Hallucinating Facial Images and Features

Bo Li[1], Hong Chang[2,3], Shiguang Shan[2,3], Xilin Chen[2,3], Wen Gao[2,4]
1 Harbin Institute of Technology, Harbin, China
2 Digital Media Research Center, Institute of Computing Tech., CAS, Beijing, China
3 Key Laboratory of Intelligent Information Processing, CAS, Beijing,China
4 Digital Media Institute, Peking University, Beijing, China
bli@jdl.ac.cn; {sgshan, changhong, xlchen}@ict.ac.cn; wgao@pku.edu.cn

Abstract

In facial image analysis, image resolution is an important factor which has great influence on the performance of face recognition systems. As for low-resolution face recognition problem, traditional methods usually carry out super-resolution firstly before passing the super-resolved image to a face recognition system. In this paper, we propose a new method which predicts high-resolution images and the corresponding features simultaneously. More specifically, we propose "feature hallucination" to project facial images with low-resolution into an expected feature space. As a result, the proposed method does not require super-resolution as an explicit preprocessing step. In addition, we explore a constrained hallucination that considers the local consistency in the image grid. In our method, we use the index of local visual primitives [5] as features and a block-based histogram distance to measure the similarity for the face recognition. Experimental results on FERET face database verify that the proposed method can improve both visual quality and recognition rate for low-resolution facial images.

1. Introduction

Facial image analysis has been widely studied for many years, with more specific research problems including face detection, recognition, expression analysis and animation. However, the performance of most existing systems is affected by the resolution of facial images. For example, low-resolution (LR) facial images, which are captured by surveillance cameras when the distance between the human and the camera is large, limit the performance of face recognition systems. In order to solve LR face recognition problems, traditional methods usually employ super-resolution (SR) as a preprocessing step to get a high-resolution (HR) image and then pass the super-resolved face image to some face recognition system.

General image SR methods can be adopted in the preprocessing step. One of the simplest ways to increase the resolution is interpolation. However, the performance of the interpolation algorithm is poor because the smoothness prior does not fit at region boundaries. Many more effective SR algorithms have been proposed during the past decades. Especially, learning-base methods gained a great of interest in computer vision and pattern recognition community [1]. Methods of this class usually learn nonparametric models from a training set to predict HR image.

As for facial image SR, more specific methods have been studied. Baker and Kanade [2] developed a learning-base method named "face hallucination". This method uses the prior of the facial images in a training set to infer the missing high-frequency components from an input LR image. Liu et al. [3] proposed a two step statistical modeling approach that integrates a global parametric model and a local nonparametric model. However, both of the two renders are sensitive to image alignment, scale, and noise. Instead of reconstructing a HR image explicitly, the method proposed in [4] carries out SR in the eigenface space and only outputs the weights along the principal components for face recognition purpose.

In this paper, we propose a novel method for simultaneous image and feature hallucination based on neighbor embedding [5]. While our method resembles [4] in performing SR in feature space, it is even superior due to globally nonlinear feature reconstruction. On one hand, the hallucinated HR facial images possess good visual quality. On the other hand, the hallucinated features are even more effective for face recognition. It is worth noting that we make use of local visual primitives (LVPs) [6] in feature representation and propose local constrained neighbor selection in image/feature reconstruction. Both

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

strategies play important roles in making our algorithm successful. Preliminary experimental results show that our method can render visually superior HR estimate and improve the performance of LR face recognition systems.

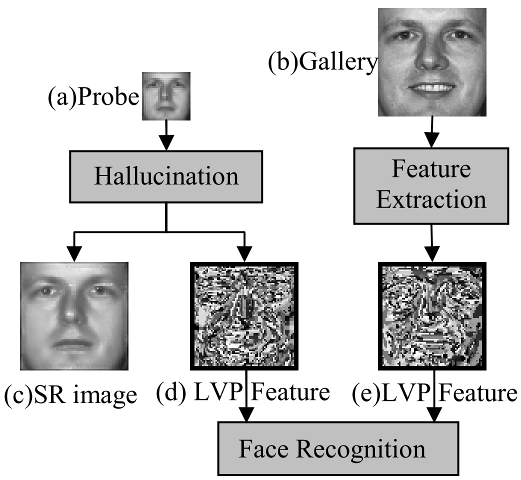


Figure 1. The overview of our approach.

2. Hallucinating images and features

An overview of our approach is shown in Fig. 1. Given a LR facial image as input, our hallucination consists of two parallel threads, image hallucination and feature hallucination. In the former thread, the input LR image is used to predict the corresponding HR image with the help of a training set. In the latter one, the corresponding feature of the HR image is reconstructed for the face recognition task. In the immediate subsection, we introduce the problem formulation, including the training set and feature representation.

2.1. Problem formulation

Our original training set is constituted by a set of image pairs of HR facial images and their LR counter-parts. For each training pair, we perform interpolation on the LR image to generate the low frequency (LF) part of the HR image. The HR/LR image pair and the LF part are denoted as $\mathbf{H}$, $\mathbf{L}$ and $\mathbf{I}_l$, respectively. In fact, a SR algorithm only need to estimate the differences between HR image and corresponding LR one, which is the missing high frequency (HF) component denoted as $\mathbf{I}_h$.

Similar with some previous learning-base SR methods, our algorithm is performed on patch-wise way. The patches are extracted from $\mathbf{H}$, $\mathbf{I}_h$ and $\mathbf{I}_l$ respectively. Besides, we record the location of the patch center in the image. Let $\mathbf{h}^i$, $\mathbf{p}_h^i$, $\mathbf{p}_l^i$ and x^i represent HR patch, HF patch, LF patch, and location in image, respectively, the i-th training sample can then be denoted as $\mathbf{t}^i = (\mathbf{h}^i, \mathbf{p}_h^i, \mathbf{p}_l^i, x^i)$. For convenience, we use the above notations to represent (concatenated) feature vectors as well as patches and training sample.

To learn the relationship between LR images and the corresponding HR feature vectors, we need to extract HR features from HR images in the training set. In this paper, we use LVP index matrix $\mathbf{M}$ to represent features for face recognition tasks. More specifically, we first perform K-means clustering algorithm to learn a set of cluster centroids for the HR patches with the size of m by m (m = 7) pixels in training set. We refer to these centroids as LVPs. In our experiments, we use 256 LVPs denoted as $\Omega = \{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_{256}\}$. For each site x in a HR image, we compute the index of the LVPs which is most similar to the patch centering at x and assign this index to the corresponding entry of $\mathbf{M}$. In Fig. 1, (e) are the corresponding LVP feature for (b). We extract features for all HR images in the training set. Then our training sample is now represented as $\mathbf{t}^i = (\mathbf{h}^i, \mathbf{p}_h^i, \mathbf{p}_l^i, x^i, m^i)$. Here, m^i is the feature value at site x^i in the corresponding LVP index matrix. Finally, a training set $\mathbf{T} = \{\mathbf{t}^i\}_{i=1}^{K}$ to be used for image and feature hallucination is established, where K is the number of samples. In the next subsection, we explain in detail our image and feature hallucination algorithm.

2.2 Proposed algorithm

Given a LR facial image, our hallucination algorithm simultaneously estimates the HR facial image and the expected feature for the recognition system. Different from most previous methods, the features for face recognition do not need to be extracted from the super-resolved image.

In our work, we adopt the idea of neighbor embedding algorithm [6], which is inspired mainly from a manifold learning method, Locally Linear Embedding (LLE) [7]. The algorithm is based on a manifold assumption, which assumes that patch pairs from the LR manifold and the corresponding HR manifold possess similar local geometry. Here the local geometry is characterized by the reconstruction weights that are used to reconstruct the LR patch by its neighborhoods on LR manifold. The target HR patch is rendered by replacing the LR neighbor patches with HR counterparts, while preserving the reconstruction weights. However, the assumption is not always hold well because the mapping from LR to HR is one-to-multiple. For example, a LR mouth sub-image may be the most similar with some LR eye sub-image while their HR counterparts may be distinctly different from each other. In order to address this problem, we introduce, in this paper, a new constraint to reduce the confusion caused by the one-to-multiple mappings. The basic idea is that the patches in neighborhood should have similar semantic property with the query patch in HR manifold, as well as being close to the patch on the LR manifold.

It is well-known that facial images have similar semantic structures. When we roughly align the HR facial images with the positions of two eyes, it means that the nearer the position, the more similar the semantic property. So we can use the spatial distance to measure the semantic consistency between the patches centering at different positions in image grids.

Based on the statements above, we propose a new energy function with local constraints as:

$$E(\mathbf{w}^x, \mathbf{p}_h^x) = \psi(\mathbf{p}_l^x, \mathbf{w}_i^x) + \alpha\phi(\mathbf{p}_h^x, \mathbf{p}_h^y) \tag{3}$$

subject to $\| x^i - x \| < \sigma_x$. Where x represents a site in the input image and y its special neighbor sites, x^j is the site in the j-th training image, and α is a tuning parameter.

The first term in (3) is the reconstruction error function that is expressed as:

$$\psi(\mathbf{p}_l^x, \mathbf{w}_j^x) = \| \mathbf{p}_l^x - \sum_{j\in N(x)} \mathbf{w}_j^x \mathbf{p}_l^j \|^2 . \tag{4}$$

Where, $\mathbf{p}_l^x$ is the query LR patch at site x, and $N(x)$ represents its neighborhood set on the LR manifold, with $\mathbf{p}_l^j$ being a neighbor in set $N(x)$. $\mathbf{w}_j^x$ is the reconstruction weight to be estimated. The second term in (3) evaluates the smoothness between the target HR patch $\mathbf{p}_h^x$ and its special neighbors $\mathbf{p}_h^y$. We use Sum Squared Difference (SSD) to define the overlapping properties between $\mathbf{p}_h^x$ and $\mathbf{p}_h^y$, as denoted by the function $D_{overlap}(\mathbf{p}_h^x, \mathbf{p}_h^y)$ in the following equation:

$$\phi(\mathbf{p}_h^x, \mathbf{p}_h^y) = \sum_y D_{overlap}(\mathbf{p}_h^x, \mathbf{p}_h^y) . \tag{5}$$

From the above definitions, we can see that the energy function in (3) is convex to $\mathbf{w}^x$ and $\mathbf{p}_h^x$. Hence, global optimal $\mathbf{p}_h^x$ can be achieved by minimizing (3).

The proposed hallucination algorithm is summarized in Table. 1. We implement step 4 with a voting method. More specifically, each neighbor casts votes, with its feature value m^i, for the possible target feature. The value, which obtained the most votes, is chosen as the feature estimate at site x. In our experiments we set k = 3 and α =0.5. After this hallucination procedure a super-resolved face image and its feature matrix are predicted simultaneously.

Table 1. Hallucination Algorithm

Input: LR patches $\{\mathbf{p}_l^x\}_{x=1}^X$.

Output: HR patches $\{\mathbf{p}_h^x\}_{x=1}^X$ and feature matrix **M**.

For each query LR patch $\mathbf{p}_l^x$,

1. Find k neighbors $\mathbf{p}_l^i$ using KNN by Euclidean distance from training set while subjecting to $\| x^i - x \| < \sigma_x$. Here, x^i is the corresponding site for $\mathbf{p}_l^i$ in training vector $\mathbf{t}^i$.
2. Calculate reconstruction weight vector $\mathbf{w}^x$ by minimizing the following function:
$$E = \| \mathbf{p}_l^x - \sum_{j\in N(x)} \mathbf{w}_j^x \mathbf{p}_l^j \|^2 + \alpha \sum_y D_{overlap}(\sum_{j\in N(x)} \mathbf{w}_j^x \mathbf{p}_h^j, \mathbf{p}_h^y) .$$
3. Predict the target HR $\mathbf{p}_h^j$ patch by
$$\mathbf{p}_h^x = \sum_{j\in N(x)} \mathbf{w}_j^x \mathbf{p}_h^j .$$
4. Predict the entry m^x of the feature matrix **M** at site x by m^i that is the corresponding feature value for each neighbor.

Repeat the step 1 to step 4 until all HR patches and entries of the feature matrix are estimated.

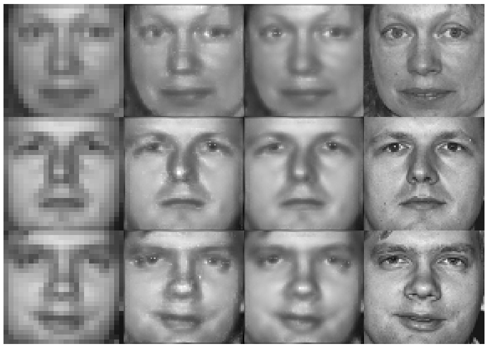
3. Experiments

In this Section, We performed two experiments to demonstrate the efficacy of the proposed method. The first experiment is face hallucination. We investigate the effectiveness of the local consistency. The second experiment is LR face recognition, where we compare our method with some traditional methods.

We carry out experiments on FERET [8] face Database. All 72×72 HR facial images used in experiments are aligned with the position of two eyes. The HR image is smoothed and down-sampled to a LR 24×24 image. The training set has 1002 frontal face images. We evaluate the method based on the standard gallery (1196 images of 1196 subjects) and the *fafb* probe sets (1195 images).

3.1. Face hallucination

We show several face hallucination results on several examples. If no constraint is added to the procedure of the face hallucination, the resulting face images, as shown in Fig. 2(b), are confused by artifacts. The reasons lie in the situation: the neighborhoods selected for reconstruction lie at the sites which are



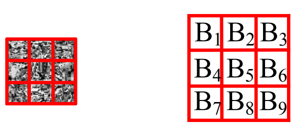
(a) (b) (c) (d)

Figure 2. Comparison of the proposed method with neighbor embedding algorithm. (a) 24×24 LR images. (b) Neighbor embedding in [6]. (c) The proposed method with σ_x = 9. (d) 72×72 HR

distinctly far from the target site. In Fig. 2(c), the results of our method with $\sigma_x = 9$ are more close to the original HR images, because the local constraint can retrain the uncertainty caused by the mapping from LR to HR.

3.2. Face recognition

We study the recognition performance using HR face images, LR face images, and hallucinated face images and hallucinated features. In all experiments, we use the HR LVPs with size of 7×7 while LR LVPs with size of 3×3. For the similarity measurement, we employ a distance between the block-based histograms. As show in Fig. 3, the LVP feature matrix is separated in blocks. For each block B_i, we compute a histogram of the LVP index and denote as H_i. Then the face representations become $R = (H_1, H_2, ..., H_N)$, N is the block number. We use $S(R, R') = 1 - |R - R'|$ to measure the similarity between two representations. Please refer to [5] for details about LVP for face recognition.



(a) Blocks (b) Block girds

Figure 3. Blocked LVP feature matrix.

From Fig. 4, we can observe that the proposed method improve the performance of face recognition on LR images. The performance using the hallucinated features for face recognition is better than that using LR image in all case. The proposed method is better than using hallucinated images when the latter achieve the best recognition accuracy at block number N=36.

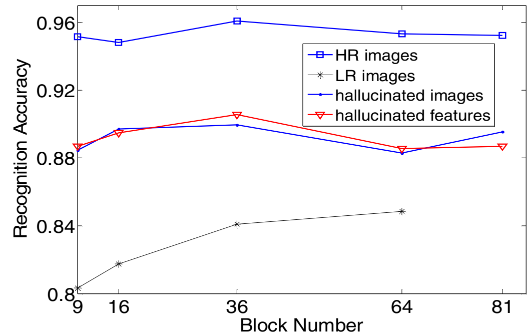


Figure 4. Recognition ratio using HR face images, LR images, hallucinated images, and hallucinated features by σ_x = 9 with different block numbers.

4. Conclusion

In this paper, we propose a novel face hallucination method for LR face recognition purpose. In our method, The HR facial feature is hallucinated from LR facial image directly. In addition, we introduce a local constraint to prevent the confusion caused by the one-to-multiple from LR to HR. The results of face hallucination and face recognition have shown that the proposed method can improve the visual quality and the recognition rate for LR facial image.

Acknowledgement

This paper is partially supported by National Natural Science Foundation of China under contract No.60673091; Hi-Tech Research and Development Program of China under contract No.2006AA01Z122.

Reference

[1] W. Freeman, E. Pasztor, and O. Carmichael. Learning low-level vision. *IJCV*, 40(1):25–47, 2000.

[2] S. Baker and T. Kanade. Hallucinating faces. *Automatic Face and Gesture Recognition*, Mar. 2000.

[3] C. Liu, H. Shum, and C. Zhang. A two-step approach to hallucinating faces: Global parametric model and local nonparametric model. *CVPR*, pp. 192–198, 2001.

[4] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes,and R. M. Mersereau. Eigenface-domain super-resolution for face recognition. IEEE Trans. Image Processing, vol. 12, pp. 597–606, 2003.

[5] X. Meng, S. Shan, X. Chen, W. Gao. Local Visual Primitives (LVP) for Face Modelling and Recognition. *ICPR*, pp. 536-539, 2004.

[6] H. Chang, D. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. CVPR, volume 1, pp. 275-282, 2004.

[7] S.T. Roweis, L.K. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, vol. 290, pp. 2323–2326, 2000.

[8] P. Philips, H. Moon, P. Pauss, and S. Rivzvi. The feret evaluation methodology for face-recognition algorithms. *CVPR*, pp. 137–143, 1997.