Color Face Recognition by Auto-regressive Moving Averaging

Mehmet Celenk and Inad Al-Jarrah
School of Electrical Engineering and Computer Science
Stocker Center, Ohio University
Athens, Ohio

Abstract

Human face identification is a main computational step for many information-processing applications including security checkpoints, surveillance systems, video conferencing, and picture telephony. A new approach is presented for recognizing human faces and discriminating expressions associated with them in color images. It is a statistical technique based on the process of drawing facial silhouettes and characterizing them by auto-regressive moving average (ARMA), which, is, in turn, infinite impulse response (IIR) filtering. First, a facial image is transformed from its (R, G, B) space to its principal component representation. A line-drawing profile of the face image is created from its principal component using the zero-crossings of a Laplacian of Gaussian (LoG) filter. The face line-silhouette is then partitioned into 5×5 non-overlapping blocks, each of which is filtered by a non-causal IIR filter. The IIR coefficients are approximated by the ARMA parameter vector **a**. By computing the ensample average of **a** over the whole image area, we obtain the ARMA feature vector of the facial pattern. Face discrimination is achieved by the non-metric similarity measure $S = |\cos \angle(\mathbf{a}.\mathbf{b})|$ for two face patterns whose feature vectors (**a** and **b**) consist of the aforementioned ARMA coefficients. Experimental results obtained from a small database indicate that the ARMA modeling is capable of discriminating facial color images, and has the ability of distinguishing facial expressions.

Background Information

To locate and understand different faces and facial expressions comes naturally to human beings. Using the computer to achieve this objective is not an easy task. A complete face-identification system includes two computational modules:[6,7] Detecting the location and size of a face is the first unit, involving image processing operations. It is proven to be complex because of the unknown location, position, and arbitration of faces in a given random image. The second stage is to obtain the facial features for recognition of faces and classification of emotional expressions.[8] Analytic (structural) and holistic (statistical) are two approaches taken for facial feature extraction.[9] Structural features are local properties governed by such distribution models as a point distribution model (PDM),[1,2] a gray-level model (GLM),[3] or a combination of both.[4] These are not affected very much by variation due to irrelevant data (e.g., hair), and foreground and background characteristics. Statistical features are, on the other hand, more global descriptors. Consequently, they are expected to be less immune to the aforementioned variations. In the literature, hybrid methods are also proposed and used to compensate for the effect of perspective variations.[9]

This paper introduces a new approach to the problem of human face recognition and classification using color images. The method described here relies on the statistical pattern classification power of the non-causal IIR filters.[10] The filter coefficients are computed in three steps. First, the discrete Karhunen-Loeve (KL) transform[5] is applied to the R, G, and B dimensions to represent the image by its principal component (i.e., projection onto the normalized eigenvector corresponding to the largest eigenvalue). The principal component represented image is converted to a line-drawing form using a LoG filter.[10-12] The line-drawing image is then partitioned into 5×5 non-overlapping blocks, each of which is modeled by the ARMA parameters' vector **a**. Although the dimension of **a** is 25, for a symmetric IIR filter, the unknown number of coefficients is 12 since $a(0,0)=1$ and $a(k,l) = a(-k,-l)$, respectively. By taking the ensample average of **a** over the whole image, we obtain the corresponding IIR filter representation of a facial pattern.

For face discrimination and classification, we use a non-metric similarity measure $S=|\cos\angle(\mathbf{a}.\mathbf{b})|$[13] for images whose IIR coefficient vectors (**a** and **b**) are the ensample average of their corresponding ARMA parameters. Notice that the adapted similarity measure (S) is the absolute value of the cosine of the angle between the vectors **a** and **b**. Its value varies within the range of 0 to 1 (i.e., $0 \leq S \leq 1$). Clearly, for two identical images, the measurement value of S appears to be 1; while for two completely different pictures it becomes 0. Obviously, these limits are for two extreme cases as obtained in accordance with the definition for S, which can hardly be achieved in practice due to noise and other image imperfections.

Effectiveness of the IIR filtering and ARMA modeling is analyzed by using a small database of facial images. The database contains three different sets of images for three different persons. Each set consists of pictures of three different facial expressions (i.e., normal face, happy face, and sad face) of each individual. Our computer results have shown that the method can recognize the faces of different people. It appears to distinguish different facial expression of the same person.

In the remaining part of this paper, we first discuss a line-drawing operation for facial images based on the LoG edge detection. We then describe the IIR filtering and the ARMA modeling. This is followed by a computer implementation and experimental result section. The paper ends by the conclusions and further research topics.

Line-Drawing of Face Profiles

Line-drawing representation of human faces is obtained by applying a Laplacian of Gaussian filter to the principal component represented facial image. The kernel of LoG is given by[15]

$$LoG(r,c) = \left(\frac{\partial^2}{\partial r^2} + \frac{\partial^2}{\partial c^2} \right) \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\left(\frac{r^2+c^2}{\sigma^2}\right)} \tag{1}$$

or, equivalently,

$$LoG(r,c) = \frac{-1}{2\pi\sigma^4} \left(2 - \frac{r^2+c^2}{\sigma^2} \right) e^{-\frac{1}{2}\left(\frac{r^2+c^2}{\sigma^2}\right)} \tag{2}$$

where σ is the standard deviation of the Gaussian kernel in a selected window of size r × c. The LoG produces a continuous (unbroken) edge profile of a facial pattern if we consider the LoG's zero-crossing points. We select this operation before performing IIR filtering because it is expected to emphasize edges in the faces than other details. This is accomplished by ignoring the blocks that do not include any edge points. If a block has all zero values after performing the LoG filter, it makes no contribution to the average value of the feature vector (i.e., **a**). As a result, it will not influence the similarity measure (S), respectively. Moreover, we assign a higher weight to the blocks, which are closer to the center of the image. This helps reducing the background effects in the classification operation. We expect better results if the important facial parts such as eyes and mouth are detected and higher weights are assigned to them. In the described implementation, we select a LoG window of size 13 × 13 and a σ of value equal to 2.[12] The line drawing-profiles of the facial image used in this implementation are shown in Figure 2.

IIR Filtering and ARMA Modeling

Consider a two-dimensional (random) sequence defined as

$$\hat{I}(m,n) = \sum_k \sum_l a(k,l) I(m-k,n-l)\,, \qquad (k,l) \in W \tag{3}$$

$$I(m,n) = \hat{I}(m,n) + \eta(m,n) \tag{4}$$

Here, the random field *I(m,n)* represents the center value of the image window W, *I*(*m,n*) denotes the corresponding predicted values of the same pixel, η(*m,n*) is the prediction error (or input to an IIR filter), and *a*(*k,l*) is the auto-regressive parameters or IIR filter coefficients. There are four types of the auto-regressive predictor (filter) models, which are strictly causal, causal, semi-causal, and non-causal. The difference between these four types is that each uses a different region (W) for prediction. In our case, we employ the non-causal model, which predicts an unknown pixel value using the whole neighboring pixels as shown in Figure 1. This means that the predicted pixel is in the center of the window being used in the model, which is of size 5 × 5 in our implementation.

	-2	-1	0	1	2
-2	X	X	X	X	X
-1	X	X	X	X	X
0	X	X	**O**	X	X
1	X	X	X	X	X
2	X	X	X	X	X

*Figure 1. A 5 × 5 non-causal prediction window (W) for the center pixel (**O**).*

For the non-causal case, Equation (3) becomes

$$I(m,n) = \sum_{k=-p}^{p} \sum_{l=-q}^{q} a(k,l) I(m-k,n-l)\,, \qquad (k,l) \in W \tag{5}$$

with the corresponding window W (see Figure 1) being

$$W = \{-p \le k \le p, -q \le l \le q, (k,l) \ne (0,0)\}$$

Using the minimum mean square error criterion in our predication case,[14] we get

$$Ra = -[\ 0\ \sigma_n^2\ 0\ ]^T \tag{6}$$

Here R is the corresponding (2p+1)x(2q+1) image autocorrelation matrix, **0** is the zero vector of appropriate dimension, $\boldsymbol{\sigma}_{\mathbf{n}}^2$ is the image variance vector in the specified window, and ***a*** is the unknown filter coefficients' (or feature) vector. For a symmetric window of size p=q, **a** is in the form of

$$a^T = [a(p,p),\ldots,a(0,0),\ldots,a(-p,-p)] \tag{7}$$

The dimension of ***a*** is given by $2(p+1)^2$. For the prediction window of Figure 1, we have p = q = 2. As a result, the size of ***a*** will become 25. However, we use only the first 12 elements due to fact that ***a*** is symmetric with the restriction that a(0,0)=1.

Similarity Measure

A simple means of measuring the discrimination capability of the IIR filtering and the ARMA modeling is carried out by a non-metric similarity measure of[13]

$$S = |\cos \angle(\boldsymbol{a}.\boldsymbol{b})| \tag{8}$$

between two facial patterns whose representations are denoted by the aforementioned vectors **a** and **b** of the form

$$a^T = [a(-p,-p),\ldots,a(0,0),\ldots,a(p,p)]$$

$$b^T = [b(-p,-p),\ldots,b(0,0),\ldots,b(p,p)]$$

The adapted similarity measure (S) is the absolute value of the cosine of the angle between **a** and **b**, which is easy to calculate. It varies within the range of 0 to 1 (i.e., $0 \leq S \leq 1$). As mentioned earlier, if we have for two identical images, then the value of S reaches its maximum limit of 1 since $\mathbf{a}=\mathbf{b}$ for a perfect match. For two completely different pictures S becomes 0 since $\mathbf{a} \perp \mathbf{b}$ for a total mismatch. These limits are, of course, for two extreme cases, and can hardly be achieved due to sensory noise, image acquisition imperfections, and facial variations.

One of the shortcomings of this similarity measure is that it is a non-metric function. It has a maximum when two feature vectors are oriented in the same direction with respect to the origin. This means that it is useful when cluster regions tend to develop along the principal axes. Its effectiveness is governed if there is sufficient separation of cluster regions with respect to each other as well as with respect to the coordinate system origin.[13] This has to be verified in the 12-dimensional feature space. This remains to be a research effort on our part.

Computer Implementation

The face recognition and classification technique presented here is implemented in a PC using the known tools of Matlab-6. A relatively small database of facial images is created to test the effectiveness of the method. In this database, there are nine images of three different persons. The imaging database is shown in Figure 2. For each individual, there are three face images, each of which is for a different facial gesture and mimics of a person. The original images and their line-drawing versions are indexed by numbers, with number 1 (i.e., a1, b1, and c1) corresponding to the normal faces, number 2 (i.e., a2, b2, and c2) representing the happy expressions, and number 3 (i.e., a3, b3, and c3) denoting the sad faces.

Line-drawing version of each facial image in Figure 2 is represented by its IIR coefficients' vector. For any two images with their vectors in the form

$$a^T = [a(-2,-2), a(-1,-2), \ldots, a(2,2)]$$

$$b^T = [b(-2,-2), b(-1,-2), \ldots, b(2,2)]$$

the S measure of Equation (8) is carried out in 12-dimensional feature space for all images in Figure 2 and the measurement results are listed in Table 1.

From Table 1, we can conclude that the method can distinguish different facial pattern classes, containing distinct facial poses of different individuals. The method was able to distinguish the first three images (a1, a2, a3) among themselves. The S values among the images (a1, a2, a3) are equal to or higher than 0.93. The closest S value between the set (a1, a2, a3) and the other two sets (b1, b2, b3) and (c1, c2, c3) is 0.79. We can reach a similar conclusion for the image set (c1, c2, c3), that is, the S values within the set are higher than the S values with any other images from the set (a1, a2, a3) and (b1, b2, b3), respectively. However, for the set (b1, b2, b3) images there appears some discrepancy. We notice that the S value between the images b2 and c1 is equal to 0.88 whereas the S value between the images b2 and b3 is only 0.86. This means that the image c1 appears to be more similar to b2 than b3 does. Clearly, this is not a desirable result. Further improvement to the method is needed in order increase the classification accuracy by means of adopting better feature vectors representing each image.

TABLE 1. The similarity (S) values among the facial images of three individuals in Figure 2.

Image	a1	a2	a3	b1	b2	b3	c1	c2	c3
a1	1.0	.95	.93	.45	.50	.35	.62	.68	.72
a2	.95	1.0	.98	.49	.55	.40	.67	.72	.77
a3	.93	.98	1.0	.51	.57	.42	.69	.74	.79
b1	.45	.49	.51	1.0	.95	.91	.83	.77	.75
b2	.50	.55	.57	.95	1.0	.86	.88	.82	.78
b3	.35	.40	.42	.91	.86	1.0	.74	.68	.63
c1	.62	.67	.69	.83	.88	.74	1.0	.94	.90
c2	.68	.72	.74	.77	.82	.68	.94	1.0	.95
c3	.72	.77	.79	.73	.78	.63	.90	.95	1.0

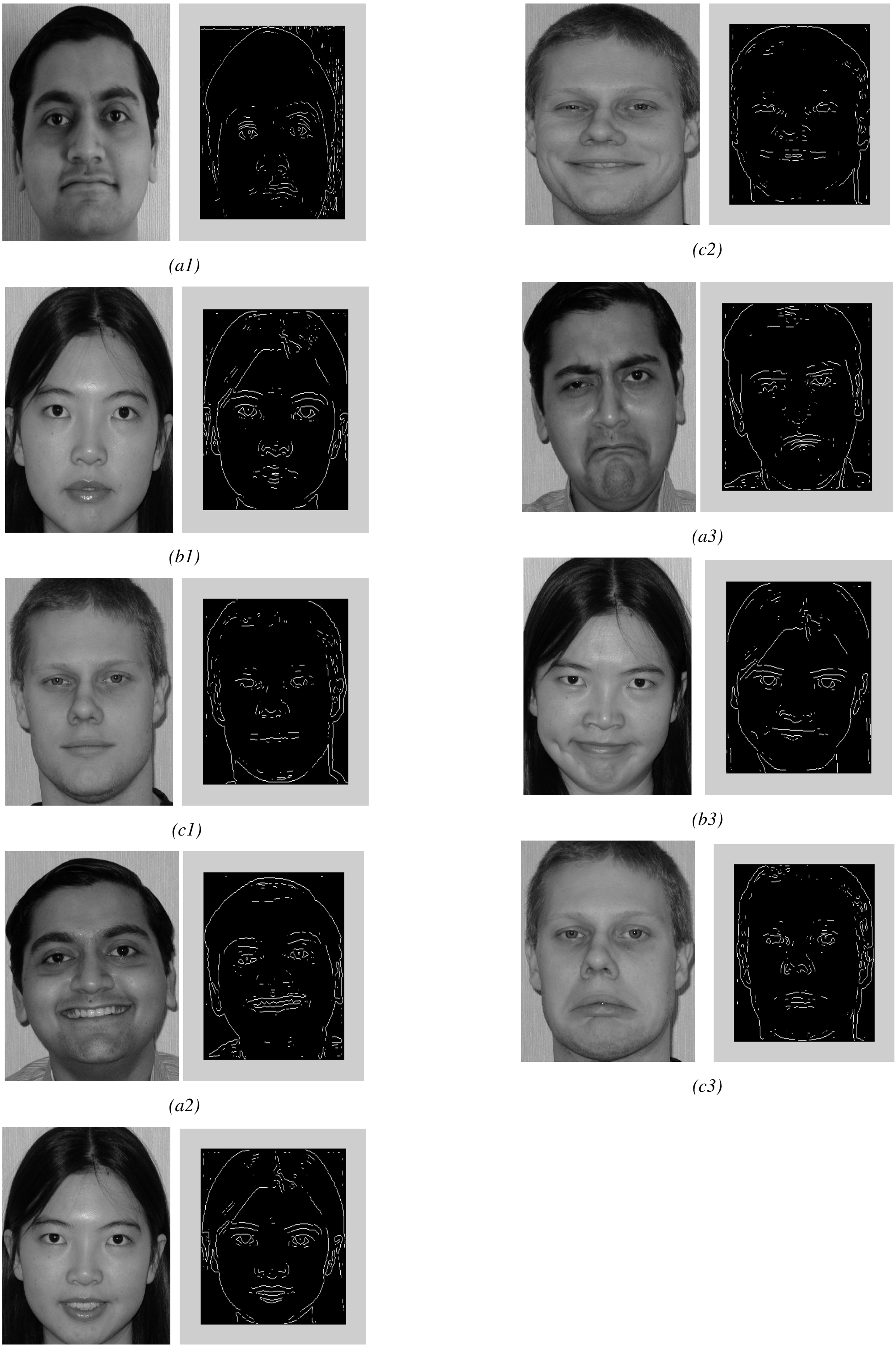
Discussion and Conclusions

In this paper, we have described a face recognition method, which has been developed based on IIR filtering and ARMA modeling using the line-drawings of facial images. The presented algorithm is probabilistic, which makes use of both local and global facial features. In computer simulation, three people have been identified correctly. There appears to be a problem associated with the facial expression recognition. Further study is needed to improve the within class accuracy as well as the between class effectiveness of the method. In order to achieve this, we should improve the feature vectors representing a face image and a facial pose. IIR coefficients- averaging needs to be performed in a more systematic way than the simple ensample averaging method adopted herein.

The selected similarity measure needs further improvement in order to avoid shortcomings of a non-metric function. We have used a small size database of facial images. A larger database is required for a through assessment of the selected ARMA based feature generation. These will be some of our future research effort.

References

1. T. F. Cootes, et al., "Active shape models – Their training and application," *Computer Vision Image Understanding* **61**(1), pp.38-59, 1995.
2. T. F. Cootes, et al., "Use of active shape models for locating structures in medical images," *Image Vision Computing* **12**(6), pp.355-365, 1994.
3. T. F. Cootes and C. J. Taylor, "Locating faces using statistical feature detector," *Int. Conf. on Automatic Face and Gesture Recognition*, Oct. 14-16, 1996, Vermont, USA.
4. C.-L. Huang and Y.-M. Huang, "Facial expression recognition using model-based feature extraction and action parameters classification," *J. of Visual and Image Representation* **8**(3), pp.278-290, 1997.



(a1)

(b1)

(c1)

(a2)

(b2)

(c2)

(a3)

(b3)

(c3)

Figure 2. *Three sets (a1, a2, a3), (b1, b2, b3), and (c1, c2, c3) of face images and their line-drawing representations.*

5. S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 1999.
6. L.-F. Chen, et al., “A new LDA-based face recognition system which can solve the small sample size problem,” *Pattern Recognition* **33**, pp.1713-1726, 2000.
7. L.-F. Chen, et al., “Why recognition in a statistics-based face recognition system should be based on the pure face portion: a probabilistic decision-based proof,” *Pattern Recognition* **34**, pp.1393-1403, 2001.
8. M. Pantic and L. J. M. Rothkrantz, “Automatic analysis of facial expression: The state of the art,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(12), pp.1424-1445, 2000.
9. K.-M. Lam and H. Yan, “An analytic-to-holistic approach for face recognition based on a single frontal view,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(7), pp.673-686, 1998.
10. J. S. Lim, *Two-dimensional Signal and Image Processing*, Prentice Hall, 1990.
11. D. Marr and E. Hildreth, “Theory of edge detection,” *Proc. R. Soc. Lond. B* **207**, pp.187-217, 1980.
12. M. Celenk and S. Bobik, “Edge-suppressed color image indexing and retrieval,” *Proc. of First Int. Conf. on Color in Graphics and Image Processing*, October 1-4, 2000, Saint-Etienne, France, pp.250-255.
13. J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, Addison-Wesley, Reading, 1974.
14. M. Celenk and I. Al-Jarrah, “Multiresolution ARMA modeling of facial color images,” *Image Processing: Algorithms and Systems*, Proceedings of SPIE, Vol.4667, 21-23 January, 2002, San Jose, CA.
15. R. M. Haralick and K. G. Shapiro, *Computer and Robot Vision Volume I*, Addison-Wesley, 1992.