Face Hallucination by the Aid of Weighted Image Pairs Patches

S. M. Mostafavi , Sabzevar Tarbiat Moallem University
J. Haddadnia, Sabzevar Tarbiat Moallem University
P. Moallem , University of Isfahan
mostafavi@ieee.org
Paper Reference Number: 0123-160
Name of the Presenter: S. M. Mostafavi

Abstract

This paper is concerned with a practical face super resolving (hallucinating) approach. The image reconstruction is based on high and low resolution image pairs. A patch is removed from the low resolution input image and is compared with the low resolution training images patches. By this comparison a weight for each low resolution images patch is gained. After estimating the corresponding weights, the learning databases high resolution images patch which is a pair for the low resolution images patch is used to hallucinate the input low resolution image. The high resolution output image is constructed by integrating the high resolution hallucinated patches. Unlike neither statistical nor manifold approaches this method uses the exact available data in its learning database to hallucinate the output image. Experimental results show that the proposed method generates higher quality hallucinated images demonstrating performance superiority and novelty in terms of single modal face hallucination.

Key words: Face hallucination, Face Super-resolution, Image Patches

1. Introduction

The everyday increasing demand on surveillance cameras are quite visible nowadays. Banks, Stores and parking lots are some examples of the major users that can be mentioned. When using security cameras or camera recording by far distances we usually face low resolution - low quality facial images which do not have sufficient information for recognition or other usages and require a boosted upgrade in resolution different automatic or manual processes. And there are always the high frequency details that carry the tyupical information used in processing technics. Surveillance and monitoring systems like many other video based applications must extract and enhance small faces from a sequence of low-resolution frames - Tom and Katsaggelos (2001), and Liu et. Al (2005). Direct interpolating of the input image is the simplest way to increase image resolution with such algorithms as cubic spline or nearest neighbor. But on the other hand the performance of direct interpolation is usually blurry and poor since no new information is added in the process. Super resolution has diverse applications such as in medical images, aerial images, surveillance cameras expanding web images, improving old pictures, converting NTSC television images into HDTV and etc. There are two main methods in resolving low resolution images the first one is using reconstruction-based methods and the other one is using learning-based methods. The former method uses a sequence of images as the input and reconstructs and output image by the multiple inputs and the latter method uses learning, created by given previous given information from a database to form a higher resolution image from the input low resolution image.

Super resolution is the process of increasing the resolution in images. If the image that is being super resolved is a human facial image, the process in this special domain is called face hallucination. The term "face hallucination" was coined the first time by Baker and Kanade (2000) and (2002). Their technic precisely functions by disintegrating an image into a pyramid of features including the first and second derivatives of the Gaussian pyramid and the Laplacian pyramid, and then searching the nearest neighbors of each pixel in this pyramid through a training dataset. In order to generate the target HR image, the high-frequency pyramid features for the input pixels are selected based on a MAP formulation of the low resolution improvement problem.

The proposed method relies on the second method. Our main aim in face hallucinating is to reach a high quality super resolved image from the low resolution single frame input image. Since face detection is one of important research issue by itself and needs its own algorithmic and conditional considerations, it will be remained beyond the scope of this paper and the input images will all be previously extracted and cropped manually. Extensively surveyed papers on face recognition are presented by Yang et al. (2005) and Kakumanua et al. (2007). The given cropped facial image is the input of the hallucination algorithm. Besides Baker and Kanades model there have been some other models solving the low resolution problem. A non-parametric patch-based prior along with the Markov random field (MRF) model to generate the desired HR images was executed by Freeman et al. (2002). Liu et al. (2001) propose to integrate a holistic model and a local model for SR reconstruction. The performed two-step approach is executed by integrating a global parametric model with Gaussian assumption and linear inference and a nonparametric local model based on MRF. Both of the two methods used complicated probabilistic models and were based on an explicit resolution-reduction-function, which is sometimes unavailable in practice - Liu et al. (2005). Inspired by a well-known manifold learning method, Locally Linear Embedding (LLE), Chang et al. (2004) implemented the Neighbor Embedding algorithm based on the assumption that small patches in the low- and high-resolution images form manifolds with similar local geometry in two distinct spaces and the local distribution structure in sample space is preserved in the down-sampling process, where the structure is encoded by patch-reconstruction weights. Wang and Tang (2005) suggested a face hallucination method using principal component analysis (PCA) to represent the structural similarity of face images. However, this method only utilizes global information without paying attention to local details. Park et al. (2008) proposed a technic derived from example-based hallucination methods and morphable face models. An example-based Bayesian method for 3D-assisted pose-independent facial texture SR was introduced by Mortazavian et al. (2009). Bayesian framework was implemented to reconstruct a high-resolution texture map given a low-resolution face image. Their method utilizes a 3D morphable model to map facial texture from a 2D face image to a shape- and pose- normalized texture map and vice versa.

The remainder of this paper is organized as follows. Section 2 briefly reviews the low resolution problem more technically. Section 3 and 4 explain the proposed hallucination method. Finally the experimental results are brought up in section 5.

2. Creating the High and Low Resolution Patches

The two dimensional input image will be converted into a vector of the pixels at the first stage. Each image pair in the database is created by one HR image with its low quality image that is the smoothed and down-sampled image of the HR pair. Both high and low resolution images are disintegrated into fixed size image patches while having an overlap region with their neighbor patch. Each low resolution image patch will be a pair for its corresponding HR patch. The ratio of the size of the low resolution patch to HR patch is proportional to the ratio of the HR image to the low resolution image. Each patch will be reshaped into a vector for further process. Each facial image will be considered as a matrix like $\{X^M(i,j)\}_{M=1}^N$ made of overlapping image patches where N is the total number of patches in image X. The patch in row i and column j in the patch matrix is mentioned by $X^M(i,j)$. Assume that each square patch fills $n \times n$ pixels. For the low resolution image $\{X^M(i,j)\}_{M=1}^N$ if n is an odd digit, the patch $X_L^M(i,j)$ will be overlapping $n \times [(n-1)/2]$ pixels with its neighbor patch and for its proportional HR image the patch $X_H^M(i,j)$ which fills $qn \times qn$ pixels will have overlapping on $(qn) \times [q(n-1)/2]$ pixels. And if n is an even digit, the patch $X_L^M(i,j)$ will be overlapping $n \times (n/2)$ pixels with its neighbor patch and for its proportional HR image the patch $X_H^M(i,j)$ which fills $qn \times qn$ pixels will have overlapping on $qn \times (qn/2)$ pixels.

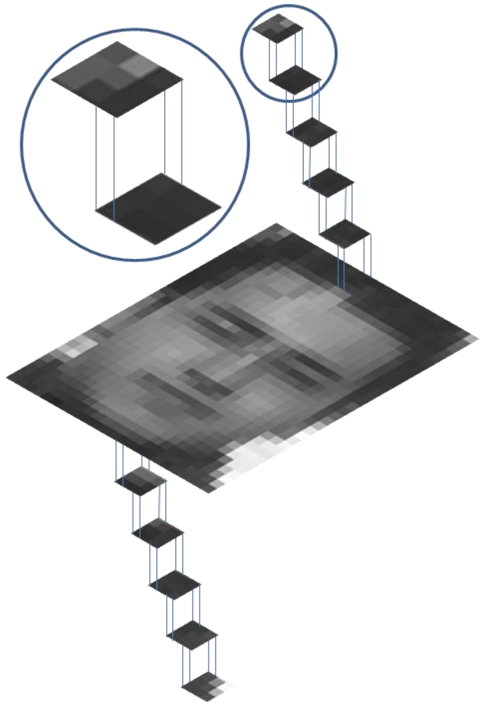


Fig 1: Creation of the patches from the facial image

3. Reconstruction Based on Patches

Assume that X^p stands for the available learning database images having $p=1,2,\ldots,P$ where P is the number of these images. Each patch in the learning database will be in the form of $\{X^{pM}(i,j)\}_{M=1}^{N}$. The reconstructions weight matrix will be as $w_p(i,j)$ where $w_p(i,j)$ is the representative for each patches share in the (i,j) location for reconstructing the input image located in the same mentioned location.

Each mentioned weight in its (i,j) location is considered in a way that the sum of the total weights will become unit. Suppose that Y is the image which will be super resolved and consider Y as patches in the form of $\{Y^M(i,j)\}_{M=1}^{N}$. For the images in the learning database, the images available at the location (i,j) like $X^{pM}(i,j)$, will be a placed patch for the image patch of $Y^M(i,j)$. We expect that each patch like $Y^M(i,j)$ in the human face image of $\{Y^M(i,j)\}_{M=1}^{N}$ to be expressed as Eq. 1.

$$Y^M(i,j)=\sum_{p=1}^{P} w_m(i,j)X^{pM}(i,j)+e \qquad (1)$$

Where e remarks the reconstruction error. From Eq. 1 we can realize that the optimum weights for reconstructing the image depends on minimizing the quantity of the error e .

$$w(i,j) = \underset{w_p(i,j)}{\arg\min} \left\| Y^M(i,j) - \sum_{p=1}^{P} w_m(i,j) X^{pM}(i,j) \right\|^2 \qquad (2)$$

Where $w(i,j)$ is a P dimensional weight vector for each reconstruction weight of $w_p(i,j)$.for each value of $p = 1,2,\ldots,P$ the value of U can be calculated by Eq. 3.

$$U = Y^M(i,j).A^T - X \qquad (3)$$

Where A is a column vector of ones and U is matrix which its columns are the values of the $X^{pM}(i,j)$ patches. The local matrix of V can be found by $V = U^T U$. Eq. 2 is a constrained least squares problem where its solution can be as :

$$w(i,j) = \left(V^{-1}A\right) / \left(A^T V^{-1} A\right) \qquad (4)$$

A better way in obtaining $w(i,j)$ is solving the linear system $V.w(i,j) = A$ followed by changing the scales in order to a sum of one in the patches weights. $w(i,j)$ can be used to reconstruct the new image patch $Y_R^M(i,j)$ from Eq. 5 as:

$$Y_R^M(i,j) = \sum_{p=1}^{P} w_p(i,j) X^{pM}(i,j) \cong Y^M(i,j) \qquad (5)$$

Where $Y_R^M(i,j)$ is a vector converted into a matrix used to construct the whole image. The whole image is constructed by integrating the patches considering their original locations. The values of the overlapping pixels are calculated by averaging the pixel values between the two overlapping neighbor patches.

All the HR patches $Y_H^M(i,j)$, are integrated to reconstruct the HR whole final output image $\{Y_H^M(i,j)\}_{M=1}^{N}$ considering their original locations. The values of the HR pixels which are in the overlapping regions are calculated by averaging the pixel values between the two overlapping HR neighbor patches.

The whole hallucinating process can be described in the following steps:

- The input image will be reshaped into patches.
- Each low resolution patch is compared to the proportional low resolution patches in the learning database according to their position and location and receives one weight by each comparison.
- The concluded weights crated by comparing the low resolution images are used as weights proportionally on the HR pair.
- The HR output is created by integrating the HR output images created in the third step.

4. Experimental Results

In order to create the learning image database 600 frontal face images were used from multiple online available databases like Essex, Yale, ORL, AR and LFW . Using multiple databases increases the functionality and makes the system robust. Although different available races, illuminations, facial poses and other conditions that has effects on the image reconstruction are gathered together in the face database, the proposed system functions hale. Because most automated face processing tasks are possible with 96 × 128 pixel images –

Baker and Kanade (2000). We developed a resolution enhancement algorithm, specifically for faces, that can convert a small number 24 × 32 pixel image of a face into a single 96 × 128 pixel image. It shall be mentioned using this method we can still have an increase in resolution from a desired input image size to a desired output by changing the database image pairs but based on the image pair sizes the quality of super resolving may vary to a higher or lower facial image output. One thing the quality of the output image is based on, is the magnification coefficient which is 4 in the mentioned method.
To quantifiably measure the performance of the proposed algorithm, we evaluated the peak signal-to-noise ratio (PSNR) as measurements between the ground truth face images and the hallucinated facial images.

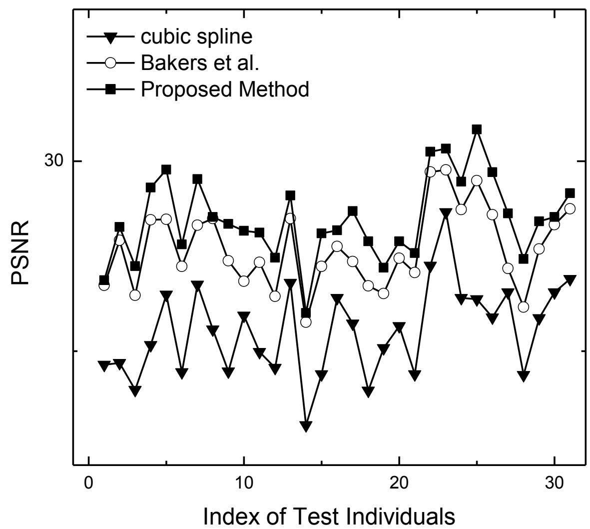
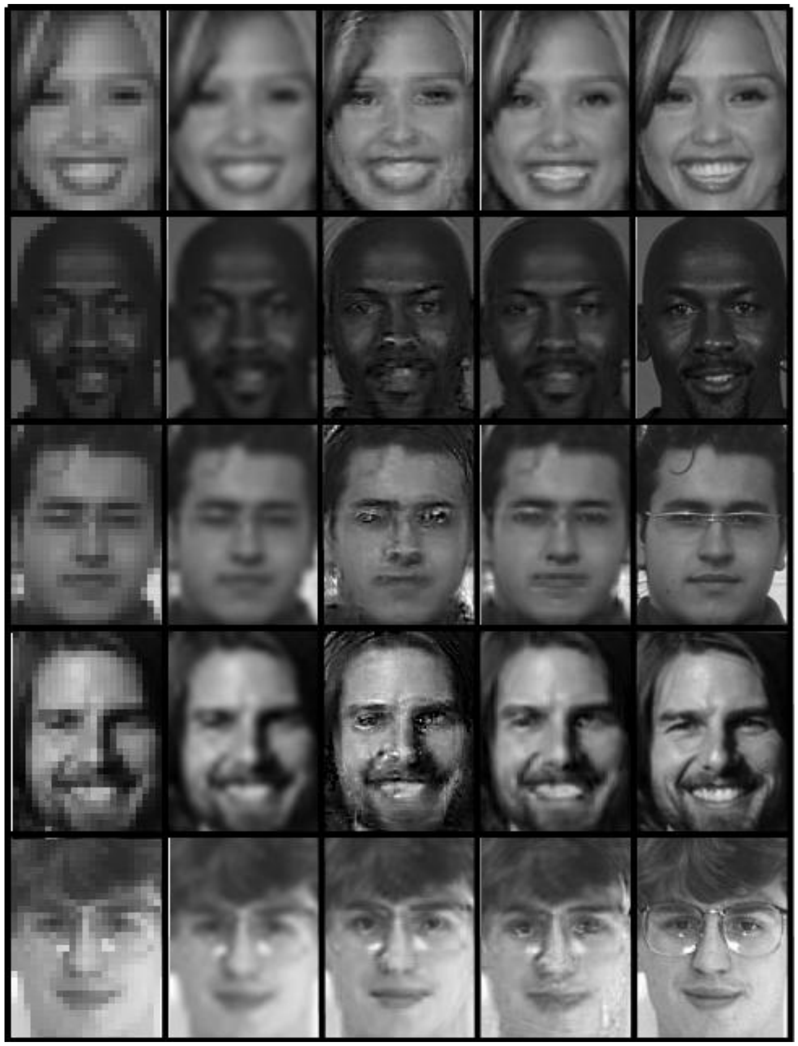


Fig 2: PSNR values of the hallucinated results of three different methods of 31 randomly selected individuals



(a) (b) (c) (d) (e)

Fig 3: Comparing the proposed method a)low resolution facial image b)cuic spline c)Bakers method d)proposed method e)original HR image

The Mean Square Error (MSE) and the PSNR are the two error metrics used to compare image compression quality which have clear physical meanings and are simple to compute. The MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represents a measure of the peak error. The PSNR values for a sample of randomly selected individuals are plotted in Fig. 5. From Fig. 5, our proposed method has better score than the other two techniques in terms of PSNR which was conformed also by the visualization of Fig. 3.

Comparing Figs. 3 and 5, our proposed method has better score than the other two techniques in terms of PSNR and SSIM and also by the means of human visualization.

5. Conclusion

A new face hallucination method based on the acquired weights form image pair patches was described. High and low resolution image pairs in the image database were the key for in the learning stage. A patch is removed from the low resolution input image and is compared with the low resolution training images patches. By this comparison a weight for each low resolution images patch is gained. After estimating the corresponding weights, the learning databases HR images patch which is a pair for the low resolution images patch is used to hallucinate the input low resolution image. The HR output image is constructed by integrating the HR hallucinated patches. Experimental results demonstrate performance superiority and novelty in terms of single modal face hallucination.

Fig 4: A real world image brought from the CMU image database with the hallucinated facial images based on the proposed method

References

Tom, B. & Katsaggelos, A. (Feb. 2001) Resolution enhancement of monochrome and color video using motion compensation, *IEEE Trans. Image Process., vol. 10, no. 2, pp. 278–287.*

Liu, W., Lin, D. & Tang, X. (Jul. 2005) Neighbor combination and transformation for hallucinating faces, *in Proc. IEEE Conf. Multimedia and Expo, Amsterdam, The Netherlands, , pp. 145–148.*

Baker, S. & Kanade, T. (2000) Hallucinating Faces, *in Proc. of Inter. Conf. on Automatic Face and Gesture Recognition, pp. 83-88.*

Baker, S. & Kanade, T. (2002) Limits on super-resolution and how to break them, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(9):1167–1183.*

Yang, M.-H., Kriegman, D. J. & Ahuja, N. (Jan. 2002) Detecting faces in images: A survey, *IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 1, pp. 34–58.*

Kakumanua, P. Makrogiannisa, S.& Bourbakis, N. (Mar. 2007) *A survey of skin-color modeling and detection methods," Pattern Recognit., vol. 40, no. 3, pp. 1106–1122.*

Freeman, W. T., Jones, T. R. & Pasztor, E. C. (2002) Example-Based Super-Resolution, *IEEE Comput. Graph. Appl., vol. 22, no. 2, pp. 56-65.*

Liu, C., Shum, H. & Zhang, C. (2001) A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Nonparametric Model, *in Proc. of CVPR, Vol. 1, pp. 192- 198.*

Liu,W., Lin, D.H., Tang, X.O., (2005) Hallucinating faces: Tensor Patch super-resolution and coupled residue compensation,*in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,United States, pp.478–484.*

Chang, H., Yeung, D. Y., & Xiong, Y. (2004)Super-resolution through neighbor embedding, *in Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 275–282.*

Wang, X. & Tang, X. (2005) Hallucinating face by eigentransformation, *to appear in IEEE Trans. on Systems, Man, and Cybernetics, Part-C, Special issue on Biometrics Systems.*

Park, J. S. & Lee, S. W. (2008) An example-based face hallucination method for single-frame, low-resolution facial images. *IEEE Transactions on Image Processing, 17(10):1806–1816.*

Mortazavian , P., Kittler, J.V. & Christmas, W.J. (2009) 3D-Assisted Facial Texture Super-resolution, *BMVC09 xx-yy*

Penev, P. S. & Sirovich, L. (2000) The Global Dimensionality of Face Space, *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 264-270.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli ,E. P. (2005) Image quality assessment: From error visibility to structural similarity. IEEE Trans. on Image Processing, 27(4):619–624.

Faces 94 Faces 95 Faces 96 and Grimace - Face Recognition Database, *University of Essex, UK* available online *http://cswww.essex.ac.uk/mv/allfaces/grimace.html*

AT&T "*The Database of Faces*" (formerly "*The ORL Database of Faces*") available online *http://www.cl.cam.ac.uk/*Research/DTG/attarchive/facedatabase.html

The Yale Face Database available online *http://cvc.yale.edu/projects/yalefaces/yalefaces.html*

AR Face Database - available online - *http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html*

Labeled Faces in the Wild - available online - http*://vis-www.cs.umass.edu/lfw/*