PHOTOGENIC FACIAL EXPRESSION DISCRIMINATION

Luana Bezerra Batista and Herman Martins Gomes
Departamento de Sistemas e Computação

João Marques de Carvalho
Departamento de Engenharia Elétrica
Universidade Federal de Campina Grande
Campina Grande, Paraíba, Brasil, 58.109-970

Keywords: Facial Expression Recognition, Photogeny, Principal Component Analysis, Multi-Layer Perceptron.

Abstract: Facial Expression Recognition Systems (FERS) are usually applied to human-machine interfaces, enabling services that require identification of the emotional state of the user. This paper presents a new approach to the facial expression recognition problem, by addressing the question of whether or not it is possible to classify previously labeled photogenic and non-photogenic face images, based on their appearance. A Multi-Layer Perceptron (MLP) is trained with PCA representations of the face images to learn the relationships between facial expressions and the concept of a good photography of the face of a person. In the experiments, the generalization performances using MLP and Support Vector Machines (SVM) were analyzed. The results have shown that Principal Component Analysis (PCA) combined with MLP represent a promising approach to the problem.

1 INTRODUCTION

Facial expressions are a manifestation of the emotional state, cognitive activity, intention, personality and psychopathology of a person (Donato et al., 1999).

According to Mehrabian (Mehrabian, 1968), the verbal part of a spoken message contributes only with 7% to the effect of the message as a whole; the voice intonation contributes with 38%, while facial expressions alone are responsible for 55% of the message information. These values clearly show that facial expressions play a major role in human communication (Pantic and Rothkrantz, 2000).

Facial Expression Recognition Systems (FERS) are generally applied to human-machine interfaces (van Dam, 2000) (Pentland, 2000) (Zue and Glass, 2000). Such interfaces enable the automation of services that require appreciation of the emotional state of the user, as in transactions that involve some form of negotiation (Chibelushi and Bourel, 2003).

The two main approaches used for facial expression recognition are based on Action Units (Donato et al., 1999) and on Basic Expressions (Ekman, 1982):

- Based on global facial features, Basic Expressions (BEs) relate to the emotional states of joy, sadness, surprise, anger, fear and disgust;
- An Action Unit (AU) is one of 46 atomic elements of visible facial movements or its associated deformation, being therefore based on local face features. An expression results from the agglomeration of several AUs.

In this paper, instead of trying to infer the emotional states of an expression or extracting features related to facial movements, we formulate a different problem and approach, by designing experiments that use a new set of global and local features to discriminate between photogenic and non-photogenic expressions. According to Wikipedia online free encyclopedia (http://en.wikipedia.org/wiki/Photogenic), the definition of the term *photogenic* is:

Bezerra Batista L., Martins Gomes H. and Marques de Carvalho J. (2006).
PHOTOGENIC FACIAL EXPRESSION DISCRIMINATION.
In *Proceedings of the First International Conference on Computer Vision Theory and Applications*, pages 166-171
DOI: 10.5220/0001370901660171

> "Attractive as a subject of photography. A person that looks attractive on pictures."

Attractiveness is a very subjective concept, which may be difficult to map into a more formal definition. Some authors may link this to the concept of beauty and symmetry, but this is not the direction we want to follow. For the purpose of this work, we associated photogenic pictures to smiling and neutral faces using the common sense idea that when people are asked to pose for a picture, they usually make a smiling face (rarely they use expressions such as anger or sadness). In the future, instead of this coarse classification, we intend to refine the concept of *photogeny* by acquiring knowledge from a set of images that have been voted by a number of human observers.

The main goal of this work is, therefore, to give a new focus to the problem of facial expression recognition, by addressing the *photogeny* question. This means to investigate the relationship between the facial expressions presented by a human subject and the concept of a good photography of that person.

This paper is organized as follows: section 2 discusses previous related work; in section 3 the *photogeny* discrimination framework is described; section 4 presents the performed experiments and results; and. finally, section 5 draws conclusions and presents proposals for future work.

2 RELATED WORK

The *photogeny* problem has not yet been studied in Computer Vision literature. However, there is some related work on facial expression recognition, which will be discussed here.

In the work of Zhang et. al (Zhang et al, 1998), Gabor filters combined with Neural Networks were used to recognize BEs. Gabor filters are applied at the location of 34 fiducial points, producing a better recognition rate (92.2%) than when only geometric positions (coordinates of the fiducial points) (73.3%) are used.

In the work of Feitosa et al. (Feitosa et al, 2000), PCA and Neural Networks were used to recognize BEs on the JAFFE database (Lyons et al, 1998). RBF (Radial Basis Function) reached a recognition rate a little higher (73.2%) than MLP in their best configurations (71.8%). However, MLP was more stable than RBF regarding changes of Principal Components and among the classes.

Bartlett et al. (Bartlett et al., 2002) used Gabor filters and SVM to recognize three kinds of AUs: Blinks, Brow Raising and Brow Lowering. A nonlinear SVM applied to the Gabor representations obtained 95.9% of correct classification for discriminating blinks from non-blinks AUs.

In the work of Nakano et al. (Nakano et al, 2002), Simple Principal Component Analysis (SPCA) were used to extract features from smiles. The value of $\cos\theta$, being θ the angle between the eigenvector and the gray scale vector of each image, was calculated and used as input to a MLP. The average rate of correct classification discriminating between true (natural) and false (plastic/forced) smile was 92.0%.

In the work of Kapoor et al. (Kapoor et al., 2003), PCA and SVM were used to recognize facial action units related to upper facial muscle movements, such as inner eyebrow raising, eye widening, etc. Using the Cohn-Kanade Facial Expression Database (Kanade et al, 2000), the system reached an accuracy of 81.22%.

Matsugu et al. (Matsugu et al., 2003) proposed a rule-based facial analysis to distinguish smiling/laughing faces from others BEs based on variations of some face parameters as the expression changes from neutral to smiling. A score is calculated to quantify the variations and thresholded for deciding whether the subject is smiling or not. Experimental results demonstrated reliable detection of smiles with correct recognition rate of 97.6%.

Shinohara and Otsu (Shinohara and Otsu, 2004) used Higher Order Local Auto-Correlation (HLAC) features and Fisher weight maps to discriminate between neutral and smiling faces. The recognition rate of the proposed method was 97.9%, while Fisherfaces method was 93.8% and HLAC without a weight map was 72.9%.

3 PHOTOGENY DISCRIMINATION FRAMEWORK

Our main goal is to train a classifier to learn the relationships between face expressions and the concept of a photogenic picture of a person.

In this section, we present a methodology designed to the *photogeny* problem. Figure 1 shows the steps composing the methodology.

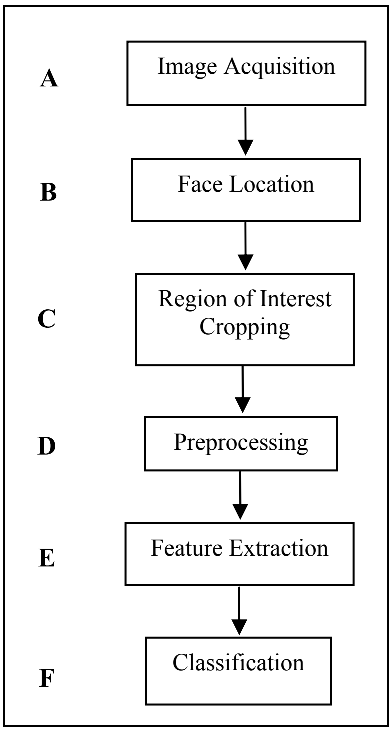

Figure 1: Photogeny discrimination framework

In the first step (block A in Figure 1), we selected a subset from the Cohn-Kanade Facial Expression Database. The pictures corresponding to neutral and happiness expressions were labeled as photogenic; whereas the pictures corresponding to the others expressions were labeled as non-photogenic. This re-labeling was based on a subjective evaluation of all images in the database. The preprocessing step (block D in Figure 1) is composed by the operations *Resizing*, *Gray Level Transformation* and *Histogram Equalization*. Whereas the others steps (blocks C, E and F) are specific to each experiment performed (see Section 4).

In this paper, we assume that the problem Face Location (block B in Figure 1) is solved. For an extensive review in this area, see the paper of Hjelmas and Low (Hjelmas and Low, 2001).

4 EXPERIMENTS

To perform the experiments, we selected a set of 324 images from the Cohn-Kanade Facial Expression Database. A total of 162 pictures were labeled as photogenic; whereas 162 pictures were labeled as non-photogenic. The subset was separated in training (75%; 244 images) and testing (25%; 80 images), so that the people contained into the training set are not contained into test set. Table 1 shows some examples of this image set.

Initially, we investigated the impact of applying Gabor filters (Lee, 1996), as feature extractors, in the following regions: (i) left side of the face, (ii) left side of the mouth, (iii) left eye and (iv) left side of the mouth and left eye. The choice of the left side is motivated by a study that shows this area is moved more extensively during facial expression changes (Borod et al. 1998). Additionally, we used the extracted features to compare the discriminating performance of the SVMs with the K-Nearest Neighbor (K-NN) classifier, for distinguishing photogenic from non-photogenic faces.

The results in Tables 2 and 3 show that SVM achieved better correct discrimination rates than K-NN (77.50% versus 71.25%, respectively).

Table 1: Examples of photogenic and non-photogenic pictures.

Photogenic	Non-photogenic

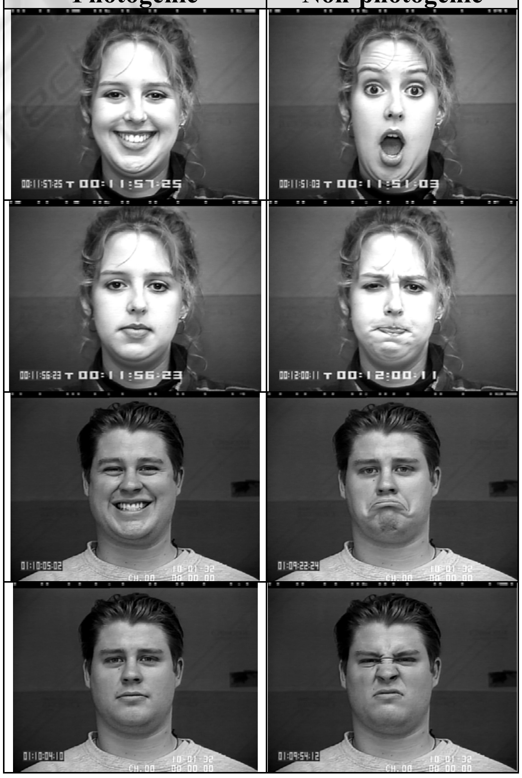

Table 2: Correct Discrimination Rates using SVM.

Regions/Classifier	SVM
left side of the face	75.00% *C*-SVC + Polinomial Kernel
left side of the mouth	**77.50%** ***C*-SVC + Polinomial Kernel**
left eye	62.50% *C*-SVC + RBF Kernel
left side of the mouth + left eye	73.75% *C*-SVC + Polinomial Kernel

Table 3: Correct Discrimination Rates using K-NN.

Regions/Classifier	K-NN
whole image	65.00% $k = 1$ or $k = 2$
left side of the mouth	71.25% $k = 2$
left eye	56.25% $k = 1$
left side of the mouth + left eye	65.00% $k = 2$

From Tables 2 and 3, we can also conclude that only the *left side of the mouth* is necessary to discriminate between the classes. Therefore, from this step on, we considered only that part of the face as our region of interest (ROI) (see Figure 2).

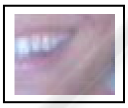

Figure 2: Region of interest.

After extracting the left sides of the mouth, the corresponding sub-images were resized to 20x25 pixels and transformed to 256 gray levels. Next, histogram equalization was performed. These operations are illustrated in Figure 3. Finally, a number of Principal Components were extracted from these images.

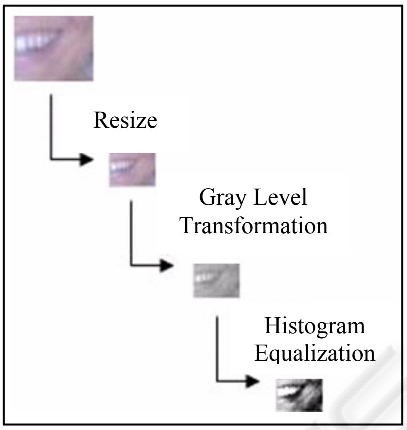

Figure 3: Preprocessing steps.

We began the experiments using SVM (Vapnik, 1999) as classifier - a kernel-based learning machine that has been successfully used for pattern recognition – in order to perform a later comparative study with MLP. The number of Principal Components (PCs) was varied from 3 to the maximum. That is, we used 3, 5, 8, 11, 16, 28, 56, 90, 133 and 242 (which contribute more than 2%, 1.25%, 1%, 0.75%, 0.5%, 0.25%, 0.1%, 0.05%, 0.025% and 0%, respectively, to the variance in the data set) components to train 10 SVMs.

Each SVM was trained with parameters automatically obtained from the "grid.py" script, available at LibSVM toolbox (Chang and Lin, 2005). This script is a model selection tool for C-SVC classification using RBF kernel. It uses the cross validation method to estimate the accuracy of each parameter combination; finding, therefore, the best parameters for a specific problem.

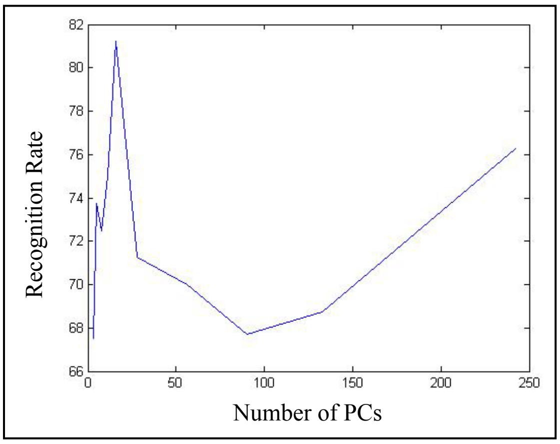

Figure 4: Number of PCs *versus* Recognition Rate.

From Figure 4, we can observe that using only 16 PCs - which contribute more than 0.5% to the variance in the data set – the best recognition rate is reached, that is, 81.25%.

Once obtained the number of PCs necessary to discriminate between the 2 classes studied in this article, we performed another experiment using a MLP as classifier. The number of hidden neurons was varied from 1 to 10, while the number of PCs was fixed in 16. Figure 5 shows that the best recognition rate, 87.5%, was obtained using 4 neurons on the hidden layer. Table 4 presents the confusion matrix for this experiment.

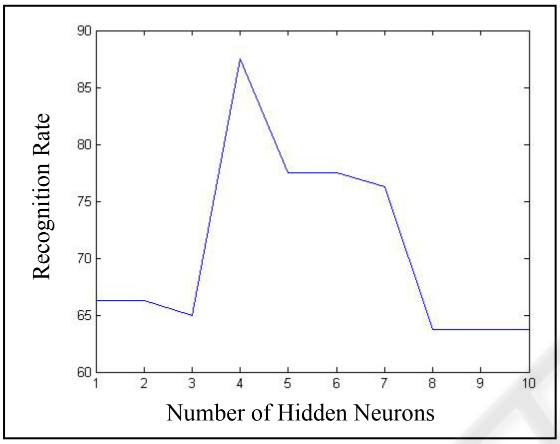

Figure 5: Number of Hidden Neurons *versus* Recognition Rate

Table 4: Confusion Matrix.

	Photogenic	Non- Photogenic
Photogenic	**36**	4
Non- Photogenic	6	**34**

From this result, it is possible to conclude that the combination PCA with MLP is more suitable to the *photogeny* problem than the utilization of Gabor filters with SVM.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we present a novel methodology that linked facial expressions with the concept of a photogenic picture of the face of a person. PCA was used to extract features from the images while a Neural Network was tested as classifier.

In the experiments reported, a comparison between MLP and SVM was performed; and different numbers of Principal Components and hidden neurons were tested. The experiments have shown that both PCA and MLP are promising, having achieved good recognition rates, similar to the ones in the existing work on specific class facial expression recognition. However, it is important to emphasize that we cannot perform a direct comparison with other previous methods, since the idea here is to deal with the problem of *photogeny*, not facial expression recognition.

The work of Elkman (Ekman, 1982) constitutes a solid foundation for many facial expression analysis works. One important difficulty with the classification of photogenic pictures is due to the high subjectivity involved in labeling the datasets. Therefore, our ultimate goal is to define the basis for this new area. This paper represents an initial effort towards this goal and is restricted to a more intuitive/obvious subset of photogenic faces (neutral and happy).

As future work we intend to incorporate in the experiments images containing facial expressions of people with the eyes closed. Another future work is to create a custom-built larger image database and use a voting scheme to assign labels (e.g. photogenic, non-photogenic) to the images. Finally, we intend to use Bayesian Regularization (Foresee and Hagan, 1997) in order to obtain the best MLP architecture.

ACKNOWLEDGEMENTS

The authors would like to thank Hewlett Packard for their support and collaboration and Professor Walfredo da Costa Cirne Filho for his useful comments and suggestions for improving this paper.

The authors also would like to thank Professor Jeffrey Cohn for granting access to the Cohn-Kanade AU-Coded Facial Expression Database.

REFERENCES

Bartlett, M., Littlewort, G., Braathen, B., Sejnowski, T. and Movellan, J., 2002. An Approach to Automatic Analysis of Spontaneous Facial Expressions. *Neural Information Processing Systems*.

Borod, J, Koff, E., Yecker, S., Santschi-Haywood, C. and Schmidt, J., 1998. Facial asymmetry during emotional expression: Gender, valence, and measurement technique. *Neuropsychologia*, vol. 36, no. 11, pp. 1209-1215.

Chang, C. and Lin, J. LIBSVM v2.8: a library for support vector machines, 2005. Available at online http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Chibelushi, C. and Bourel, F., 2003. Facial Expression Recognition: A Brief Tutorial Overview. In *On-Line Compendium of Computer Vision*.

van Dam, A., 2000. Beyond WIMP. In *IEEE Computer Graphics and Applications,* vol. 20, no. 1, pp. 50-51.

Donato, G, Bartlett, M., Hager, J., Ekman, P. and Sejnowski, T., 1999. Classifying Facial Actions*. In IEEE Trans.Pattern Analysis and Machine Intelligence*. vol. 21, no. 10, pp. 974-989.

Ekman, P., 1982. *Emotion in the Human Face*, Cambridge University Press.

Feitosa, R., Vellasco, M., Oliveira, D., Andrade, D. and Maffra, S., 2000. Facial Expression Classification using RBF and Back-Propagation Neural Networks. In *4th World Multiconference on Systemics, Cybernetics and Informatics and the 6th International Conference on Information Systems Analysis and Synthesis*, pp. 73-77.

Foresee, F. and Hagan, M., 1997. Gauss-Newton approximation to Bayesian regularization. In *International Joint Conference on Neural Networks*, pp. 1930-1935.

Haykin, S., 1998. *Neural Networks: A Comprehensive Foundation*, 2nd Edition, Prentice Hall.

Hjelmas, E. and Low, B., 2001. Face Detection: A Survey. In *Image and Vision Understanding*, vol.83, pp. 236-274.

Hornik, K., Stinchcombe, M. and White, H., 1989. Multilayer Feedforward Networks are Universal Approximators. In *Neural Networks*, vol. 2, pp. 359-366.

Kanade, T., Cohn, J. and Tian, Y., 2000. Comprehensive Database for Facial Expression Analysis. *In 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 46-53.

Kapoor, A., Qi, Y. and Picard, R, 2003. Fully Automatic Upper Facial Action Recognition. In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*.

Lee, T., 1996. Image representation using 2d Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 959–971.

Lyons, M, Akamatsu, S., Kamachi, M. and Gyoba, J., 1998. Coding Facial Expressions with Gabor Wavelets. *IEEE International Conference on Automatic Face and Gesture Recognition*.

Matsugu, M, Mori, K., Mitarai, Y. and Kaneda, Y., 2003. Facial expression recognition combined with robust face detection in a convolutional neural network. *International Joint Conference on Neural Networks*, vol. 3, pp. 2243 - 2246.

Mehrabian, A., 1968. Communication without Words. In *Psychology Today*, vol. 2, no. 4, pp 53-56.

Nakano, M., Mitsukura, Y., Fukumi, M. and Akamatsu, N., 2002. True Smile Recognition System using Neural Networks. In *International Conference on Neural Information*, pp. 1-5.

Pantic, M. and Rothkrantz, L., 2000. Automatic Analysis of Facial Expressions: The State of the Art. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp.1424-1445.

Pentland, A., 2000. Looking at People: Sensing for Ubiquitous and Wearable Computing. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 107-119.

Shinohara, Y. and Otsu, N., 2004. Facial Expression Recognition Using Fisher Weight Maps", *International Conference on Automatic Face and Gesture Recognition*, pp. 499-504.

Vapnik, V., 1999. *The Nature of Statistical Learning Theory*, 2nd Edition, Springer-Verlag, New York.

Zhang, Z., Lyons, M., Schuster, M. and Akamatsu, S., 1998. Comparison between geometry-based and Gabor wavelets based facial expression recognition using multi-layer perceptron". *IEEE International Conference on Automatic Face and Gesture Recognition*.

Zue, V. and Glass, J., 2000. Conversational Interfaces: Advances and Challenges. In *IEEE*, vol. 88, no. 8, pp. 1166-1180.