LEARNING ADAPTIVE LOCAL DISTANCE METRIC FOR FACE HALLUCINATION
Yuanpeng Zou, Fei Zhou*, Qingmin Liao

Visual Information Processing Laboratory/Shenzhen Key Laboratory of Science and Technology Department of Electronics Engineering/Graduate School at Shenzhen, Tsinghua University, China

ABSTRACT
In this paper, we propose a novel method for face hallucination by learning a new distance metric in the low-resolution (LR) patch space (source space). Local patch-based face hallucination methods usually assume that the two manifolds formed by LR and high-resolution (HR) image patches have similar local geometry. However, this assumption does not hold well in practice. Motivated by metric learning in machine learning, we propose to learn a new distance metric in the source space, under the supervision of the true local geometry in the target space (HR patch space). The learned new metric gives more freedom to the presentation of local geometry in the source space, and thus the local geometries of source and target space turn to be more consistent. Experiments conducted on two datasets demonstrate that the proposed method is superior to the state-of-the-art face hallucination and image super-resolution (SR) methods.

Index Terms— Face hallucination, face super-resolution, metric learning, position-patch

1. INTRODUCTION

Face image SR is a domain-specific image SR problem, with the goal to generate HR face images from LR inputs. Face image SR algorithms improve the quality of the LR face images, and thus numerous applications in computer vision can be found them, such as face recognition in video surveillance systems [1]. Face image SR can be roughly grouped into two categories: multi-image reconstruction based face image SR [2] and single-image learning based face image SR [3]. In this paper, we focus on the latter one, which is also known as face hallucination [4].

Face hallucination is an ill-posed inverse problem. To address this problem, prior knowledge on face structure and regularizations for the reconstruction have been investigated. In [5], the HR face was constructed by the Principal Component Analysis (PCA) coefficients estimated from the LR subspace. In [6], the face image was decomposed into facial variations and texture in a morphable face model, then each component was separately handled. However, because of the limitation of global linear subspace representation, these global based methods tend to produce smooth results with less details. To improve the results of global based methods, local patch-based methods are proposed. Local patch-based methods cut face image into small local patches, and handle these patches individually. In [7], the patch to be hallucinated was only constructed by the training patches coming from the same position in the face images. Based on the method in [7], a locality constraint term was introduced to regularize the reconstruction weights in [8]. Local patch based methods focus more on the local details of face image, and thus plausible results are produced.

Face hallucination aims to upscale LR face images to HR face images, and it involves two feature spaces with distinct dimensionality: LR face image/patch space (source space) and HR face image/patch space (target space). Local patch based methods usually assume that small image patches in the LR and HR images form manifolds with similar local geometry in these two distinct spaces [9]. Based on this assumption, the input LR patch and the desired HR patch can be represented as a linear combination of its k-nearest neighbors in its own space with the same representation coefficients. However, due to the complexity of degradation process from HR face image to LR face image, this assumption does not hold well in practice [10]. Therefore, if we use the same distance metric (such as Euclidean distance) to measure the local geometry in each space, the estimated representation coefficients in these two spaces will not identical.

Recently, metric learning, which learns a specific distance metric turned to a particular task, has been proved to be very useful in many machine learning tasks [11]. Motivated by the idea of metric learning, we propose to learn a new distance metric for the source space. The learned distance metric gives more freedom to the representation of local geometry in the source space, and thus the local geometry of source and target spaces turn to be more consistent. To put it simply, the local geometry of the target space is "imitated" in the source space with the help of the learned new distance metric.

The remainder of this paper is organized as follows. In Section 2, we review some works related with the method...
proposed in this paper. In Section 3, the proposed method is presented in detail. We conduct experiments in Section 4. Finally, we conclude this paper in Section 5.

2. RELATED WORKS

Our work relates with local patch-based methods for face hallucination and supervised distance metric learning in machine learning. Representative works covering these two issues are briefly reviewed below.

Inspired by the locally linear embedding (LLE) in manifold learning, it was assumed in [9] that the LR patch and its corresponding HR counterpart could be represented by a linear combination of k-nearest neighbors in their own spaces with the same weights. Based on [9], in [7], position prior of face image was employed, the patch to be hallucinated was only constructed by the training patches coming from the same position in face images. Based on the method in [7], the sparse prior of the reconstruction weights was emphasized, and therefore a sparsity constraint was introduced in [12]. However, these works rely on the consistency between source and target space, which does not hold in practice. To address this inconsistency problem, in [10], the source and target space were projected to a middle term space, where the consistency between the source and target space is maximized. However, the local geometry of the target space is not fully preserved in this middle term space. Recently, a multilayer iterative neighbor embedding method was proposed in [13], which gradually updates the training set, to iteratively renew the local geometry of the source space. Nevertheless, the update process is not learnt from external training set, and thus the result of the iteration can not be guaranteed to be highly consistent with the target space.

Supervised metric learning aims to learn a distance metric for a specific task from side supervision information [11]. In [14], a metric was learnt to make the k-nearest neighbors belonging to the same class separated from examples in other classes by a large margin. In [15], a distance metric over cheap features was learnt by the supervision of standard metric over more sophisticated source features. By capturing all pairwise similarities and dissimilarities of training samples, a robust distance metric for person re-identification was learnt in [16]. For details about metric learning, please refer to [11] for a survey.

3. PROPOSED METHOD

In this section, we introduce our method of learning a local adaptive distance metric for face hallucination. We appoint that matrices and vectors are respectively represented in bold capital latters and bold lowerscases. We follow [7] to cut face images into position-patches, \( I_L = \{x_1, x_2, \ldots, x_N\} \) and \( I_H = \{y_1, y_2, \ldots, y_N\} \) are the LR and HR training sets at position index \((i, j)\) (for simplicity, the position index \((i, j)\) for each patch feature is omitted). \( N \) is the size of training set, \( x_1, x_2, \ldots, x_N \) and \( y_1, y_2, \ldots, y_N \) are the LR and corresponding HR patch features.

In this paper, the distance metric to be learnt is set as the Mahalanobis distance which is widely applied in metric learning [11]. For \( x_i, x_j \in \mathbb{R}^N \), their Mahalanobis distance is represented as:

\[
d(x_i, x_j) = \sqrt{(x_i - x_j)^T M (x_i - x_j)},
\]

where \( M \) is an \( N \times N \) positive semi-definite matrix.

3.1. New distance metric for face hallucination

Similar to other local patch-based methods, we use the linear combination of vectors in \( I_L \) to approximate the testing LR patch vector \( x \). However, the approximation is not in the Euclidean distance space, but in a new distance metric space characterized by \( M \). The loss function with respect to reconstruction weights \( \omega \) and \( M \) can be expressed as:

\[
f(M, \omega) = \left( x - \sum_{i=1}^{N} \omega_i x_i \right)^T M \left( x - \sum_{i=1}^{N} \omega_i x_i \right) + \lambda_1 \sum_{i=1}^{N} (x - x_i)^T M (x - x_i) \omega_i^2 + \lambda_2 ||M||_F^2,
\]

\[
s.t. \sum_{i=1}^{N} \omega_i = 1.
\]

The first term is the reconstruction error characterized by the new distance metric, the second term is a locality constraint term which has been used in image classification [17], aiming at emphasizing the locality of the test patch in the new distance metric space, and the third term is Frobenius norm regularization for \( M \). \( \lambda_1, \lambda_2 \) are the regularizations parameters for the second and third terms, respectively. \( \omega = [\omega_1, \omega_2, \ldots, \omega_N]^T \) is an \( N \times 1 \) vector.

3.2. Learning the distance metric \( M \)

In (2), there are two variables to be optimised. To obtain the optimal \( M \), the optimal \( \omega \) in the source space should be known in advance. Since we want that the local geometry of the source and target spaces to be consistent, the optimal reconstruction weights in these two spaces should be identical. Thus the optimal \( \omega \) can be estimated in the target space, where the distance metric is set as the Euclidean distance.

We build another training set \( I_L^t = \{x_1^t, x_2^t, \ldots, x_K^t\} \) and \( I_H^t = \{y_1^t, y_2^t, \ldots, y_K^t\} \) to learn the distance metric \( M \) at position index \((i, j)\). \( K \) is the size of training set, \( x_1^t, x_2^t, \ldots, x_K^t \) and \( y_1^t, y_2^t, \ldots, y_K^t \) are the LR and corresponding HR patch features at position index \((i, j)\). For each
\( y^i \) (\( i = 1, 2, \ldots K \)) in \( T^i \), the optimal reconstruction weights \( \omega^i = [\omega^i_1, \omega^i_2, \ldots \omega^i_N]^T \) can be estimated from:

\[
\omega^i = \arg \min_{\omega^i} \left\{ \left( y^i - \sum_{j=1}^{N} \omega^i_{ij} y_j \right)^T \left( y^i - \sum_{j=1}^{N} \omega^i_{ij} y_j \right) \right\} + \lambda_1 \sum_{j=1}^{N} (y^i_j - y_j)^T (y^i_j - y_j) (\omega^i_{ij})^2; \tag{3}
\]

\[- \sum_{j=1}^{N} \omega^i_{ij} = 1.
\]

The close-form solution of (5) can be expressed as:

\[
M = \frac{1}{K} \sum_{i=1}^{K} M_i, \tag{6}
\]

and \( M_i = \frac{1}{N^2} \left( x^i - \sum_{j=1}^{N} \omega^i_{ij} x_i \right) \left( x^i - \sum_{j=1}^{N} \omega^i_{ij} x_i \right)^T + \sum_{j=1}^{N} (x^i_j - x_i) (x^i_j - x_i)^T (\omega^i_{ij})^2. \tag{7}
\]

The above process shows that the optimal reconstruction weights in the target space can be the supervision to learn the distance metric \( M \) in the source space.

### 3.3. Reconstruction weights with the new distance metric

Once \( M \) is learned, for the test patch \( x \), the optimal reconstruction weights \( \omega \) in the source space can be obtained:

\[
\omega = \arg \min_{\omega} f(M, \omega), s.t. \sum_{i=1}^{N} \omega_i = 1. \tag{7}
\]

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{Some results obtained in the FERET dataset. (a) Bl. (b) MLNE \[13\]. (c) NEFC \[18\]. (d) A+ \[19\]. (e) Proposed method. (f) Groundtruth. (The up-sampling factor is 4)}
\end{figure}

Let \( \frac{\partial f}{\partial \omega} = 0 \), and \( \omega \) has an analytical solution:

\[
\omega = \frac{Q^{-1} \mathbf{1} \mathbf{1}^T Q^{-1}}{\mathbf{1}^T Q^{-1} \mathbf{1}}, \tag{8}
\]

where \( Q = (x \cdot \mathbf{1}^T - X)^T M (x \cdot \mathbf{1}^T - X) + \lambda_1 A \), \( A = diag([x_1, x_2, \ldots x_N], \mathbf{1} = [1, 1, \ldots 1] \), \( x = [x_1, x_2, \ldots x_N] \), \( \mathbf{1} = [1, 1, \ldots 1] \)

We can see that under the supervision of the local geometry in the target space, the local geometry in the source space is imitated by the new distance metric \( M \). Compared with the Euclidean distance, \( M \) gives more freedom to the representation of local geometry in the source space, and thus the consistency between the source and target spaces is improved. Please note that the proposed method is different from methods in \[10\] and \[13\]. In \[10\], the source and target space are projected to a middle term space, but the local geometry in the middle term space still diverges from that in the target space. In \[13\], the source space is gradually updated, to make it more consistent with the target space. However, the result of the iteration can not be guaranteed.

### 4. EXPERIMENTS

In this section we describe the experiments conducted on the FERET dataset \[20\] and FEI dataset \[21\]. All the face images are manually aligned. We compare our method with some state-of-the-art face hallucination and image SR methods, including the Bicubic interpolation (BI), the multilayer iterative
neighbor embedding (MLNE) in [13], and the neighbor embedding for facial components (NEFC) in [18], and the Adjusted Anchored Neighborhood Regression (A+) in [19]. All the LR images are formed by smoothing and down-sampling HR images (the down-sampling factor is 4) in the same way. The LR patch size is $3 \times 3$ and the overlap of the LR patches is 1. We tune the parameters of other methods to get their best results. The objective evaluation criterion is the average PSNR and SSIM [22].

4.1. Experiments in the FERET database

In this experiment, one set containing 600 images at frontal pose is utilized, including 500 for training ($N = 300$, $K = 200$), 50 images for the validation set for parameters settings and 50 images for the test set (there are no overlaps between these sets). The size of the HR face image is $120 \times 120$, $\lambda_1$ and $\lambda_2$ are respectively set as 0.00001 and 0.0001 to get the best results in the validation set. Table 1 shows the numerical results on the test set, and some examples are presented in Fig. 1. We can see that the proposed method achieves the best numerical results, and recovers the most facial details.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>FERET(PSNR/SSIM)</th>
<th>FEI(PSNR/SSIM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>27.78/0.8116</td>
<td>28.29/0.8289</td>
</tr>
<tr>
<td>MLNE [13]</td>
<td>31.81/0.8780</td>
<td>32.71/0.9165</td>
</tr>
<tr>
<td>NEFC [18]</td>
<td>31.96/0.8806</td>
<td>32.78/0.9168</td>
</tr>
<tr>
<td>A+ [19]</td>
<td>30.58/0.8578</td>
<td>31.32/0.8945</td>
</tr>
<tr>
<td>Proposed</td>
<td>32.25/0.8861</td>
<td>33.06/0.9224</td>
</tr>
</tbody>
</table>

and some examples are presented in Fig. 2. We can see that the proposed method also achieves the best numerical results, and recovers the most facial details.

4.2. Experiments in the FEI database

In this experiment, we choose all the 400 images at frontal pose, specifically, $N = 300$, $K = 60$, and 40 images for the testing set (there are no overlaps in these sets). The size of HR face images is $120 \times 100$, and $\lambda_1$ and $\lambda_2$ are set as in Section 4.1. Table 1 shows the numerical results in the testing set.

4.3. Learned distance metric vs Euclidean distance

In this experiment we use the Euclidean distance in the source space, i.e. $M$ is set as $E$ (an identity matrix with the same size as $M$). Experiments are conducted on above two datasets, and the average PSNR (dB) and SSIM of the test set are 31.70/0.8742 and 32.62/0.9112, respectively. Compared with the original results listed in Table 1, the improvement is quite prominent. We can conclude that compared with the Euclidean distance, the learned $M$ indeed makes the source and target spaces more consistent.

4.4. Discussion

A+ handles the face image holistically and considers no face prior, it produces the worst results. MLNE does not involve supervision for the update of the source space, and thus the update can not be guaranteed to the target space. NEFC heavily relies on the assumption of spaces consistency, which is not true in practice. By introducing a new distance metric in the source space, our method deals with the inconsistency problem with supervision from the target space. Therefore, our method produces the best results.

5. CONCLUSION

In this paper, a novel approach for face hallucination by learning a new distance metric in the source space is proposed. With the supervision of the true local geometry in the target space, the new distance metric is learned to imitate the local geometry in the source space. The learned distance metric gives more freedom to represent the local geometry in the source space, and thus the consistency between the source and target spaces is improved. Experiments conducted on two datasets demonstrate that the proposed method is superior to the state-of-the-art face hallucination and image SR methods.
6. REFERENCES


