AN EFFICIENT LOCAL-STRUCTURE-BASED FACE-HALLUCINATION METHOD

Zhuo Hui and Kin-Man Lam

Centre for Signal Processing, Department of Electronic and Information Engineering,
The Hong Kong Polytechnic University, Kowloon, Hong Kong

ABSTRACT

In this paper, we propose a novel patch-based face-hallucination algorithm, which is based on the local structure kernels established via the relation between interpolated low-resolution (LR) images and their corresponding high-resolution counterparts. In our algorithm, the local linear embedding (LLE) algorithm is used to extract local structures, and the kernels are then constructed based on non-overlapped patches in the interpolated LR images. The information about local structures as described by the kernels is propagated to the corresponding regions of the HR images. Sub-pixel distortions are refined by solving a constrained problem at pixel level via iterative procedures. Experimental results show that our proposed method can provide a good performance in terms of reconstruction errors and visual quality.

Index Terms— face hallucination, local structure, eigentransformation

1. INTRODUCTION

In most digital-imaging applications, high-resolution (HR) images are preferred or needed, especially in human face recognition, video surveillance, high-definition TV, etc. However, the facial images captured are often of low-resolution (LR) and blurred. Face-hallucination techniques have been introduced, and reconstruction- and learning-based algorithms [1] are usually employed to reconstruct HR images from LR observations. The former class also refers to the interpolation-based method, which aims to generate HR images based on the information given in a single image. However, in recent years, learning-based methods which employ training sets to generate the target HR images have achieved great success, and thereby demonstrating the advantages of edge preserving and details' reconstruction. In [2], Freeman et al. proposed the framework which utilizes the Markov random field to search for suitable patches by using maximum a posteriori (MAP). Spatial consistency between adjacent patches is then refined by using belief propagation. Fan et al. [3] further improved the algorithms by introducing primitive manifold learning in extracting the local-structure information. A certain number of neighbors are searched based on the Euclidean distance for a certain input visual feature, and a linear regression is used to characterize the relationship between the testing visual features and those of the neighbors. The same weights are projected to the corresponding HR patches in order to obtain the target reconstructed results. Ma et al. [5] approached the problem via extracting the relation between LR and HR patches in the same region based on the LLE algorithm [4].

However, both [3] and [5] are focused on reconstructing a single patch at a given position without taking into account the neighboring patches defined by geometric distances. This may cause structural errors or distortions in the feature shape of the local prior. Although [3] searches neighboring patches for a testing input, the neighborhood is defined using the Euclidean distance, i.e. the L_2 norm. Hence, the selected neighbors are possibly the patches at the same position of the reference samples. A similar method is used in [5]. In other words, both [3] and [5] aim to extract information to reconstruct the local prior based on reference samples with the least Euclidean distance, instead of based on the local information in the LR input image.

In our proposed method, the face-hallucination problem is tackled by using a two-stage patch-reconstruction approach. The first stage is to estimate an initial HR version of a LR face image based on a patch-based eigentransformation method. The estimated HR face images reconstructed will exhibit some local-structure distortion. Thus, in the second stage, based on the initially estimated results, the local structure is refined by the local prior extracted from the interpolated LR input and training images. We define a local region composed of nine patches, and we use the LLE method to derive the local prior for each patch with respect to its neighbors in the same region. Finally, we project the local priors back to the corresponding patches obtained in the first stage via an iterative reconstruction.

The remainder of this paper is organized as follows. In Section 2, the details of our proposed method are presented, as are its advantages. In Section 3, the experimental results based on our proposed method and several related algorithms are demonstrated, and finally a conclusion is given.

2. PROPOSED ALGORITHM

2.1 Stage I: Reconstruction using patch-based eigen-transformation

As described in the previous section, we implement two-stage reconstruction. In the first stage of reconstruction, we target the reconstruction of the HR counterparts from their LR face images using only patch-based eigentransformation. Depending on the magnification factor under consideration, pixels in the interpolated input image and the corresponding ground-true HR image may have a small misalignment, which results in local-structure distortions; we neglect this misalignment at this stage, and compensate for it in the second stage. Moreover, facial images possess great similarity in terms of overall structure, i.e. all face images have the same facial features with a similar relative position. By concatenating pixels in a patch in lexicographical order to form a vector, a pixel-to-pixel correspondence between two patches at the same position in both an interpolated LR face image and the HR counterpart can be established. This correspondence between an interpolated LR patch and the corresponding HR patch can be viewed as a

mapping function or a kernel, denoted as T_i, i.e.

$$l_i = T_i h_i , \tag{1}$$

where i denotes the index of the patches, and l_i and h_i represent the interpolated LR patch and the HR patch at the same position, respectively.

Such a correspondence exists between all the patches in the interpolated LR image and its ground-true HR image, but the kernel function to be estimated is different for different patch positions. As in [7], patch-based eigentransformation is employed in our algorithm so that the HR patches can be reconstructed without requiring an estimation of the kernel T_i. Consider a training set of face images which contains pairs of LR and HR face images. The LR training samples are interpolated to the same size as the HR face images, and these interpolated LR images are denoted as $\{l^1, l^2, ..., l^N\}$, where N is the number of training pairs. The corresponding HR face images are denoted as $\{h^1, h^2, ..., h^N\}$. Each image in the training set is divided into overlapped patches. The i^{th} patches of the interpolated LR images and the corresponding HR images are denoted as $\{l_i^1, l_i^2, ..., l_i^N\}$ and $\{h_i^1, h_i^2, ..., h_i^N\}$, respectively. These two sets of interpolated LR patches and HR patches are represented as two matrices, having the patches arranged as columns, as follows:

$$L_i = \begin{bmatrix} l_i^1 & l_i^2 & \cdots & l_i^N \end{bmatrix} \text{ and } H_i = \begin{bmatrix} h_i^1 & h_i^2 & \cdots & h_i^N \end{bmatrix}. \tag{2}$$

Using eigentransformation [6], the i^{th} interpolated patch of a LR face image can be considered as a linear combination of the interpolated LR patches as follows:

$$l_i^t = \sum_{j=1}^{N} c_i^j l_i^j = L_i c_i , \tag{3}$$

where $c_i = \begin{bmatrix} c_i^1 & \cdots & c_i^N \end{bmatrix}^T$ and T is the transpose operation. The coefficients c_i contributed by each patch in the training set can be explicitly expressed as follows:

$$c_i = V_i \Lambda_i^{-1/2} w_i , \tag{4}$$

where V_i and Λ_i are the eigenvectors and the corresponding eigenvalues of the covariance matrix $L_i^T L_i$, and w_i are the weights for the input interpolated LR patch l_i^t projected onto the eigenspace spanned by the eigenvectors of $L_i L_i^T$.

According to LLE [4], we can assume that a LR patch contains the intrinsic feature of the corresponding HR patches. Therefore, the prior knowledge reflected in the LR samples can be used to reconstruct the HR data. Hence, the coefficients that each interpolated patch in the training set contributed to the testing input can be applied to their HR counterparts, which can be estimated as follows:

$$\hat{h}_i^t = \sum_{j=1}^{N} c_i^j h_i^j = H_i c_i . \tag{5}$$

As this is a patch-based approach, blocky artifacts may appear in the reconstructed HR images. In order to reduce the artifacts, the patches involved are overlapped by 50% with adjacent patches, and the pixels in the overlapped regions are merged based on the algorithm proposed in [12]. When all the estimated HR patches have been estimated, the initial estimated results $\hat{I}_t^H$ are generated.

2.2 Stage II: Patch-based local structure refinement

As mentioned previously, the sub-pixel misalignments are ignored in the first stage. Hence, in this stage, we aim to derive local kernels, which can help to constrain the patch-based reconstruction and thus reduce the existing misalignments. Based on the work in [2] and [3], the interpolated LR images can retain their structural information. This means that we can initially estimate the structural information about the target HR image from its interpolated counterpart. Moreover, according to the framework of LLE [4], the structural information of a dataset can be preserved via the embedding, extracted based on the neighbors. However, unlike the neighbors defined in [3] and [4], we use locally derived metrics instead of the Euclidean norm.

2.2.1. Determination of local prior

In our method, we consider nine non-overlapped patches, arranged in 3×3 patches, to form a local region Ω in deriving the embedding, i.e. the contribution of the eight neighboring patches to the patch at the center. This embedding $w_{i,j}^t$ can be determined as follows:

$$\varepsilon = \arg\min_{w_{i,j}^t} \left| l_i^t - \sum_{j=1}^{J-1} w_{i,j}^t l_j \right|^2 \; s.t. \sum_j w_{i,j}^t = 1 , \tag{6}$$

where $l_j \in \Omega$ and j = 1, 2, …, J–1. Ω represents the region used to extract the local-structure information, l_i^t is the testing patch with index i, and J (=9) refers to the number of patches in Ω. Note that, to make the representation simple, we still use l_i and l_i^t to represent a patch in general and an interpolated LR patch, respectively. However, the patch size used in this stage is smaller than that used in the first stage. In this stage, the patch size is 3×3; this can make a better estimation of the local structures.

According to [4], we define the local Gram matrix G as follows:

$$G_i = \left(l_i^t P^T - \eta \right)^T \left(l_i^t P^T - \eta \right), \tag{7}$$

where P is a (J–1)-dimensional column vector of ones, η is a $D \times (J-1)$ -dimensional matrix with its column being the neighbors defined in the region Ω, and D refers to the number of pixels in the local neighboring patch l_j. The least-square problem can be solved by setting the weights as follows:

$$w_{i,j}^t = \frac{G_i^{-1} P}{P^T G_i^{-1} P} . \tag{8}$$

However, the above equation requires an explicit inversion of the local Gram matrix. Hence, according to the previous work [3] and [5], we can simply solve the system of linear equations as follows:

$$\sum_j G_i w_{i,j}^t = 1 . \tag{9}$$

Having determined the local structure of a patch located at the center of a region Ω, in order to reduce the blocky artifacts, the region is then shifted by one pixel to its next position, where the corresponding local structure is extracted. Therefore, the patches at the center of adjacent regions overlap with each other, as illustrated in Fig. 1.

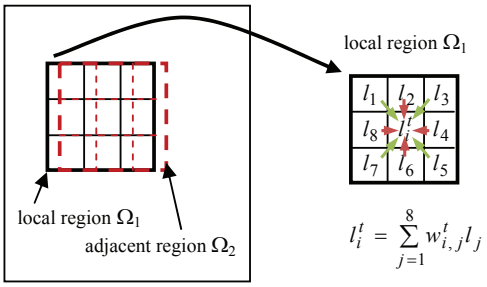
However, as previously illustrated, structural distortions also exist between interpolated LR images and the corresponding HR ones. To tackle these structural distortions, we search similar interpolated LR samples from the training set with respect to the input face, and find the corresponding HR faces. The distortions can be compensated by referring to the HR training samples, which should possess similar local structures to the input face in the same region under consideration. The neighboring LR samples are selected based on the structural similarity, defined using as follows:

$$s_k = \frac{\sigma_{tk}^2}{\sigma_k \sigma_t + \tau}, \tag{10}$$

where τ is a small value in order to avoid the denominator being zero, σ_{tk} is the correlated variance between the testing region and the reference region. It is defined as follows:

$$\sigma_{tk}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(l^t(i)-\mu_t)(l^k(i)-\mu_k). \tag{11}$$

μ and σ denote the mean and standard deviation for the corresponding regions used, N is the number of pixels in the region under consideration, and $l^t(i)$ and $l^k(i)$ represent the i^{th} pixel in the testing region and the k^{th} reference region, respectively.



(a) An interpolated LR image (b) Determination of the local prior

Fig.1 Extraction of local-structure information: (a) The current local region Ω_1 is divided into 3×3 patches, where the local embedding of the patch at the center will be determined. The next adjacent region Ω_2 is the region Ω_1 shifted to the right by one pixel. (b) The local prior $w_{i,j}^t$ is determined for the patch l_i^t.

The larger the value of s_k, the more similar are the local structures of the two regions concerned. Hence, based on the value of s_k, we search for K pairs of interpolated LR and HR samples, and then derive the corresponding weights for both the interpolated LR and HR samples in the same region. We define $(w_{i,j}^k)_L$ and $(w_{i,j}^k)_H$ as the weights that patch j contributes to the center patch i in the region under consideration for both the interpolated LR and HR samples, respectively, and the ratio between $(w_{i,j}^k)_H$ and $(w_{i,j}^k)_L$ as $\gamma_{i,j}^k$, i.e.

$$\gamma_{i,j}^k = \frac{(w_{i,j}^k)_H}{(w_{i,j}^k)_L}. \tag{12}$$

Moreover, it is expected that the estimated ratio for the target HR image is more dependent on the samples with a larger structural-similarity index s_k. Hence, we establish a penalty function with its coefficients proportional to the value of s_k, as follows:

$$p_k = \frac{s_k}{\sum_{k=1}^{K} s_k + \tau}. \tag{13}$$

Finally, we use a Gaussian distribution to characterize the relation between the reference sample and the target difference ratio $\gamma_{i,j}^t$ expected to be generated as follows:

$$\gamma_{i,j}^t \propto \exp\left(\frac{(\boldsymbol{\gamma}_{i,j})^T \mathbf{P}_\Omega (\boldsymbol{\gamma}_{i,j})}{2\theta^2}\right), \tag{14}$$

where $\boldsymbol{\gamma}_{i,j} = \left[\gamma_{i,j}^1 \ \ldots \ \gamma_{i,j}^K\right]^T$ and $\mathbf{P}_\Omega = \text{diag}\{p_1, ..., p_K\}$, and Ω is the region defined to derive the local prior. The value of $\gamma_{i,j}^t$ is the maximum value obtained via the Gaussian distribution. Then, the refined weights are computed as follows:

$$\hat{w}_{i,j}^t = \gamma_{i,j}^t w_{i,j}^t. \tag{15}$$

2.2.2 Iteration Reconstruction

In the previous section, we have derived the local prior using (15). In this stage, we aim to propagate the local kernel to the initially estimated results so that the distortions in the structure can be compensated. This objective can be expressed as a least-square problem, as follows:

$$\arg\min_{\hat{h}_i^t} \left\| \hat{h}_i^t - \sum_{\hat{h}_j^t \in \Omega} \hat{w}_{i,j}^t \hat{h}_j^t \right\|^2, \tag{16}$$

where $\hat{h}_j^t$ refers to the neighboring patches of the initially estimated patch $\hat{h}_i^t$ in region Ω, and $\hat{w}_{i,j}^t$ is the local prior weights learnt in previous stage. According to damped basic iterative method illustrated in [8], we can approach the least-square problem by simply iterative process to update the value of $\hat{h}_i^t$. The distortions in structure can be compensated as the pixels in a certain local prior corresponding to the initial estimated results are set the same as the local prior extracted in the second stage. The updating rule for structure compensation can be described as follows:

$$d_{i,v} = \hat{h}_{i,v}^t - \sum_{\hat{h}_j^t \in \Omega} \hat{w}_{i,j}^t \hat{h}_{j,v}^t, \text{ and} \tag{17}$$

$$\hat{h}_{i,v+1}^t = \hat{h}_{i,v}^t - \beta d_{i,v}, \tag{18}$$

where $d_{i,v}$ is the difference between the current estimated patch and the best estimated patch in terms of structure information, v is the iteration index, $\hat{h}_{i,0}^t$ is the initially estimated results, and β is the updating step size.

3. EXPERIMENTAL RESULTS

In our experiments, frontal-view face images with neutral expression were selected from the CMU [9] database. 68 distinct facial images were available, and the leave-one-out

strategy was used to evaluate the performance of our proposed method. Based on the framework in [10], the faces are aligned based on the positions of the two eyes. The size of the facial images in the training set is 100×100, and the LR images are obtained through a 7×7 Gaussian kernel and a down-sampling factor of 4 in both the vertical and horizontal directions. The patch size used in the first stage is 10×10 and the patch size used in the second stage is 3×3; the local region selected each time is 9×9. The maximum number of iterations used and the step size β are set at 10 and 0.05, respectively, which are determined by experiments.

In order to evaluate the effectiveness of proposed local-structure kernel, we compare our proposed method with bicubic interpolation method, typical patch-based algorithm [2], as well as with those methods that employ the relations between patches to estimate the HR faces, including the neighbor-embedding algorithm [3] and the position-patch algorithm [5].

Fig. 2 shows the results using the different algorithms. Those algorithms employing local-structure constraints can achieve a better visual performance than typical patch-based method [2]. This is mainly due to the fact that the local kernel learned from our proposed method can more accurately reflect the structural information than using belief propagation only can. Compared with [3] and [5], our proposed method is more robust in terms of retaining the feature shapes and learning new features. The neighbor embedding method will sometimes exhibit artifacts, as the neighbors are defined according to the Euclidean distance. Patches which do not match the local prior may also be identified and embedded into the reconstruction results. In [5], the position patches are not further constrained by adjacent neighbors; this may lead to dissimilarity to the original features. According to the faces shown in the second row, the edges of the nose cannot be well reconstructed using the position-patch-based method. The quantitative measurements for the respective methods are tabulated in Table 1, in terms of both PSNR and SSIM [11].

4. CONCLUSIONS

In this paper, we have proposed a novel face-hallucination method based on local-structure constraints learned from interpolated LR images, as well as the difference between the interpolated LR and HR samples. We define a local region composed of nine patches, and derive the embedding weights for each patch with respect to its neighbors in a region. The embedding weights are applied to the initially estimated results, which are obtained using patch-based eigentransformation. Finally, we use an iterative process to propagate the local kernels learned for the patches to the initially estimated results. Experimental results show that our method can produce a good performance in terms of both visual quality and reconstruction errors.

	PSNR (dB)	SSIM
Bicubic interpolation	24.5329	0.6731
Freeman's method	25.3661	0.6749
Neighbor embedding	28.6860	0.7709
Position patch	27.1966	0.7232
Proposed method	**28.9086**	**0.7803**

Table.1 PSNR and SSIM of the respective methods based on the CMU database.

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

Fig. 2. HR face reconstruction results rendered by different algorithms based on the CMU database: (a) Bicubic interpolation, (b) Freeman's method [2], (c) Neighbor-embedding method [3], (d) Position-patch method [5], (e) Our proposed method, and (f) Ground-true images.

5. REFERENCES

[1] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.,* pp. 1167-1183, 2002.

[2] W. T. Freeman, T. R. Jones and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graphics Appl.,* pp. 56-65, 2002.

[3] W. Fan and D. Y. Yeung, "Image hallucination using neighbor embedding over visual primitive manifolds," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on,* 2007, pp. 1-7.

[4] L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds," *The Journal of Machine Learning Research,* vol. 4, pp. 119-155, 2003.

[5] X. Ma, J. Zhang and C. Qi, "Hallucinating face by position-patch," *Pattern Recognition,* vol. 43, pp. 2224-2236, 2010.

[6] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* vol. 35, pp. 425-434, 2005.

[7] Z. Hui and K. M. Lam, "Two-stage Patch-based Multi-View Face Super-resolution", In Proc. APSIPA Annual Summit Conference (ASC), 2011.

[8] P. Wesseling, "Introduction to multigrid methods," 1995.

[9] T. Sim, S. Baker and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.,* pp. 1615-1618, 2003.

[10] X. Xie and K. M. Lam, "An efficient illumination normalization method for face recognition," *Pattern Recog. Lett.,* vol. 27, pp. 609-617, 2006.

[11] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? A new look at signal fidelity measures," *Signal Processing Magazine, IEEE,* vol. 26, pp. 98-117, 2009.

[12] P. J. Burt and E. H. Adelson, "Merging images through pattern decomposition," in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 1985, pp. 173.