AMHSR ORIGINAL ARTICLE

Improving the quality of images synthesized by discrete cosine transform regression-based method using principle component analysis

Kian Hamedani[1] MSc, Valiallah Saba[1] PhD

[1]Radiation Research Center, Faculty of Paramedicine, AJA University of Medical Sciences, Tehran, Iran.

ABSTRACT

Purpose: Different views of an individuals' image may be required for proper face recognition. Recently, discrete cosine transform (DCT) based method has been used to synthesize virtual views of an image using only one frontal image. In this work the performance of two different algorithms was examined to produce virtual views of one frontal image.

Materials and Methods: Two new methods, based on neural networks and principle component analysis (PCA) were used to make virtual views of an image. The results were compared with those of the DCT-based method. Two distance metrics, i.e. mean square error (MSE) and structural similarity index measure (SSIM), were used to measure and compare image qualities. About 400 data were used to evaluate the performance of the new proposed methods.

Results: The neural networks fail to improve the quality of virtually produced images. However, principle component analysis improved the quality of the synthesized images about 3%.

Conclusion: Principle component analysis is better than both DCT-based and neural network methods for synthesizing virtual views of an image.

Keywords: neural networks; face recognition; principle component analysis; discrete cosine transform; mean square error; stractural simillarity index measurment.

INTRODUCTION

Face recognition was a big challenge in pattern recognition for a long time. A general statement of this problem can be formulated as follows: Given still or video images of a scene, identify person or persons in the scene.[1]

Recognizing faces from different views is a challenging problem and most of the face recognition systems fail to have high recognition rates in these cases. This problem is mainly because of the quality of non-frontal images captured by CCTV cameras that are not good enough due to unsuitable positioning of these cameras.[2] So providing good quality non-frontal images is an important task for such a system. How can this problem be solved when there is only one frontal image of every client? To answer this question researchers have started to synthesize virtual views of every client from their frontal images. Generally,

virtual face synthesizing methods are divided into two methods: 2-dimentional (D) and 3D methods. In 2D methods only 2D information of images is used, but in 3D methods, 3D information of texture and shapes are also used to construct virtual faces.[3]

2D methods are easier to implement but since they cannot cover large changes in view, they have some difficulties. A simple method for synthesizing virtual views is to map frontal view image by geometric transformations like rotation and scaling to non-frontal views.[3] In another attempt to synthesize face in 2D, a method was proposed in which virtual views were synthesized by statistical approximation method.[4] A 2D method was proposed in which 2D images were divided into overlapping blocks.[2] Then discrete cosine transform (DCT) of every block in both frontal and non-frontal images were calculated. Afterwards, the mapping

transformation between every corresponding block was calculated using a multivariate regression.

Poggio and Vetter introduced a computational method for 3D face synthesizing named Linear Classes method.[6] In this method the variations of every class is learned from the variations of other people in the same class and these variations are applied to the new face in order to synthesize virtual images. Briefly, in this method every 3D face is considered to be as the combination of some other faces and these faces are combined together linearly with some coefficients. The rotated faces can be synthesized by calculating these coefficients.[5,6]

Fitting a 2D image into a 3D morphable model was another approach which was proposed.[7] In 3D face model estimation the human head geometry was used as a prior knowledge and then shape and texture coefficients were calculated. The drawback of this method was that it was computationally expensive and it might fail to converge.

In this paper two methods were evaluated to improve the quality of 2D synthesized images.[2] The first method is based on neural networks to find the nonlinear mappings instead of linear regression which is used in DCT-based method. The other method was based on principle component analysis instead of DCT.

MATERIALS AND METHODS
DCT-based face synthesizing using multivariate linear regression

A method was presented for synthesizing virtual non-frontal images using one frontal image.[2] The Facial Recognition Technology (FERET) dataset[8] was used to evaluate this method. For synthesizing the images in this method, the frontal images with 64×64 pixels were divided into 8×8 blocks while these blocks had 50% overlap with the adjacent block. The same process was done for non-frontal images. The 2D DCT of every block in all images was calculated. In order to reduce dimensionality from 64 to 16 pixels, the top left 4×4 DCT coefficients of every block were kept and the other coefficients were set to zero. Then every block's DCT coefficients were rearranged in a row.[2] A non-frontal view was achieved using the following equation:

$$v_{(x,y)}^{\theta} = \left[1 \ (v_{(x,y)}^{frontal})^T \right] W_{(x,y)}^{\theta} \qquad (1)$$

In this equation, *W* is a transformation matrix for a specific view and location. This method has been illustrated in Figure 1.

If *A* is considered as the frontal view and *B* as the non-frontal view then the following equations can be used:
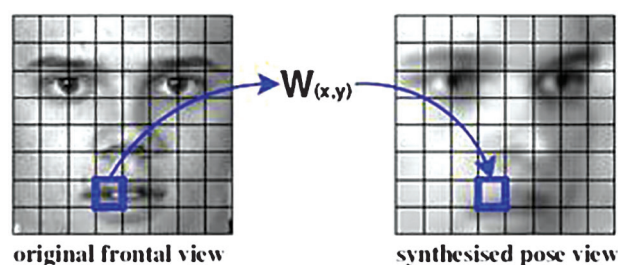

Figure 1. Transforming a frontal view to a specific view using *W* transforming matrix.2.

$$B_{(x,y)} = A_{(x,y)} W_{(x,y)}^{\theta} \qquad (2)$$

$$\begin{bmatrix} b_1^T \\ b_2^T \\ . \\ . \\ . \\ b_N^T \end{bmatrix} = \begin{bmatrix} 1a_1^T \\ 1a_2^T \\ . \\ . \\ . \\ 1a_N^T \end{bmatrix} \begin{bmatrix} w_{1,1} & ... & w_{1,D} \\ w_{2,1} & ... & w_{2,D} \\ . & . & . \\ . & . & . \\ . & . & . \\ w_{D+1,1} & ... & w_{D+1,D} \end{bmatrix} \qquad (3)$$

N is the number of vectors and D is the dimension of vectors.[2] Based on the sum of the least square regression, W can be calculated for every view by the following equation:

$$w_{(x,y)}^{[\theta]} = \left(A_{(x,y)}^T A_{(x,y)} \right)^{-1} A_{(x,y)}^T B_{(x,y)} \qquad (4)$$

By applying the W to the DCT coefficient matrix of every new frontal face, the DCT coefficient matrix of the non-frontal view was achieved. The inverse DCT transform was used to synthesize the blocks. Consequently, by putting these blocks next together the new virtual image corresponding to every view was synthesized.[2]

Improving the quality of synthesized images using nonlinear transform

In the last sub-section the authors used sum of least square regression criterion, which is a linear method for mapping every block in the frontal view image, to the corresponding block in non-frontal image. Using a nonlinear transformation instead of the mentioned linear transformation might increase the quality of the synthesized images. Hence, the neural network was selected to calculate the appropriate mapping between every block in frontal and non-frontal images.

Similar to the last section, the images were divided into 8×8 blocks with 50% overlap. So there were 225 blocks in every image and 225 neural networks had to be trained

for learning the mapping between the corresponding blocks in frontal and non-frontal images. The images of 100 individuals were used for training and 50 others for testing. The specification of the used network was as follows: 17 nodes in the input layer, 64 neurons in the first hidden layer, 128 neurons in the second hidden layer and 17 neurons in the output layer. The applied algorithm for training was scaled conjugate gradient back propagation. This algorithm updates weight and bias values according to the scaled conjugate gradient method.

Using principle component analysis instead of DCT

In signal and image processing, DCT is often applied because of its lossy data compression and strong energy compaction. DCT concentrates most of the signal information in a few low-frequency components.[7,9] However, it does not consider the data distribution in space; this is the drawback of DCT. So here principle component analysis was used instead of DCT for dimension reduction. Principle component analysis reduces the dimension of data considering its distribution in space since the Eigen values represent the distribution of the source data's energy.[10] All the stages in this method were similar to those of the DCT method except that principle component analysis was used instead of the DCT.

But perception was not a good method for approving the image quality improvement. So some similarity measure tools were used to quantitatively evaluate the image quality improvement. The mean square error (MSE) and structural similarity index measure (SSIM), two of the well-known metrics, were used to evaluate and compare the new methods with the DCT-based method.

Using MSE for measuring image quality

Image quality is a characteristic of an image that measures the perceived image degradation. MSE has been widely used during last 20 years for image quality measurement. MSE is defined as:

$$\text{MSE} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (x_{ij} - y_{ij})^2 \qquad (5)$$

In this equation x is the original image and y is the synthesized image. M and N are the width and height of an image.

Using SSIM for measuring image quality

In order to improve traditional methods like MSE, SSIM was proposed. MSE has been proven to be inconsistent with human visual system.[11,12] The values of SSIM method lie between (0 and 1). A higher value of SSIM corresponds to a better image synthesize. The SSIM is applied to all images then the average value of all SSIM results is calculated.

RESULTS

After training the network in the first method, the 225 proper weight matrices were achieved for transforming the blocks in frontal images in order to correspond to non-frontal image blocks. Then these weights were applied to the blocks of frontal test images to synthesize the other views of the image. To evaluate the performance of the neural network method, eight different views of 15°, 25°, 40°, 60°, -15°, -25°, -40 and -60° were synthesized. Some of the synthesized images and the MSE of this method are shown in Figures 2 and 3.

In the second proposed method, principle component analysis was used instead of DCT; Figure 4 shows an instance of a synthesized image using principle component analysis. Eight different views including 15°, 25°, 40°, 60°, -15°, -25, -40 and -60° were synthesized. The results indicated a considerable enhancement in the quality of the synthesized images.

The results of applying MSE to the synthesized images of the both methods are shown in Figure 5. Principle component analysis had the least MSE compared to the other two methods, i.e. neural networks and DCT. The

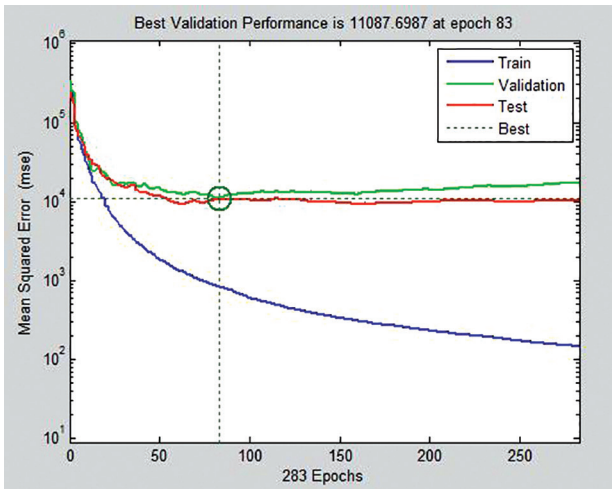**Figure 2.** Synthesized images using nonlinear transform by neural networks.

Figure 3. MSE of the neural networks.

Figure 4. left column: original image, middle column: image synthesized by DCT, right column: image synthesized by principle component analysis.
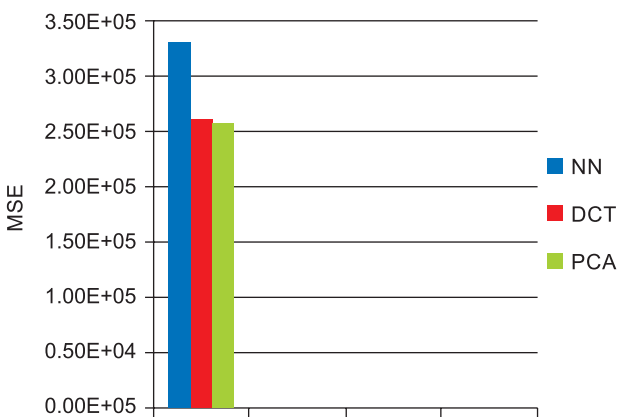
Figure 5. MSE results comparison.
NN, neural networks; PCA, principle component analysis; DCT, discrete cosines transform

results of MSE indicate that using principle component analysis instead of DCT can improve the image quality about 2.9%.
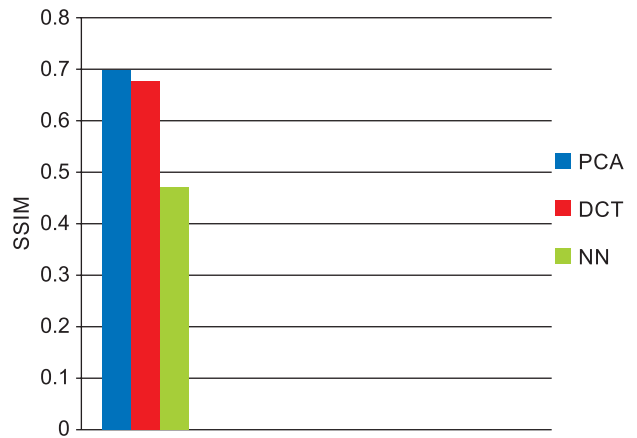
Figure 6. SSIM results comparison.
NN, neural networks; PCA, principle component analysis; DCT, discrete cosines transform

In this study the results of SSIM were as follows: principle component analysis: 0.6980, DCT: 0.6766, NN: 0.4724. Figure 6 shows the SSIM results. These results indicate that using principle component analysis instead of DCT can improve the image quality about 3.1%.

DISCUSSION

In this paper, the performances of two algorithms including neural networks and principle component analysis have been examined to produce the virtual views of a frontal image. The results of using the nonlinear transformation and neural networks were not satisfying. In contrast to the expectations, the quality of the synthesized images did not get better and even got worse. In this neural networks training the error function has a high value and is always higher than the regression error (Figure 3). The reason of this problem could be that neural networks need to have data's which are normalized with their mean and variance for an appropriate training. Since normalizing with mean and variance is an irreversible task, we could not normalize the training data. In the other words, if we normalize the data we cannot bring them back to their real values after the training.

The principle component analysis method was able improve the quality of synthesized images because this algorithm reduced the dimension of data considering its distribution in space. The idea of decomposed eigenface is to make the face synthesize and recognition robust to variations such as lightning and rotation. According to Shakunaga and Kazuma, the performance of eigenfaces is approved in recognizing faces' principle component analysis since least mean square error is an optimal linear scheme. [13] It is possible to transform each original image

into eigenface using principle component analysis.

The Eigen values represent the distribution of the source data's energy in the space.[14] Both SSIM and MSE approve the performance of principle component analysis over two other methods. DCT on the other hand is a lossy data compression[15] method which results in producing images with lower qualities. One other benefit of using principle component analysis over DCT, resulting in images with higher qualities, is that principle component analysis is adaptive. This means that principle component analysis is defined with respect to the dataset because it is necessary to estimate covariance matrix. However, DCT is absolute and is only defined by the input size.

The answers achieved from principle component analysis are unique because of their autonomy from any hypothesis regarding the probability of data distribution. Principle component analysis is also a non-parametric method for data compression in which no prior knowledge is incorporated that makes it possible for principle component analysis to compress information without loss exactly opposite to DCT which is a lossy method of data compression.

One reason that the quality of images synthesized by principle component analysis is better than images synthesized by DCT is because data compression in principle component analysis does not result in loss of data. Wong and colleagues also used DCT for synthesizing virtual faces but because of the mentioned reasons the quality of their synthesized images using principle component analysis was better. Generally, principle component analysis is a locally dependent dimension reduction tool and yields to better results.[2]

The deformations caused from head rotation are nonlinear[15] so we tried to use non-linear mappings using neural networks. Still this did not improve the results because neural networks are irreversible and it is not possible to normalize their data which is an important condition for having good results from neural networks. Reversible nonlinear dimension reduction tools like Laplacian Eigenmaps[16] can be employed for dimension reduction instead of linear dimension reduction tools like principle component analysis and DCT. They can also be used to evaluate the quality of synthesized images using non-linear dimension reduction tools and to compare them with the results achieved from DCT and principle component analysis. This is because it seems there are some features in the head rotation images which lie on a non-linear space and they could only be discovered using nonlinear dimension reduction tools such as Laplacian Eigenmaps.

CONCLUSION

In this paper we suggested two methods for improving the quality of images synthesized by DCT-based method regression. In the first method the neural networks was used to find a non-linear mapping between the DCT coefficients of the frontal and non-frontal blocks. This method could not improve the quality of the synthesized images. In the second method, principle component analysis was used instead of DCT. Based on the MSE and SSIM metrics results, this method improved the quality of the synthesized images about 3%. It is recommended that reversible nonlinear dimension reduction tools such as Laplacian Eigenmaps be employed for dimension reduction instead of linear dimension reduction tools, i.e. principle component analysis and DCT, in future studies.

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Zhao W, Chellappa R, Phillips PJ, et al. Face recognition: A literature survey. *ACM Comput Surv*. 2003;12:399-458.

2. Wong Y, Sanderson C, Lovell BC. *Regression based non-frontal face synthesis for improved identity verification*. Paper presented at: 13th International Conference on Computer Analysis of Images and Patterns; September 2-4, 2009; Münster (North Rhine-Westphalia), Germany.

3. Shan T, Lovell B, Chen S. *Face recognition robust to head pose from one sample image*. Paper presented at: 18th International Conference on Pattern Recognition; August 20-24, 2006; Hong Kong.

4. Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. *IEEE Trans Pattern Anal Mach Intell*. 2001;23:681-5.

5. Beymer D, Poggio T. *Face recognition from one example view*. Paper presented at: 5th International Conference on Computer Vision; June 20-23, 1995; Cambridge, USA.

6. Poggio T, Vetter T. *Recognition and structure from one 2D model view: observations on prototypes, object classes and symmetries*. Cambridge: Artificial Intelligence Laboratory, MIT publications; 1992.

7. Blanz V, Vetter T. Face recognition based on fitting a 3D morphable model. *IEEE Trans Pattern Anal Mach Intell*. 2003;25:1063-74.

8. FERET database. http://www.itl.nist.gov/iad/humanid/feret/. Published June 16, 2004. Accessed October 27, 2012.

9. Ahmed N, Natarajan T, Rao KR. Discrete cosine transform. *IEEE T Comput*. 1974;C22:90-3.

10. Ding C, He X. *K-means clustering via principal component analysis*. Paper presented at: 21st International Conference on Machine Learning; July 4-8, 2004; Banff, Alberta, Canada.

11. Wang Z, Bovik AC, Sheikh HR, et al. Image quality assessment: From error visibility to structural similarity. *IEEE T Image Process*. 2004;13:600-12.

12. Shang X. *Structural similarity based image quality assessment: pooling strategies and applications to image compression and digit recognition* [MS Thesis]. Arlington: University of Texas; 2006.

13. Shakunaga T, Shigenari, K. *Decomposed eigenface for face recognition under various lighting conditions.* Paper presented at: IEEE Computer Society Conference on Computer Vision and Pattern Recognition; December 8-14, 2001; Kauai, HI, USA.

14. Zhang L, Samaras D. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. *IEEE Trans Pattern Anal Mach Intell*. 2006;28:351-63.

15. Hamedani K, Ali Seyyed Salehi S. *Video-based face recognition using manifold learning by neural networks.* Paper presented at: 20th Iranian Conference on Electrical Engineering; May 15-17, 2012; Tehran, Iran.

16. Hong Q, Zhang P, Wang D, et al. An explicit nonlinear mapping for manifold learning. *IEEE T Cyb*. 2013;43:51-63.

Corresponding Author:
Valiallah Saba, PhD
Address: Radiation Research Center, Faculty of Paramedicine, AJA University of Medical Sciences, Fatemi St., Etemadzade Ave., Tehran, Iran.
Postal Code: 1411718541
Tel: +98 21 85952438
Fax: +98 21 88632961
Cell Phone: +98 9124069382
E-mail: vsaba@aut.ac.ir