■ 2004년 정보과학 논문경진대회 수상작

하향식 기계학습의 반복적 오차 역투영에 기반한 고해상도 얼굴 영상의 복원

(Reconstruction of High-Resolution Facial Image Based on Recursive Error Back-Projection of Top-Down Machine Learning)

박 정 선 †　　이 성 환 ††

(Jeong-Seon Park)　　(Seong-Whan Lee)

요 약 본 논문에서는 하향식 기계 학습 및 반복적 오차 역투영을 이용하여 한 장의 저해상도 얼굴 영상으로부터 고해상도 얼굴 영상을 복원하는 방법을 제안한다. 먼저 얼굴 영상을 독립된 형태 기저와 질감 기저의 선형 중첩으로 표현하고, 주어진 저해상도 얼굴 영상을 형태 기저와 질감 기저의 선형 중첩으로 최대한 근사하게 표현할 수 있는 계수를 추정한다. 이 추정된 계수를 고해상도 얼굴 영상의 형태 기저와 질감 기저의 선형 중첩 계수로 사용함으로써 고해상도 얼굴 영상을 복원한다. 또한, 복원된 고해상도 얼굴 영상의 정확도를 개선하기 위하여 학습 기반 오차 역투영 과정을 반복적으로 적용한다.

다양한 실험을 통하여, 제안된 방법이 저해상도 얼굴 영상으로부터 고해상도 얼굴 영상을 효과적으로 복원함을 입증하였으며, 이 방법을 사용하여 원거리 감시 시스템에서 획득된 저해상도 얼굴 영상을 고해상도 얼굴 영상으로 합성함으로써, 얼굴 인식 시스템의 성능을 높일 수 있음을 확인하였다.

키워드 : 저해상도 영상, 하향식 기계 학습, 영상 복원, 얼굴 인식, 오차 역투영

Abstract This paper proposes a new reconstruction method of high-resolution facial image from a low-resolution facial image based on top-down machine learning and recursive error back-projection. A face is represented by a linear combination of prototypes of shape and that of texture. With the shape and texture information of each pixel in a given low-resolution facial image, we can estimate optimal coefficients for a linear combination of prototypes of shape and those that of texture by solving least square minimizations. Then high-resolution facial image can be obtained by using the optimal coefficients for linear combination of the high-resolution prototypes. In addition, a recursive error back-projection procedure is applied to improve the reconstruction accuracy of high-resolution facial image.

The encouraging results of the proposed method show that our method can be used to improve the performance of the face recognition by applying our method to reconstruct high-resolution facial images from low-resolution images captured at a distance.

Key words : Low-resolution facial image, top-down machine learning, image reconstruction, face recognition, error back-projection

1. Introduction

There is a growing interest in surveillance system at a distance for security areas such as international airports, borders, sports grounds, and safety areas. And various researches on face recognition have been carried out for a long time. But there still exist a number of difficult problems such as

· This research was supported by the 2005 Seoul R&BD Program. We would like to thank the Max-Planck Institute for providing the MPI Face Database.

† 정 회 원 : 전남대학교 멀티미디어학과 교수
jpark@chonnam.ac.kr

†† 종신회원 : 고려대학교 컴퓨터학과 교수
swlee@image.korea.ac.kr

논문접수 : 2004년 7월 1일
심사완료 : 2006년 11월 22일

estimating facial pose, facial expression variations, resolving object occlusion, changes of lighting conditions, and in particular, the low-resolution images captured at a distance.

Handling low-resolution images is one of the most difficult and commonly occurring problems in various image processing applications, such as scientific, medical, astronomical, or weather image analysis, image archiving, retrieval and transmission as well as video surveillance or monitoring [1].

Numerous methods have been reported in the area of estimating or reconstructing high-resolution images from a series of low-resolution images or a single-frame low-resolution image. Super-resolution is a typical example of techniques reconstructing a high-resolution image from a series of low-resolution images [2-4], while interpolation produces a large image from only one low-resolution image.

In this paper, we concerned with building a high-resolution facial image from only one low-resolution facial image. Our task is distinguished from previous works that built high-resolution images mainly from scientific images or image sequence of video data.

The proposed method is a top-down, object-class-specific and model-based approach. It is highly tolerant to sensor noise, incompleteness of input images and occlusion by other objects [5]. The top-down approaches for interpreting images of various objects are now attracting considerable interest among many researchers [6,7]. The motivation for top-down machine learning lies on its potential of deriving high-level knowledge from a set of prototypical components.

This paper proposes a novel method for reconstructing high-resolution facial image from only one low-resolution image based on top-down machine learning. The 2D morphable face model [7] is used in top-down machine learning, and a mathematical procedure for solving least square minimization(LSM) is applied to the model. More over, a recursive error back-projection procedure is applied to compensate the residual errors of the reconstructed high-resolution results.

This paper is organized as follows. In section 2, we explain the proposed reconstruction method based on top-down machine learning, the problem of high-resolution reconstruction and solution to solve the least square minimization of reconstruction error. Then, in the next section, we describe the proposed recursive error back-projection procedure which is composed of estimation of high-resolution data, simulation of low-resolution data, and compensation of residual errors. In section 4, experimental results with low-resolution facial images are provided along with an analysis of these results. Finally, we make our conclusions and provide a discussion of future works in section 5.

2. Reconstruction of High-resolution Facial Image Using Top-down Machine Learning

Suppose that sufficiently large amount of facial images are available for off-line training, then we can represent any input face by a linear combination of facial prototypes [9]. Moreover, if we have a pair of low-resolution facial image and its corresponding high-resolution image for the same person, we can obtain an approximation to the deformation required for the given low-resolution facial image by using the coefficients of examples. Then we can obtain high-resolution facial image by applying the estimated coefficients to the corresponding high-resolution example faces as shown in Figure 1.

2.1 Extended 2D morphable face model

The proposed method is based on the morphable face model introduced by Poggio et al. [8] and developed further by Vetter et al. [9,10]. Assuming that the pixel-wise correspondence between facial images has already been established [10], the 2D shape of a face is coded as the displacement field from a reference image. The texture is coded as the intensity map of the image which results from mapping the face onto the reference face as shown in Figure 2(a).

Then, the shape of a facial image is represented by a vector $S=(d_1^x, d_1^y, \cdots, d_k^x, d_k^y, \cdots, d_N^x, d_N^y)^T \in R^{2N}$, where N is the number of pixels in image, (d_k^x, d_k^y) the x, y displacement of a pixel that corresponds to a pixel x_k in the reference face and can be denoted by $S(x_k)$. And the shape normalized texture is repre-

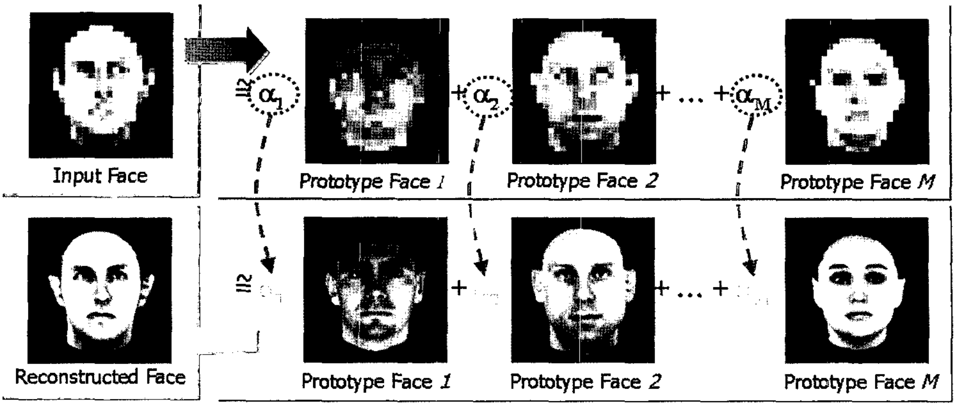


Figure 1 Basic idea of the proposed reconstruction method using top-down learning

sented as a vector $T=(i_1,\cdots,i_k,\cdots,i_N)^T\in R^N$, where i_k is the intensity or color of a pixel that corresponds to a pixel x_k among N pixels in the reference face and can be denoted by $T(x_k)$.

Next, we transform the orthogonal coordinate system by PCA(Principal Component Analysis) into a system defined by eigenvectors s_p and t_p of the covariance matrices C_S and C_T on the set of training faces. C_S and C_T are computed over the differences of the shape and texture, $\tilde{S}=S-\bar{S}$ and $\tilde{T}=T-\bar{T}$. Where $\bar{S}$ and $\bar{T}$ represent the mean of shape and that of texture, respectively.

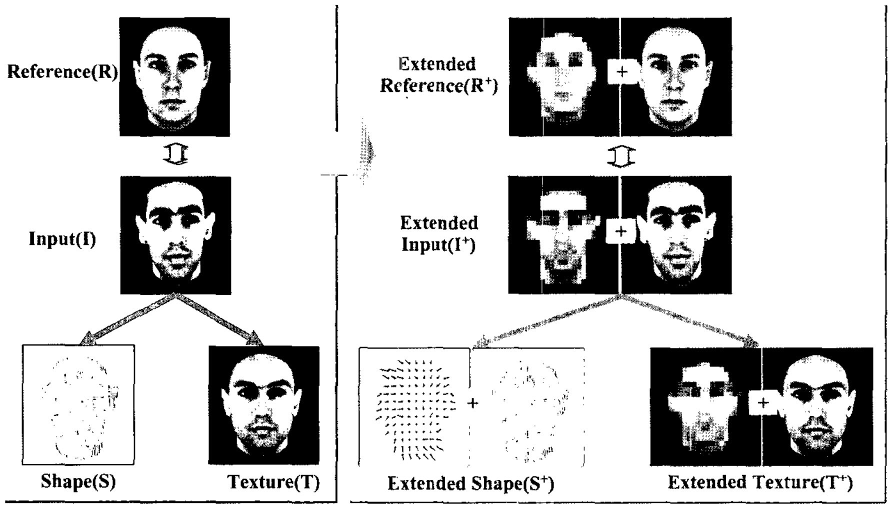
Then, a facial image can be represented by

$$S=\bar{S}+\sum_{p=1}^{M}\alpha_p s_p,\quad T=\bar{T}+\sum_{p=1}^{M}\beta_p t_p \tag{1}$$

where $\alpha,\beta\in R^M$.

Our goal is to reconstruct a high-resolution face from only one low-resolution image based on top-down machine learning. We extended the previous 2D morphable face model by the combination of low-resolution face and high-resolution one as shown in Figure 2(b).

Let us define $S^+=(d_1^x,d_1^y,\cdots,d_L^x,d_L^y,d_{L+1}^x,d_{L+1}^y,\cdots,d_{L+H}^x,d_{L+H}^y)^T$ to be a new shape vector by simply concatenating a low-resolution shape vector and a high-resolution shape vector, where L is the number of pixels in low-resolution image and H is



(a) 2D morphable face model (b) extended 2D morphable face model

Figure 2 Comparisons of existing 2D morphable face model and the our extended 2D morphable face model the number of pixels in high-resolution image. Similarly, let us define $T^{+}=(i_1,\cdots,i_L,i_{L+1},\cdots,i_{L+H})^T$ to be a new texture vector. Then, by applying PCA to both shape S^{+} and texture T^{+}, the face image in Eq.(1) can be extended as

$$S^{+}=\overline{S^{+}}+\sum_{p=1}^{M}\alpha_p s_p^{+}, \qquad T^{+}=\overline{T^{+}}+\sum_{p=1}^{M}\beta_p t_p^{+}. \tag{2}$$

2.2 Reconstruction of high-resolution face

Our goal is to find an optimal parameter set ($\hat{\alpha}$) which best estimates the high-resolution image from a given low-resolution image.

Before explaining the proposed reconstruction procedure, we define two types of warping processes, forward and backward warping as shown in Figure 3. Forward warping warps a texture expressed in reference shape onto each input face by using its shape information. This process results in an original facial image. Backward warping warps an input face onto the reference face by using its shape information. This process results in a texture image expressed in the reference shape. The mathematical definition and more details about the forward and backward warping can be found in reference [9].

The reconstruction procedure of high-resolution facial image consists of 4 steps, starting from a low-resolution facial image to a high-resolution face as shown in Figure 4. Here, the displacements of the pixels in an input low-resolution face which correspond to those in the reference face are known.

Step 1. Obtain the texture of a low-resolution facial image by backward warping.

Step 2. **(a)** Estimate a high-resolution shape from the given low-resolution shape.
(b) Improve the estimated high-resolution shape by our recursive error back-projection procedure.

Step 3. **(a)** Estimate a high-resolution texture from the low-resolution texture obtained at Step 1.
(b) Improve the estimated high-resolution texture by our recursive error back-projection procedure.

Step 4. Synthesize a high-resolution facial image by forward warping the improved texture with the improved shape

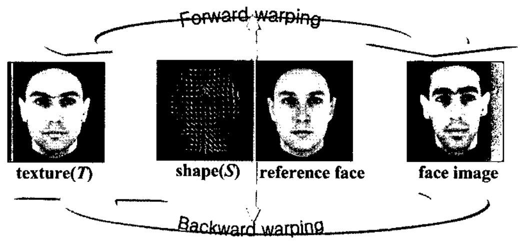


Figure 3 Examples of forward warping and backward warping

Step 1 and Step 4 are explained from the previous studies of morphable face models [6,9]. Step 2(a) and Step 3(a) are carried out by similar mathematical procedure except that the shape of a pixel is 2D vector and the texture is 1D(or 3D for RGB color image) vector. Therefore, we will describe only the Step 2(a) of estimating a high-resolution shape vector from a low-resolution shape vector.

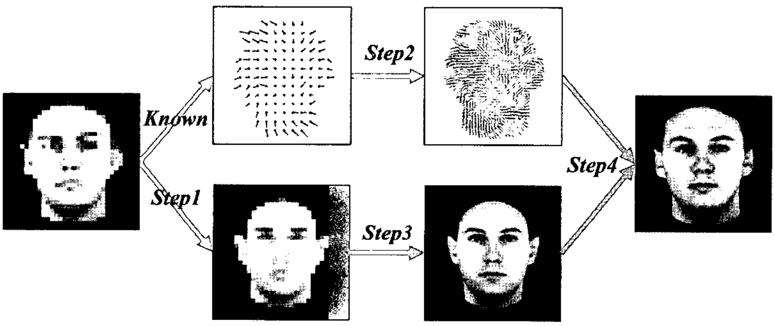


Figure 4 Reconstruction procedure from a low-resolution facial image to a high resolution one

2.3 Problem definition for high-resolution shape estimation

Since there is a shape vector from only low-resolution facial image, we need an approximation to the deformation required for the low-resolution shape by using coefficients of the bases(prototype shapes) as shown in Figure 1. The goal is to find an optimal set α_p that satisfies

$$\widetilde{S^+}(x_j)=\sum_{p=1}^{M}\alpha_p s_p^+(x_j), \quad j=1,2,\cdots,L, \tag{3}$$

where x_j is a pixel in the low-resolution facial image, L the number of pixels in low-resolution image, and M the number of bases.

Generally there may not exist a set of α that perfectly fits the $\widetilde{S^+}$. So, the problem is to choose $\hat{\alpha}$ so as to minimize the error. For this, we define an error function, $E(\alpha)$, the sum of squared errors which measures the difference between the known displacements of pixels in the low-resolution input image and its represented ones.

$$E(\alpha)=\sum_{j=1}^{L}\left(\widetilde{S^+}(x_j)-\sum_{p=1}^{M}\alpha_p s_p^+(x_j)\right)^2 \tag{4}$$

where $x_1,\cdots,x_L$ are pixels in the low-resolution shape vector.

Then the problem of reconstruction is formulated as finding $\hat{\alpha}$ which minimizes the following error function :

$$\hat{\alpha}=\arg\min_{\alpha} E(\alpha) \tag{5}$$

2.4 Solution by least square minimization

The solution to Eq.(4)-(5) is nothing more than a least square solution. Eq.(3) is equivalent to the following equation

$$\begin{pmatrix} s_1^+(x_1) & \cdots & s_M^+(x_1) \\ \vdots & \ddots & \vdots \\ s_1^+(x_L) & \cdots & s_M^+(x_L) \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_M \end{pmatrix}=\begin{pmatrix} \widetilde{S^+}(x_1) \\ \vdots \\ \widetilde{S^+}(x_L) \end{pmatrix}. \tag{6}$$

This can be rewritten as :

$$S^+\alpha=\widetilde{S^+}, \tag{7}$$

where

$$S^+=\begin{pmatrix} s_1^+(x_1) & \cdots & s_M^+(x_1) \\ \vdots & \ddots & \vdots \\ s_1^+(x_L) & \cdots & s_M^+(x_L) \end{pmatrix}, \tag{8}$$

$$\alpha=(\alpha_1,\cdots,\alpha_M)^T,$$

$$\widetilde{S^+}=(\widetilde{S^+}(x_1),\cdots,\widetilde{S^+}(x_L))^T.$$

The least square solution to an inconsistent $S^+\alpha=\widetilde{S^+}$ of L equation in M unknowns satisfies $S^{+T}S^+\alpha=S^{+T}\widetilde{S^+}$. If the columns of S^+ are linearly independent, then $S^{+T}S^+$ is non-singular and

$$\alpha^*=(S^{+T}S^+)^{-1}S^{+T}\widetilde{S^+}. \tag{9}$$

The projection of $\widetilde{S^+}$ onto the column space is therefore $\hat{\widetilde{S^+}}=S^+\alpha^*$. By using Eq.(9), we obtain a high-resolution shape vector

$$S(x_{L+j})=\overline{S^+}(x_{L+j})+\sum_{p=1}^{M}\alpha_p^* s_p^+(x_{L+j}), \quad j=1,2,\cdots,H \tag{10}$$

where $x_{L+1},x_{L+2},\cdots,x_{L+H}$ are pixels in the high-resolution shape vector, L and H are the number of pixels in the low-resolution facial image and that of high-resolution one, respectively.

By using Eq.(10), we can get the correspondence of all pixels. Previously we made the assumption that the columns of S^+ are linearly independent. Otherwise, Eq.(9) may not be satisfied. If S^+ has dependent columns, the solution α^* will not be unique. The optimal solution in this case can be solved by pseudoinverse of S^+ [5]. But, for our purpose of effectively reconstructing a high-resolution facial image from a low-resolution one, this is unlikely to happen.

3. Proposed Recursive Error Back-Projection Method

In order to reconstruct a high-resolution facial image from only one low-resolution image, we used an example-based learning or top-down approach. Also to improve the accuracy of the reconstruction, we applied a recursive error back-projection procedure to the results of step 2(a) and step 3(a), respectively. Recursive error back-projection or error compensation has been used to various applications such as super-resolutions [4].

In this section, we explained the procedure of our recursive error back-projection method for improving the resolution of facial images. The flowchart of the procedure we have designed for recursive error back-projection is outlined in Figure 5. The

Table 1 Notations for recursive error back-projection procedure

Notation	Definition
L^I	Input low-resolution information (shape or texture)
t	Iteration index, $t = 1, 2, ..., T$
H_t^R	Reconstructed high-resolution data at iteration t
L_t^R	Low-resolution data simulated by down-sampling the reconstructed high-resolution data at iteration t
D_t^L	Reconstruction error by measuring Euclidean distance between input and simulated low-resolution data at iteration t
T_1	A threshold value to determine whether the reconstruction is accurate or not
T_2	A threshold value to determine whether the iteration is convergent or not
L_t^E	Reconstruction error of low-resolution data by pixel-wise difference between input and simulated low-resolution data at iteration t
H_t^E	Reconstructed high-resolution error of low-resolution error at iteration t
ω_t	Weight for error compensation at iteration t

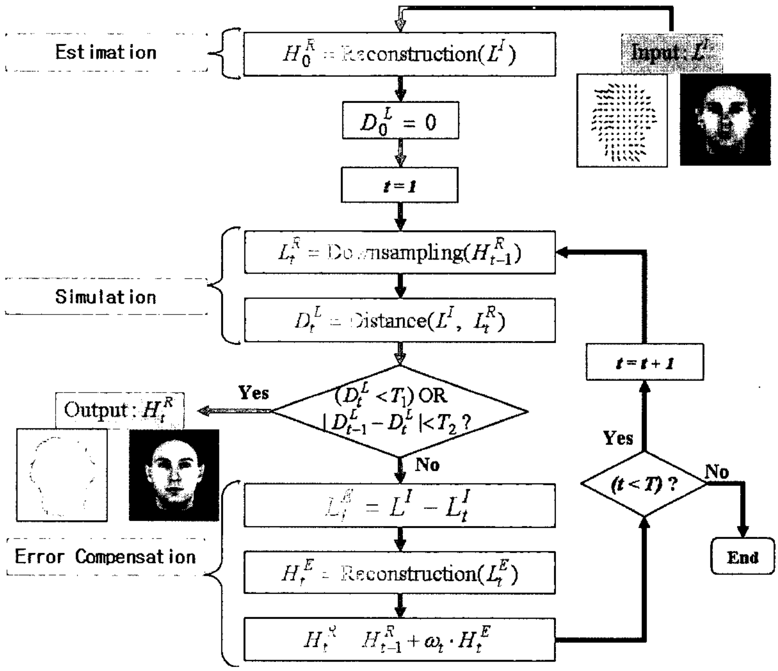


Figure 5 Flowchart of the recursive error back-projection procedure

notations showed in this figure are defined at Table 1.

First, as an initial procedure, we estimate the high-resolution data (H_0^R) from input low-resolution data (L^I) by using our solution of least square minimization described in section 2.4.

Second, as a simulation procedure, in order to verify the accuracy of our reconstruction method, simulate the low-resolution data (L_1^R) from the estimated high-resolution data by down-sampling it, then measure the distance (D_1^L) between input low-resolution and simulated one by simple Euclidean distance measure. We assume that if the reconstruction is successful, the reconstruction error(distance) will be very small. From this assumption, we determine whether the reconstruction is accurate or not by comparing the current reconstruction error and one threshold value (T_1) and whether the iteration is convergent or not by comparing amount of previous distance and current distance, and another threshold value (T_2). If one or two comparisons are satisfied, the current result of reconstruction is considered as the final output high-resolution data, otherwise following error back-projection procedure is recursively applied.

Third, as an error compensation procedure, we create the low-resolution error data between input low-resolution and simulated one by simple difference operation, estimate the high-resolution error data by our reconstruction of low-resolution error data, then compensate previously estimated high-resolution data by adding currently estimated error to it. In this procedure, we use weight value(ω_t) which is smaller than 1, in order to prevent divergence of iterations. The weight can be varied according to the current distance, that is, if the distance is large then the weight is also large. We recursively perform the same procedures until an accurate estimation is achieved or iterations are convergent.

By using these iterative procedures, we tried to improve the result of high-resolution estimation by recursively compensating error.

4. Experimental Results and Analysis

4.1 Face database

For testing the proposed method, we used 200 images of Caucasian faces that were rendered from a database of 3D head models recorded by a laser scanner(Cyberware^{TM}) [9,10]. The original images were color image set of 256×256 pixels. They were converted to 8-bit gray level and resized to 16×16 and 32×32 for low-resolution facial images by Bicubic interpolation method. PCA was applied to a random subset of 100 facial images for constructing bases of the defined face model. The remaining 100 images were used for testing the proposed method.

Next, we use a hierarchical, gradient-based optical flow algorithm to obtain a pixel-wise correspondence [9]. The correspondence is defined between a reference image and every image in the database. It is estimated by the local translation between corresponding gray level intensity patches.

4.2 Results of resolution enhancement

As mentioned before, 2D-shape and texture of facial images are treated separately. Therefore, a high-resolution facial image is synthesized by combining both reconstructed shape and reconstructed texture.

Figure 6 shows the examples of the high-resolution facial image synthesized from two kinds of low-resolution images. In the Figure, (a) shows the input low-resolution images, (b) to (e) the synthesized high-resolution images using Bilinear interpolation, Bicubic interpolation, proposed method 1(using only example-based reconstruction) and proposed method 2(enhancement of (d) by recursive error back-projection), respectively. (f) is the original high-resolution facial images.

As shown in Figure 6(d) and (e), the reconstructed images by the proposed top-down machine learning method are more similar to the original images and clearer than others. More over, the effect of improving the results by recursive error back- projection can be notified in the synthesis results of 16×16.

As shown in Figure 6, classifying the 16×16 input low-resolution faces are almost impossible, even with the use of Bilinear or Bicubic interpolations. On the other hand, reconstructed facial images by the example-based learning methods, especially the reconstructed images by the proposed method are more similar to the original faces than others. Also, similar but better results were obtained from 32×32 low-resolution facial images.

Figure 7 shows the comparison of mean reconstruction errors in shape, texture and facial image from the original high-resolution data. The horizontal axes of the figure represent the input low-resolution, two interpolation methods, the proposed reconstruction method and the improved results by the proposed recursive method. Vertical axes represent the mean displacement error per pixel about shape vectors and the mean intensity error per pixel about texture and image vector, respectively. That is, Err_S_x and Err_S_y of (a) are the x- and y- directional mean displacement errors for the shape vector, respectively. And Err_T and Err_I of (b) imply the mean intensity errors for the texture vector and for the image vector, respectively.

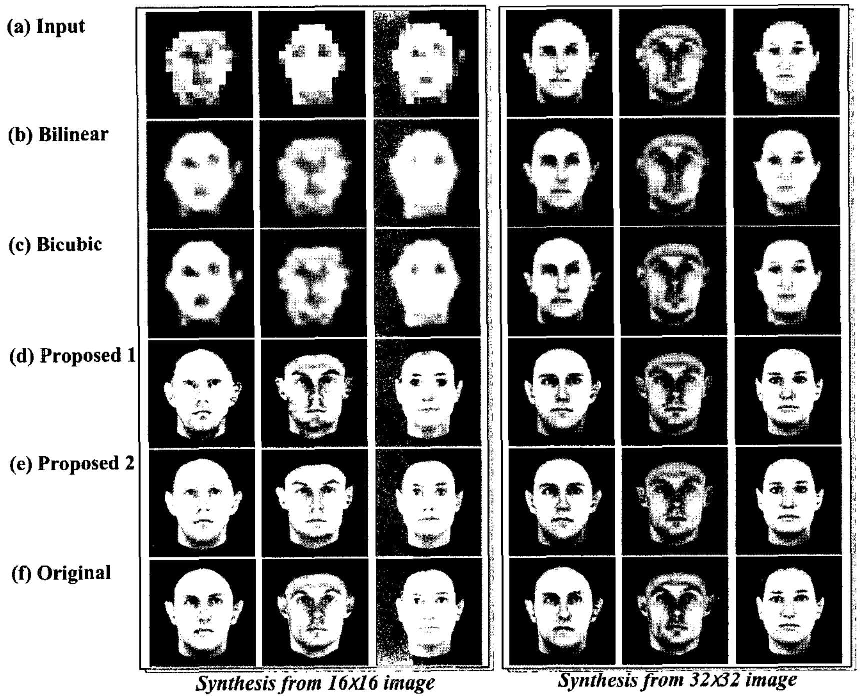


Figure 6 Examples of reconstructed high-resolution face from low-resolution face

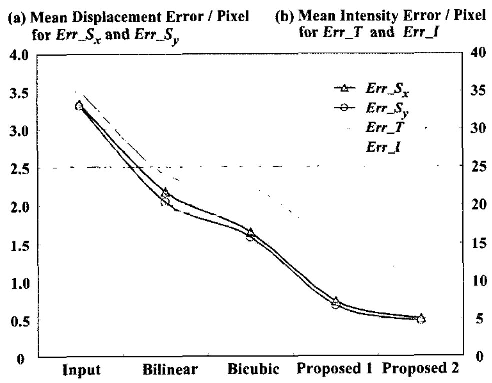


Figure 7 Comparison of mean reconstruction errors

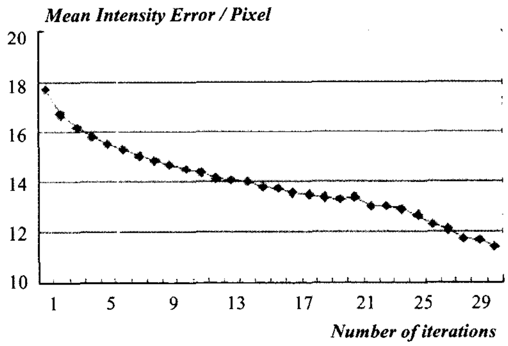


Figure 8 Changes of mean intensity error per pixel during recursive error back-projection

Also, we measure the distances between the original high-resolution facial images and their reconstructed images while increasing the number of iterations. Figure 8 shows the change in the mean intensity error per pixel between the original high-resolution image and the recursively updated images. As the trend of gradually decreasing distance shows, we can conclude that the similarity between the original high-resolution facial images and the compensated one increased as the number of iterations increased.

From the encouraging results of the proposed method as shown in Figures 6-8, we can expect that our reconstruction method can be used to improve the performance of the face recognition by reconstructing high-resolution facial images from single-frame low-resolution facial images captured at visual surveillance systems.

5. Conclusions and Further Research

In this paper, we proposed an efficient method of reconstructing high-resolution facial image based on top-down machine (or example-based) learning and recursive error back-projection. The proposed method consists of the following steps : computing linear coefficients minimizing the error or difference between the given shape/texture and the linear combination of the shape/texture prototypes in the low-resolution image, and applying the coefficient estimates to the shape and texture prototypes in the high-resolution facial image, respectively.

The experimental results are very natural and plausible like original high-resolution facial images, when displacements among the pixels in an input face which correspond to those in the reference face, are known. It is a challenge for researchers to obtain the correspondence between the reference face and a given facial image under low-resolution vision tasks.

Reference

[1] B. Tom and A. K. Katsaggelos, *"Resolution Enhancement Monochrome and Color Video Using Motion Compensation," IEEE Trans. on Image Processing*, Vol.10, No.2, pp. 278-287, February 2001.
[2] S. Baker and T. Kanade, *"Limit on Super-Resolution and How to Break Them," IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24 No. 9, pp. 1167-1183, September 2002.
[3] M. G. Kang and S. Chaudhuri, *"Super-Resolution Image Reconstruction," IEEE Signal Processing Magazine*, Vol. 23, No. 3, pp. 19-36, May 2003.
[4] F. Dekeyser, P. Perez, and P. Bouthemy, *"Restoration of Noisy, Blurred, Undersampled Image Sequences Using Parametric Motion Model," Proc. of the International Symposium on Image/Video Communications over Fixed and Mobile Networks, ISIVC 2000*, Rabat, Morocco, April 2000.
[5] P. S. Windyga, *"Fast impulsive noise removal," IEEE Trans. on Image Processing*, Vol. 10, No. 1, pp. 173-178, 2001.
[6] M. J. Jones, P. Sinha, T. Vetter, and T. Poggio, *"Top-down Learning of Low-level Vision Tasks," Current Biology[brief communication]*, Vol. 7, pp. 991-994, 1997.
[7] J.-S. Park and S.-W. Lee, *"Resolution Enhancement of Facial Image Based on Top-down Learning,"* Proc. of ACM SIGMM 2003 Workshop on Video Surveillance, November 2003, pp. 59-64.
[8] D. Beymer, A. Shashua, and T. Poggio, *"Example-Based Image Analysis and Synthesis," AI Memo 1431/CBCL Paper 80,* Massachusetts Institute of Technology, Cambridge, MA, November 1993.
[9] T. Vetter and N. E. Troje, *"Separation of Texture and Shape in Images of Faces for Image Coding and Synthesis," Journal of the Optical Society of America A.* Vol. 14, No. 9, pp. 2152-2161, 1997.
[10] V. Blanz and T. Vetter, *"Morphable Model for the Synthesis of 3D Faces,"* Proc. of SIGGRAPH'99, Los Angeles, pp. 187-194, 1999.

박 정 선
1992년 충북대학교 전산학(학사). 1994년 충북대학교 전산학(석사). 2005년 고려대학교 전산학(박사). 1994년~1999년 (주) 현대정보기술 선임연구원. 2005년~현재 전남대학교 멀티미디어학과 전임강사. 관심분야는 컴퓨터비전, 얼굴인식, 영상처리

이 성 환
1984년 서울대학교 계산통계학과(학사) 1986년 한국과학기술원 전산학과(석사) 1989년 한국과학기술원 전산학과(박사) 1989년~1995년 충북대학교컴퓨터과학과 조교수. 1995년~현재 고려대학교 컴퓨터학과 정교수. 1997년~현재 고려대학교 인공시각연구센터 소장. 2001년~2002년 MIT AI Lab. 객원교수. 관심분야는 컴퓨터비전, 패턴인식, 영상처리