many feature points are designated for high motion facial components such as lips. However such an extensive amount of feature points are not necessary in facial caricaturing, so some of them are discarded for simplicity. Moreover, some subtle and static facial feature points, such as cheeks, are also removed. Yet, two new FDPs, 4.8 and 4.7 (see figure 5a), are added for better description of eyebrows. A comparison of the FDPs in MPEG-4 standard and our proposed FDPs with 46 feature points is presented in table 1. The proposed FDPs are also presented graphically in figure 5(a). Since the main focus of research is proving that the drawing style of an artist can be captured, facial feature points are marked manually on all original images and caricatures for improved accuracy and simplicity.

Facial Components	*Number of FDPs in MPEG-4*	*Number of Proposed FDPs*
Face	15	9
Left Ear	5	5
Right Ear	5	5
Left Eyebrow	3	4
Right Eyebrow	3	4
Left Eye	7	4
Right Eye	7	4
Nose	11	6
Mouth	18	5
Total	74	46

Table 1: Comparison of MPEG-4 FDPs and Proposed FDPs.



(a) (b)

Figure 5: (a) Proposed FDPs with 46 feature points. (b) Generated mean face from ten original images.

2.3 Mean face generation

Psychologists [11] suggest that everybody has a mental visualization of a "mean face", which is an average of the faces he/she has ever seen during lifetime. The driving factor of caricature drawing, widely agreed amongst the caricaturists is, exaggerating the difference from the mean (EDFM) face [12]. In this research, the original images 1 to 10 in figure 4 are used to generate a mean face by using the freeware, Morpher of [13]. The underneath technique is referred as morphing, which is not covered here [15]. The final generated mean face is illustrated in figure 5(b).

2.4 Relationships among original image, corresponding caricature and mean face

We hypothesize the caricature generation process as follows: When a caricaturist sees a face, he/she has the ability to identify the distinctive facial features by comparing it with the mean face hidden in his/her mind. The difference between an original face, O, and a mean face, M, is defined as ΔS (The details of ΔS calculation will be covered in section 2.5).

$$\Delta S = O - M \qquad (1)$$

Then based on the original image, the artist exaggerates the distinctive facial features intentionally to form a caricature of this particular face. Therefore, the difference between the original image, O, and its corresponding caricature, C, is the change made by the artist, which is defined as $\Delta S'$ (The details of $\Delta S'$ calculation will be covered in section 2.5).

$$\Delta S' = C - O \qquad (2)$$

The relationship between ΔS and $\Delta S'$ explains why the caricaturist makes this change, which refers to the drawing rules embedded in the artist's brain that governs his/her drawing style. The relationships among the original image, the corresponding caricature and the mean face are proposed in figure 6.

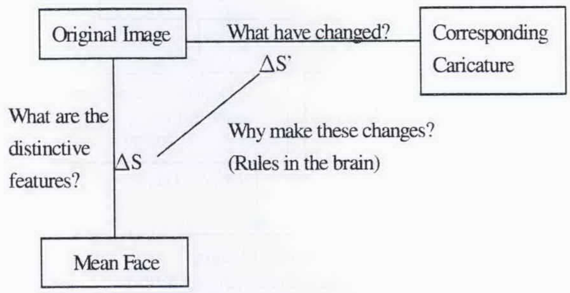


Figure 6: Relationship diagram of original image, corresponding caricature and mean face.

By capturing the relationships between ΔS and $\Delta S'$ from original image-caricature pairs (belongs to a particular artist), the drawing style of the artist can be learnt, the details will be covered in section 2.5. However, the relationship is always non-linear and difficult to describe in written language preciously. As a result, we adopt a neural network to accomplish this task.

2.5 Artificial neural network

Artificial neural network, commonly referred to as "neural network", is a type of artificial intelligence technologies that attempts to imitate the way a human brain works. There are over 10 billion biological neurons in a human brain, which cause us able to think, remember and learn. These human abilities can be simulated by connecting artificial neurons in a particular architecture to form a neural network [14].

The main reason of using neural network in this research is because of its effective learning ability. The network is capable to learn from the training set by constructing an input-output mapping for the problem automatically. Therefore, an understanding of how input is mapped to output is not necessary, which is perfectly fit for capturing the unexplainable relationship between ΔS and $\Delta S'$ that was discussed above. Moreover, a neural network has the ability to capture non-linear relationships from the training set and provides non-linear results [14]. This is suitable for capturing and mimicking non-linear exaggerations created by professional caricaturists.

The training set consists of both input and output values. To capture the relationship between ΔS and $\Delta S'$, the input and output to the neural network should be ΔS and $\Delta S'$ respectively. As illustrated in figure 7, an example of ΔS and

$\Delta S'$ of a feature point is presented. Each oval represents the contour of an eye of a face, which is defined by eight feature points. They are normalized and overlapped with each other by using iris as the reference point.

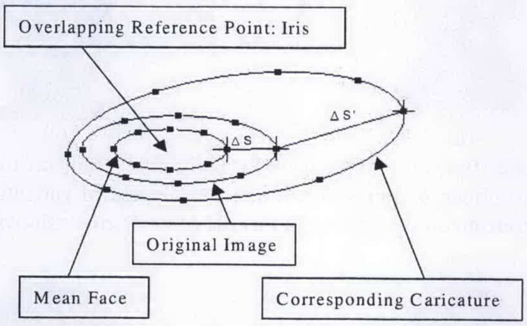


Figure 7: An example of ΔS and $\Delta S'$ of a feature point.

ΔS of an original image-caricature pair can be calculated by subtracting the X and Y coordinates of all feature points of mean face (X_M and Y_M) from their corresponding feature points of original image (X_O and Y_O) respectively.

$$X_{\Delta S} = X_O - X_M \tag{3}$$

$$Y_{\Delta S} = Y_O - Y_M \tag{4}$$

Similarly, $\Delta S'$ of the same original image-caricature pair can be calculated by subtracting the X and Y coordinates of all feature points of original image (X_O and Y_O) from their corresponding feature points of caricature (Xc and Yc) respectively.

$$X_{\Delta S'} = Xc - X_O \tag{5}$$

$$Y_{\Delta S'} = Yc - Y_O \tag{6}$$

The above calculations are repeated on all remaining original image-caricature pairs. Subsequently, the X and Y training set tables are constructed. In this research, only the first ten original image-caricature pairs (1-10 of figure 4) are used for training. The remaining two original image-caricature pairs (11-12 of figure 4) are reserved for validations; further details will be discussed in section 3. In order to reduce the requirement of computer resource during the neural network training process, X and Y training sets are trained separately in two different neural networks with the same architecture and parameters.

Once the training sets have been prepared, the next step is to define neural networks. In this project, all experiments were carried out by using MATLAB neural network toolbox [16]. We used feed-forward back-propagation network with only one hidden layer as the architectures. The number of nodes of input, hidden and output layers in each neural network was 46, which was the same as the number of feature points proposed in table 1. The neural networks were trained by using Levenberg-Marquardt algorithm [16] without momentum. The mean squared error was used as the performance functions and the performance goals were set to zero. The trainings stopped once the performances were minimized to the goals or the gradients of performances were less than the minimum gradient parameter, which means the error slope is close to zero and further training cannot reduce the error by much [16]. A summary of the neural network architecture and the training parameters are shown in table 2.

Architecture and Parameters	*Choice*
Number of neural network required per artist	2
Neural Network Type	Feed-forward back-propagation
Training Function	Levenberg-Marquardt
Performance Validation Function	Mean Squared Error
Performance Goal	0
Minimum Gradient	1e-010
Number of Layers	3
Hidden Layer Transfer Function	Tan-sigmoid
Output Layer Transfer Function	Pure-linear
Number of nodes in input layer	46
Number of nodes in hidden layer	46
Number of nodes in output layer	46

Table 2: Neural network architecture and training parameters.

After training the neural networks, the next step is validations. (The validation results and full analysis will be covered in section 3.) Subsequently, any new original image can be sampled and fed as input to the trained neural networks for testing. The outputs will be the $X_{\Delta S'}$ and $Y_{\Delta S'}$ of the testing image, which can then be used to calculate the XY coordinates of the newly generated caricature.

2.6 Mesh warping

Mesh warping is the final module of the system that deforms an original test image into a caricature [15]. The module accepts an original facial image, a source face mesh and a target face mesh. These two meshes define how to deform an image into a caricature. In this project, the source face mesh is constructed by the XY coordinates of feature points of the original image, and the target face mesh is created by the XY coordinates of feature points generated from the neural networks (see section 2.5). Then the mesh warping module converts the original test image into a caricature, following the captured drawing style of the neural networks.

3 Experimental results and analysis

In the first experiment, the neural networks are trained by using the drawings of our first artist. Afterward the ΔS of the remaining two original image-caricature pairs (11-12 of figure 4) pass through the trained neural networks separately for validations. The generated numerical results will then enter the mesh warping module for caricature deformations. The final photorealistic outputs are compared with the caricatures drawn by our first artist in figure 8 and 9 for validations.

(a) (b) (c)

Figure 8: (a) Original image 11 of figure 4. (b) Caricature of 8(a) drawn by our first artist. (c) Generated caricature of 8(a) from our system trained by the first artist's drawings.

(a) (b) (c)

Figure 9: (a) Original image 12 of figure 4. (b) Caricature of 9(a) drawn by our first artist. (c) Generated caricature of 9(a) from our system trained by the first artist's drawings.

In figure 8, by comparing the caricature drawn by our artist, 8(b), with the caricature generated by our system, 8(c), it can be shown that some of the drawing styles have been picked up successfully. First of all, the height of the forehead in 8(b) is shortened when compared with 8(a). A very similar forehead distortion happens in 8(c). Besides, similar exaggerations of the noses appear in both 8(b) and 8(c). Not only are the widths of noses increased, but also the lengths are elongated. Moreover, the changes of mouths in 8(b) and 8(c) are very close to each other. Both of them are stretched in width but not in height. Finally, the shapes of the eyes and eyebrows in 8(c) only slightly changed when compared with 8(a). These are similar to 8(b) as our artist did not modify these components vigorously, so these are also considered as the picked up drawing styles. Note that both hair and neck are not considered in our current system as they are not in the MPEG-4 FDPs standard, so these changes could not be captured.

In figure 9, by comparing the caricature drawn by our artist, 9(b), with the caricature generated by our system, 9(c), satisfactory results are demonstrated. Firstly, the ways of nose exaggeration in 9(b) and 9(c) are very close to each other, both of them are slightly bigger than the original image 9(a). On top of this, similar exaggerations of ears appear in both 9(b) and 9(c); their shapes and exaggeration ratios are almost the same. Furthermore, the sizes of mouths of 9(b) and 9(c) are both wider than 9(a) as the shapes remain unchanged. Moreover, the success of capturing drawing style is very obvious at the chins of 9(b) and 9(c), both of them are elongated heavily. Finally, both eyebrows and eyes of 9(c) are slightly wider than 9(a); these changes also match the drawing of our artist in 9(b).

(a) (b) (c)

Figure 10: (a) Original image 11 of figure 4. (b) Caricature of 10(a) drawn by our second artist. (c) Generated caricature of 10(a) from our system trained by the second artist's drawings.

(a) (b) (c)

Figure 11: (a) Original image 12 of figure 4. (b) Caricature of 11(a) drawn by our second artist. (c) Generated caricature of 11(a) from our system trained by the second artist's drawings.

In order to prove the validity of our proposed drawing style capture algorithm further, the second experiment has been carried out. The same methodology mentioned in section 2 with newly created neural networks has been applied on the drawings of our second artist. (Due to the page limitation, the training set caricatures drawn by our second artist are not shown here.) Figure 10 and 11 are the validations of the trained system. In figure 10(c), our caricature system captured the drawing style once again. The heavily elongated face and exaggerated nose are similar to the changes made by our second artist in 10(b). The slightly exaggerated mouth and elongated ears of 10(c) also match the drawings of 10(b).

Likewise, in figure 11(c), the generated face is slightly elongated when compared with 11(a), this also happens in the caricature 11(b) that drawn by our second artist. Both size and shape changes of nose in 11(c) are very close to 11(b), the success of style capturing is obvious in this component. Finally, the slightly caricatured eyebrows, the widened eyes and remain unchanged mouth are similar in 11(b) and 11(c). Although a slightly tilted face and a dent at the chin cannot be seen in 11(c), a careful investigation revealed that these two drawing styles do not appear in the training set; hence our system cannot imitate them.

In the above analysis, justify the style similarity of two drawings could be subjective, and it is difficult to convince all the viewers that the styles of two caricatures are the same. After all, exactly capture and predict the style of a particular artist is an extremely difficult task. However, by comparing the results of the first and second experiments, the success of style capturing can be seen ultimately. When comparing computer generated caricatures 8(c) and 10(c) with the drawings of artists, 8(b) and 10(b), 8(c) looks closer to 8(b) than 10(b). On the other hand, 10(c) looks similar to 10(b) rather than 8(b). Similarly, when comparing computer generated caricatures 9(c) and 11(c) with the drawings of artists, 9(b) and 11(b), 9(c) looks closer to 9(b) than 11(b). And 11(c) looks similar to 11(b) rather than 9(b). These comparisons demonstrated that the drawing rules of the first and second artists have been captured in the first and second experiments respectively, even though our system cannot predict and generate caricatures exactly the same as their drawings.

The validation experiments and comparisons of system outputs for both artists provide satisfactory results. The generated caricatures, 8(c), 9(c), 10(c) and 11(c) illustrated that the trained neural networks have captured the drawing styles of our artists. As a result, the trained system can be putted into practises.

4 Discussions

The proposed novel example-based neural network approach has captured the drawing style of a particular artist successfully, as proved in section 3. Training on a larger set of original image-caricature pairs (belonging to different people but caricatured by the same artist) will help improve prediction further. However, data collection has been a major obstacle of the project. It is unreasonable and time consuming to request an artist to draw too many caricatures for analysis, which also limits the applicability of the work. As a result, a key challenge of the project is to capture the drawing style of an artist from a limited dataset and still obtain satisfactory results.

Due to the fact that esthetic judgment of the resulting photorealistic caricatures could be a subjective task, a comprehensive subjective test is required to further analyse and evaluate more generated caricatures, which helps improve our system in the next stage.

The above research is a part of a more advanced research project that is exploring the areas of caricaturing in different genders, ages and races. Besides, the hair and accessories will also be considered.

5 Conclusions

In this paper we have identified the drawbacks of existing caricature generation systems in their inability to capture non-linear relationship and the drawing style of a particular artist. We have proposed an example-based neural network approach to address the above issues by training neural networks on a small set of caricatures made by an artist. The proposed system has been evaluated by using the drawings of two professional caricaturists, both results proved that neural networks can be used for capturing the drawing style of an artist. The system is then capable to generate photorealistic caricatures with embedded styles automatically.

In the future, the trained neural networks can further be processed by artificial intelligence technology to extract linguistic rules from them [17]. This will help in the understanding and drawing style comparisons of different artists in both, artistic and psychological ways.

Acknowledgments

We would like to thank Mr. Yang and Mr. Abuhelga for their contributions in drawing caricatures for the project.

References

[1] A. Hughes. "Learn to draw caricatures", *HarperCollins Publishers*, London, (1999).

[2] S. E. Brennan. "Caricature generation", *M.S. thesis*, Depart. Vision Studies, MIT, (1982).

[3] P. J. Benson, D. I. Perrett. "Synthesising continuous-tone caricatures", *Image and Vision Computing*, **vol. 9**, pp. 123-129, (1991).

[4] J. Nishino, T. Kamyama, H. Shira, T. Odaka, H. Ogura. "Linguistic knowledge acquisition system on facial caricature drawing system", *IEEE Int. Conf. on Fuzzy Systems*, **vol. 3**, pp. 1591–1596, (1999).

[5] B. Gooch, E. Reinhard, A. Gooch. "Human Facial Illustrations: Creation and Psychophysical Evaluation", in *ACM Transactions on Graphics*, New York, **vol. 23**, issue 1, pp. 27-44, (2004).

[6] L. Liang, H. Chen, Y. Q. Xu, H. Y. Shum. "Example-based caricature generation with exaggeration", *Proceedings of 10th Pacific Conference on Computer Graphics and Applications*, pp. 386-393, (2002).

[7] A. Pujol. "Automatic view based caricaturing", in *Proc. of ICPR*, Barcelona, **vol. 1**, pp. 1072-1075, (2000).

[8] R. N. Shet, K. H. Lai, E. A. Edirisinghe, P. W. H. Chung. "Use of neural networks in automatic caricature generation: an approach based on drawing style capture", *IEE Int. Conf. on Visual Information Engineering*, Scotland, pp. 23-29, (2005).

[9] A. M. Martinez, R. Benavente. "The AR face database", *CVC Technical Report* #24, June, (1998).

[10] MPEG Systems. "Text of ISO/IEC FDIS 14496-1: Systems", *Doc. ISO/MPEG N2501*, Atlantic City MPEG Meeting, October, (1998).

[11] G. Rhodes, T. Tremewan. "Averageness, exaggeration and facial attractiveness", in *Psychological Science*, **vol. 7**, pp. 105-110, (1996).

[12] L. Redman. "How to Draw Caricatures", *McGraw-Hill Publishers*, (1984).

[13] M. Fujimiya. (2005, September 10). "Morpher (version 3.1)" [Online]. Available: http://www.asahi-net.or.jp/~FX6M-FJMY/mop00e.html

[14] L. Fausett. "Fundamentals of Neural Network", *NY: Prentice Hall*, (1994).

[15] G. Wolberg. "Digital image warping", *IEEE Computer Society Press*, CA: Los Alamitos, (1990).

[16] The Maths Works Inc. "User's guide version 4, neural network toolbox", *MATLAB*, (2002).

[17] L. Fu. "Rule generation from neural networks", *IEEE Trans. Systems, Man and Cybernetics*, **vol. 24**, pp. 1114-1124, (1994).