Reconstruction of High-Resolution Facial Image Based on A Recursive Error Back-Projection

Joeng-Seon Park⁰ and Seong-Whan Lee
Center for Artificial Vision Research, Korea University
{jspark⁰, swlee}@image.korea.ac.kr

Abstract

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

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 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 single low-resolution image. Super-resolution is a typical example of techniques reconstructing a high-resolution image from a series of low-resolution images[2-3], where as interpolation produces a large image from only one low-resolution image.

In this paper, we are 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 video data. The proposed method is a top-down, object-class-specific and model-based approach. The top-down approach to interpreting images of variable objects are now attracting considerable interest among many researchers[4-6]. 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 of reconstructing high-resolution facial image from only one low-resolution one based on top-down machine learning. The 2D morphable face model[6] 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 improve the accuracy of high-resolution reconstruction.

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 a number of facial prototypes[5]. 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.

The proposed method is based on the morphable face model introduced by Poggio et al[4]. and developed further by Vetter et al.[6]. Assuming that the pixel-wise correspondence between facial images has already been established[6], 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.

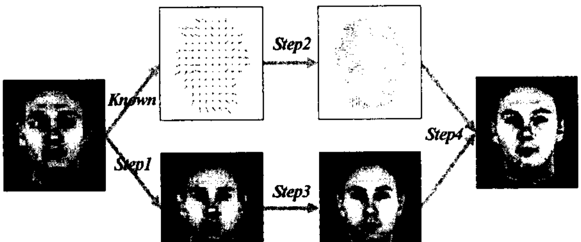

Figure 1. Reconstruction procedure from a low-resolution facial image to a high resolution one

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 1. Here the displacement of the pixels in an input low-resolution face which correspond to those in the reference face is known.

Step 1 and Step 4 are explained from the previous studies of morphable face models in many studies[6-7]. Step 2 and step 3 are carried out by similar mathematical procedure except that the shape about a pixel is 2D vector and the texture is 1D(or 3D for RGB color image) vector.

Let us define $S^+ = (d_1^x, d_1^y, \cdots, d^x_{L+H}, d^y_{L+H})^T$ to be a new shape information by simply concatenating low-resolution shape and high-resolution shape, where L is the number of pixels in low-resolution image and H is 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 information. Then, by applying PCA to both the shape S^+ and texture T^+, the face image in Eq.(1) can be expressed as

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

where $\alpha, \beta \in R^{M-1}$.

Then, our goal is to find an optimal parameter set ($\hat{\alpha}, \hat{\beta}$) which best estimates the high-resolution shape and texture from a given low-resolution image. The mathematical definition and more details about the procedure for finding an optimal parameter set can be found in reference [5].

3. Recursive Error Back-Projection Procedure

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 high-resolution reconstruction, we applied a recursive error back-projection procedure which has been used to various applications such as super-resolution[4]. In this section, we explained the procedure of the our recursive error back-projection for improving the resolution of reconstructed facial images. The flowchart of the procedure we have designed for recursive error back-projection is outlined in figure 2. The notations showed in this figure are defined as follows.

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 reconstructed one at iteration t
D_t^L	: Reconstruction error measured by 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

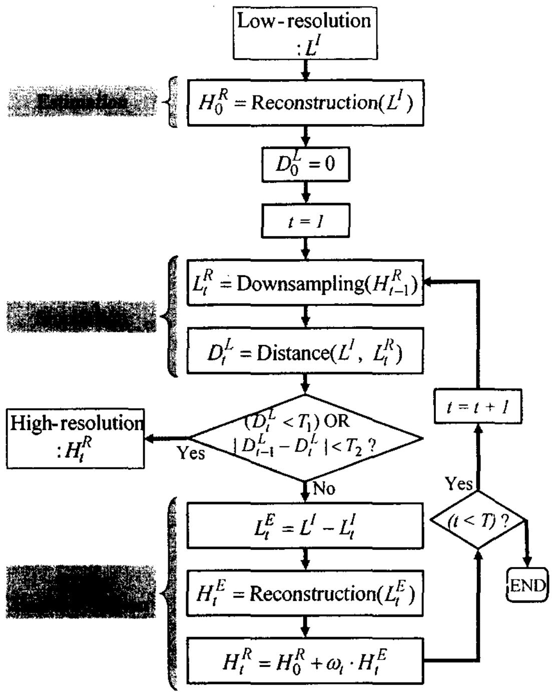

Figure 2. Flowchart of the recursive error back-projection procedure

First, as an initial procedure, we estimate the high-resolution data(H_0^R) from input low-resolution one(L^I) by using our solution of least square minimization described in reference [5].

Second, as a simulation procedure, in order to verify the accuracy of our reconstruction method, simulate the low-resolution data(L_1^R) from estimated high-resolution one 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 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 current reconstruction error and one threshold value(T_1) and whether the iteration is convergent or not by comparing amount of previous and current distance and another threshold value(T_2). If one or two of both comparisons are satisfied, current result of reconstruction is considered as output high-resolution data, otherwise following error back-projection procedure is applied.

Third, as an error back-projection procedure, we create low-resolution error between input low-resolution and simulated one by simple difference operation, estimate high-resolution error data by our reconstruction of low-resolution error data, then compensate initially estimated high-resolution data by adding currently estimated error to it, in order to improve 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 the current distance, that is, if the distance is large then the weight is also large. We recursively perform the same procedure until an accurate estimation is achieved or iterations are converged.

4. Experimental Results and Analysis

4.1 Face database

For testing the proposed method, we used about 200 images of Caucasian faces that were rendered from a database of 3D head models recorded by a laser scanner(Cyberware^{TM})[6-7]. 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 technique. PCA was applied to a random subset of 100 facial images for constructing bases of the defined face model. The other 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. The correspondence is defined between a reference facial 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 of the reconstructed shape and reconstructed texture.

Figure 3 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 described in section 2) and proposed method 2(enhancement of (d) by recursive error back-projection of section 3), respectively. (f) is the original high-resolution facial images.

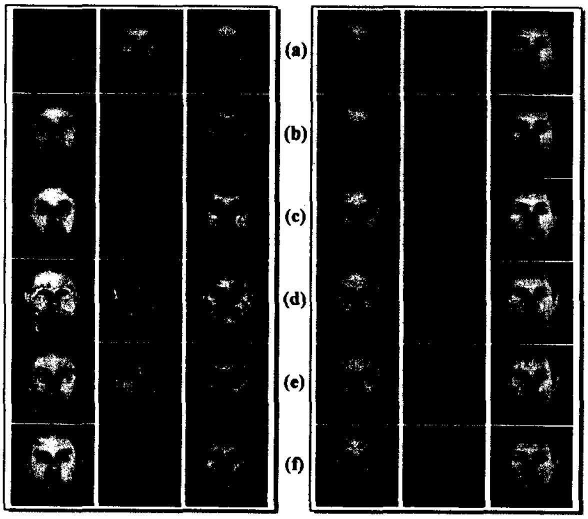

Figure 3. Examples of reconstructed high-resolution face from low-resolution one

As shown in figure 3(d), the reconstructed images by the proposed top-down machine learning are more similar to the original ones and clearer than others. More over the effect of improving the results by recursive error back-projection can be notified in the reconstruction results of 16×16.

Figure 4 shows the comparison of mean reconstruction errors in shape, texture and facial image from the original high-resolution image. Horizontal axes of the figure represent the input low-resolution, two interpolation methods, the proposed reconstruction method and the improved results by 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.

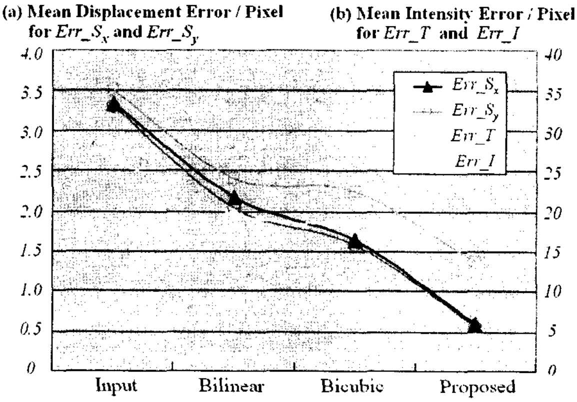

Figure 4. Comparison of mean reconstruction errors

From the encouraging results of the proposed method as shown in figure 3 and figure 4, we can expect that our reconstruction method can be used to improve the performance of the face recognition by reconstructing a high-resolution facial image from a low-resolution facial images captured at visual surveillance systems.

Reference

[1] Tom, B. and Katsaggelos, A. K., *"Resolution Enhancement Monochrome and Color Video Using Motion Compensation," IEEE Trans. on Image Processing*, Vol.10, No.2, pp. 278-287, February 2001.

[2] Baker, S. and Kanade, T., *"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] Kang, M. G. and Chaudhuri, S., *"Super-Resolution Image Reconstruction," IEEE Signal Processing Magazine*, Vol. 23, No. 3, pp. 19-36, May 2003.

[4] Jones, M. J., Sinha, P., Vetter, T. and Poggio, T., *"Top-down learning of low-level vision tasks,"* Current Biology[brief communication], Vol. 7, pp. 991-994, 1997.

[5] Park, J.-S. and Lee, S.-W., *"Resolution Enhancement of Facial Image Based on Top-down Learning,"* Proc. of ACM SIGMM 2003 Workshop on Video Surveillance, November 2003, pp. 59-64.

[6] Blanz, V. and Vetter, T., *"Morphable Model for the Synthesis of 3D Faces,"* Proc. of SIGGRAPH'99, Los Angeles, pp. 187-194, 1999.