COLOR CHANNEL ENCODING WITH NMF FOR FACE RECOGNITION

Menaka Rajapakse

Institute for Infocomm Research
21 Heng Mui Keng Terrace
Singapore 119613

Jeffrey Tan, Jagath Rajapakse

Nanyang Technological University
School of Computer Engineering
50 Nanyang Avenue, Singapore 639798

ABSTRACT

Colors act as cues for perceiving objects, particularly in complex scenes. Intuitively, color seems to play an important role in recognizing people in scenes. Recent research has evinced that color cues contribute in recognizing faces, especially when shape cues of the images are degraded. Although the input to many of the face recognition systems are color images, during preprocessing, these images are converted to gray scale images for the feature extraction. In this study, we use Non Negative Matrix Factorization (NMF) to recognize color face images. By using NMF, we encode color channels (Red, Green, Blue), thereby, projecting these feature vectors to sparse subspaces. The implemented system is tested on a subset of color images in the AR database for robustness against facial expressions and illumination variations. Furthermore, color face recognition results are compared with the results obtained for the gray scale images of the same data set. Our results show improved accuracy of color image recognition over gray level image recognition when large facial expressions and illumination variations are present.

1. INTRODUCTION

Given color and gray level images, the objects in color images can be recognized easier and faster compared to their counterparts in gray images [1] [2]. Even though colors are usually robust to noise, image degradation, changes in image size, resolution and orientation, much of the work in face recognition has been carried out only using shape cues. The use of color cues in face recognition has been kept to a minimum for the belief that these cues confer little or no information in such tasks. The reasons have been that the color cues do not affect shape from shading processes [3], and, when using high resolution images, strong shape cues are available so that the importance of color cues is not evident [4]. On the contrary to these beliefs, recent research has provided evidence that color cues actually do play a role in recognizing faces, especially when shape cues of the images are degraded [4]. In a given database, degraded images can occur due to several reasons: low resolution, image normalization (small to large), strong illumination etc.

This paper presents reduced subspace color features to recognize color face images. Non-negative Matrix Factorization (NMF) is used to extract low dimensionality feature vectors. In NMF, as the name implies, the non-negativity adds constraints to the matrix factorization, allowing only additions in the synthesis; there are no cancellations or interference of patterns via subtraction or negative feature vector values. This naturally leads to the notion of parts-based representation of images [5] [6]. With the underlying non-negative constraints, NMF is able to learn localized parts based representations. NMF has been applied to recognize gray level images [7]. To extend NMF for color face images, each image in the database is first dissected so that the individual color channels (red, green, blue) act as indexed data vectors representing each image. The large amount of feature data containing in these vectors carry redundant information. As such, the use of a data encoding technique such as NMF is desirable in achieving a reduced data space. Color, the most extensively used feature for color image indexing stands out among other features for its robustness to noise, image degradation, and changes in image resolution and orientation. In this, we exploit this multifacetedness of color feature for face recognition.

This paper is organized as follows. In section 2, we will review the basic concepts of NMF. Face recognition with NMF encoding is addressed in Section 3. Section 4 will present some experiments and results. Section 5 finally will discuss the results of the paper.

2. NON-NEGATIVE MATRIX FACTORIZATION

Given a data matrix $F = \{F_{ij}\}_{nxm}$, non-negative matrix factorization refers to the decomposition of the matrix F into two matrices W and H of size nxr and rxm, respectively, such that

$$F = WH \tag{1}$$

where the elements in W and H are all positive values. From this decomposition, a reduced representation is achieved by choosing the rank, r, such that $r < n$ and $r < m$.

In NMF, no negative values are allowed in matrix factors W and H. The non-negativity constraint is imposed in factorizing the data matrix F by limiting data manipulation only to additions; no subtractions are allowed. The reconstruction of an object is performed only by adding its representative parts collectively. Each column in the matrix W is called a basis image, and a column in the matrix H is called an encoding. An image (column) in F can be reconstructed by linearly combining basis images with the coefficients in an encoding. The encodings influence the activation of pixels in the original matrix via basis images.

Given a data matrix F, Lee and Seung [5] found a technique for factorizing the F to yield matrices W and H as given in Eq(1). Each element in the matrix F can be written as $F_{ij} = \sum_{\rho=1}^{r} W_{i\rho} H_{\rho j}$ where r represents the number of basis images and the number of coefficients in an encoding. The following iterative learning rules are used to find the linear decomposition [5]:

$$H_{\rho j} \leftarrow H_{\rho j} \sum_{i=1}^{n} \left(\frac{W_{i\rho} F_{ij}}{\sum_{k=1}^{r} W_{ik} H_{kj}} \right) \tag{2}$$

$$W_{i\rho} \leftarrow W_{i\rho} \sum_{j=1}^{m} \left(\frac{F_{ij} H_{\rho j}}{\sum_{k=1}^{r} W_{ik} H_{kj}} \right) \tag{3}$$

$$W_{i\rho} \leftarrow \frac{W_{i\rho}}{\sum_{k=1}^{n} W_{k\rho}} \tag{4}$$

The above *unsupervised* multiplicative learning rules are used iteratively to update W and H. The initial values of W and H are fixed randomly. At each iteration, a new value for W or H is evaluated. Each update consists of a multiplication and sums of positive factors. With these iterative updates, the quality of the approximation of the Eq(1) improves monotonically with a guaranteed convergence to a locally optimal matrix factorization [6].

3. FACE RECOGNITION

3.1. Color Image Representation and Training

For color images, a set of 3 data matrices F^l where $l \in \{R, G, B\}$ is constructed such that each color channel, l, of training face images occupies the columns of the F^l matrices.

Let the set of faces be $\Gamma = \{\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_m\}$, then the data matrices, $F^l = [\mathbf{f}^l{}_1 \mathbf{f}^l{}_2 \cdots \mathbf{f}^l{}_m]$, $l \in \{R, G, B\}$. Now learning is done using Eqs (2)-(4) to decompose each matrix F^l into two matrices, H^l and W^l. Let the basis images be $W^l = [\mathbf{w}^l{}_1 \mathbf{w}^l{}_2 \cdots \mathbf{w}^l{}_r]$ and encodings be $H^l = [\mathbf{h}^l{}_1 \mathbf{h}^l{}_2 \cdots \mathbf{h}^l{}_m]$. Each face $\mathbf{f}^l{}_i$ in F can be approximately reconstructed by linearly combining the basis images, and the corresponding encoding coefficients $\mathbf{h}^l{}_i = (h^l{}_{1i} h^l{}_{2i} \cdots h^l{}_{ri})^{\mathrm{T}}$. Hence, a face can be modeled in terms of a linear superposition of basis functions together with color encodings as follows:

$$\mathbf{f}_i^l = W^l \mathbf{h}^l{}_i \tag{5}$$

For each face $\mathbf{f}_i$ in the training set and test set, we calculate the corresponding encoding coefficients for each color channel. The basis images in W^l are generated from the set of training faces, Γ^{train}. The corresponding color encodings, $\mathbf{h}^l{}_i$ of each training face $\mathbf{f}_i$ is given by

$$\mathbf{h}^l{}_i = (W^l)^{\dagger} \mathbf{f}^l{}_i$$

where $W^{\dagger}$ is the computed pseudo-inverse matrix of the basis matrix W. Once trained, the face image set, Γ^{train} is represented by a set of encodings $\{\mathbf{h}^l{}_1, \mathbf{h}^l{}_2, \cdots, \mathbf{h}^l{}_m\}$ with a reduced dimension of rank r.

3.2. Color Image Testing

Given a face image $\mathbf{f}$, we can find a representative color encodings for $\mathbf{f}$ as follows:

$$\mathbf{h}^l = (W^l)^{\dagger} \mathbf{f}^l$$

Cosine angle distance measure is used to calculate the similarity between encodings of a trained image $\mathbf{h}_i \in \Gamma^{train}$ and a test image $\mathbf{h} \in \Gamma^{test}$.

The cosine angle between the two data vectors is taken as the similarity measure, $\mathbf{s}_i$:

$$\mathbf{s}_i = \sum_{l \in R,G,B} \mathbf{s}^l{}_i = \frac{\mathbf{h}^l \cdot \mathbf{h}^l{}_i}{|\mathbf{h}^l||\mathbf{h}^l{}_i|} \tag{6}$$

The similarity measure $\mathbf{s}_i$ determines the matching score between the encodings $\mathbf{h}$ and $\mathbf{h}_i$ corresponding to 2 faces $\mathbf{f}$ and $\mathbf{f}_i$. The optimum matching encoding of a trained image can be given as $\mathbf{h}_{i^*}$ where

$$i^* = \arg \max_i \ \mathbf{s}_i \tag{7}$$

3.3. Gray level Recognition

For gray level images, a single data matrix F, in which each column representing the gray levels of each pixel in each training image is constructed. Using Eqs (2)-(4), the matrix F is decomposed into two matrices, H and W. Lets denote the basis images and encodings as W and H, respectively. A given gray level image can then be modeled as $\mathbf{f}_i = W\mathbf{h}_i$.

The basis images in W are generated from the set of gray level training faces, Γ^{train}. The encodings, $\mathbf{h}_i$ of each training face $\mathbf{f}_i$ is given by $\mathbf{h}_i = W^{\dagger}\mathbf{f}_i$

Similar to the color images, once trained, the face image set, Γ^{train} is represented by a set of encodings $\{\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_m\}$ with an empirical rank r, for data reduction.

Given a gray level face image $\mathbf{f}$, we can find a representative encoding for $\mathbf{f}$ as follows:

$$\mathbf{h} = W^{\dagger}\mathbf{f}$$

A cosine distance measure is used to calculate the similarity between encodings of a trained image $\mathbf{h}_i \in \Gamma^{train}$ and a test image $\mathbf{h} \in \Gamma^{test}$. The cosine of the angle between the two data vectors is taken as the similarity measure:

$$\mathbf{s}_i = \frac{\mathbf{h} \cdot \mathbf{h}_i}{|\mathbf{h}||\mathbf{h}_i|} \tag{8}$$

4. EXPERIMENTS

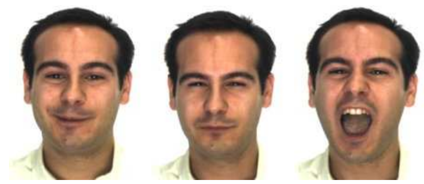
We compare the performance between NMF-encoded color channels of color images and single channel encoded gray level images for their ability to recognize faces. Furthermore, we test the retrieval capability as well as the robustness of the approach against varying facial expressions when color and gray level images are used. Moreover, a test set composed of images with varying lighting conditions is used to evaluate the robustness to withstand varying illumination in query images.

4.1. Database

We investigate the performance of NMF for a subset of AR color face database [8]. This database contains color facial images of 126 individuals (70 males and 56 females). The original dimension of images are 768x576. A total of 13 photos are taken from each individual with each shot taken under different conditions: neutral, smile, anger, scream, left light on, right light on, all side lights on, wearing sun glasses, wearing sun glasses and left light on, wearing sun glasses and right light on, wearing scarf, wearing scarf and left light on, and wearing scarf and right light on. These same shots are taken again after two weeks interval in another session.

For our experiments, only a subset of the original AR database has been extracted and used. The subset extracted contains $m = 100$ face images (50 males and 50 females). Of 13 shots taken in the first photo-taking session, only 7 were used in the experiment, for the reason where these 7 posses the desired variations so that the method can be tested for the robustness and the efficiency of recognition. These variations include changes in facial expressions: sample N(neutral), sample F(expressions), and sample L for illumination conditions in the photo-taking environment. (see Figure 1). These images were cropped and reduced to dimensions of 200x150 and subsequently converted and stored in JPEG format. To evaluate the ability of NMF to extract and analyze low-level image information, the experiments were carried out with non-normalized images.

(a) N : I_1



(b) F : I_2, I_3, I_4

(c) L : I_5, I_6, I_7

Fig. 1. Sample images in the AR database: (a) Sample N - Neutral, (b) Sample F - Facial expressions, (c) Sample L - Illumination variations

4.2. Face Recognition with Cross Validation

The goal of this experiment is to explore the recognition performance and robustness of the implementation for color face images. As for the robustness, it is desirable that the same level of recognition be maintained even when the query image is subjected to certain degrees of variations (facial expression and illumination). With this in mind, 3 training sets, each set exclusively containing the images from each sample N, F and L were assembled. Lets denote each training set as ${\Gamma_N}^{train} = \{{I^i}_1; i = 1, 2, .., m\}$, ${\Gamma_F}^{train} = \{{I^i}_4; i = 1, 2, ..m\}$, and ${\Gamma_L}^{train} = \{{I^i}_7; i = 1, 2, ..m\}$, respectively. For each training set, three different basis matrices and encodings were extracted for the each in color in the RGB scheme as explained in section 3.1. Two experiments were carried out as described below.

Experiment 1: This experiment was carried out to test the robustness of the approach when each training set consists solely of a single type of image variation: neutral, large facial expression or strong illumination. Each case is trained separately using associated training sets ${\Gamma_N}^{train}$, ${\Gamma_F}^{train}$ or ${\Gamma_L}^{train}$. Two test sets were used to test each trained data set for its robustness to facial expressions, and lighting variation. The images used in the training set were excluded in the 2 test sets. For instance, when the training set selected is ${\Gamma_N}^{train} = \{{I^i}_1; i = 1..m\}$, the 2 test sets chosen are: ${\Gamma_F}^{test} = \{{I^i}_2, {I^i}_3, {I^i}_4; i = 1..m\}$, and ${\Gamma_L}^{test} = \{{I^i}_5, {I^i}_6, {I^i}_7; i = 1..m\}$. These test sets check the ability of the approach to withstand facial expression and illumination variations.

Experiment 2: This experiment was used to measure the recognition performance when large variations are introduced to the test images. These variations (illumination and facial expressions) were captured and encoded using the training sets ${\Gamma_N}^{train}$, ${\Gamma_F}^{train}$, and ${\Gamma_L}^{train}$. Using the images from sample F, ${\Gamma_F}^{test} = \{{I^i}_2, {I^i}_3\}$, and sample L, ${\Gamma_L}^{test} = \{{I^i}_5, {I^i}_6\}$ two new query sets were formed and tested for recognition. For each recognition operation, the recognition rate in terms of percentage of correct recognition of faces in each training set was computed using the Eq(7). The resulting performance was calculated by averaging the performance for each training set.

Apart from investigating the performance of the color face recognition based on the variations in the query images, we also compared the recognition results when corresponding gray level images are used. These experiments were carried out with 2 NMF ranks of 50 and 100. The rank value serves to provide dimension reduction in NMF and were chosen arbitrarily. The results compiled for the above experiments are summarized in Table 1.

Test Images	Train Images	Recognition Rate (%)			
		r=50		r=100	
		color	gray	color	gray
I_2,I_3,I_4	I_1	94.5	92.5	94.0	93.0
I_5,I_6,I_7	I_1	79.7	72.2	86.5	85.0
I_1,I_2,I_3	I_4	92.0	90.5	91.5	72.5
I_5,I_6,I_7	I_4	81.0	68.5	92.0	72.5
I_1,I_2,I_3,I_4	I_7	74.0	67.5	79.0	74.5
I_5,I_6	I_7	94.0	93.0	94.5	93.5
I_2,I_3	I_1,I_4,I_7	87.0	83.5	87.0	82.0
I_5,I_6	I_1,I_4,I_7	84.7	78.0	87.0	83.5

Table 1. Recognition Performance for color and gray level images under various test conditions for NMF ranks(r)=50,100.

5. DISCUSSION

We explored the advantages of using color images over gray level images for face recognition by applying parts-based NMF to learn color and gray level face images from a subset of 100 images in the AR database. The images subjected to strong illumination and large facial expression variations were used in the experiments to establish the robustness of recognition when using color images over gray level images. For both color and gray level images, the method withstood well for large facial expression variations when neutral images were trained during the feature extraction. Under the same training conditions, for the images with significant lighting variations, the results achieved for color images were higher compared to the gray level images. But the over all performance of the illumination variation test was insignificant when compared with the results for the large facial expression variations indicating the sensitivity of the approach to lighting variations. When the training set consists of sample images of facial expressions and illumination variations, the test results for the images in sample F dropped for both color and gray level images while the performance for the test images in sample L showed an improvement for both types of images. Overall, for all the experiments, color image recognition results outperformed the results for the gray level images. With multiple training images (experiment 2), the maximum difference between recognition rates achieved for the ranks 50 and 100 is 3% and 5.5% for color and gray scale implementations, respectively. In terms of dimension reduction, the percentage of reduction is 50%, which is a trade-off when storage, searching and retrieval costs are of concern. As summarized in Table 1, the results of each experiment illustrates the consistency of the approach proving the stability of the method in recognizing color images when color channels are encoded with NMF.

6. REFERENCES

[1] J. W. Tanaka and L.M. Presnell, "Color diagnocity in object recognition," *Perception and Psychophysics*, vol. 61, pp. 1140–1153, 1991.

[2] G. E. Isenberg L. H. Wurm and A. Lubeker, "Color improves object recognition in normal and low vision," *Journal of Experimental Psychology:Human Perception and Performance*, vol. 19, pp. 899–911, 1993.

[3] V. Bruce and A. Young, *In the eye of the beholder. The science of face perception*, Oxford University Press, 1998.

[4] A. Yip and P. Sinha, "Role of color in face recognition," *MIT AI Memo*, vol. 035, December, 2001.

[5] D. D. Lee and H. Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[6] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Proceedings NIPS*, pp. 556–562, 2000.

[7] S. S. Li and et al., "Learning spatially localized, parts-based representation," *IEEE CVPR*, 2001.

[8] A. M. Martinezm and R.Benavente, "The ar face database," *CVC Tech. Report*, vol. 24, 1998.