Face Hallucination by Learning Local Distance Metric

Yuanpeng ZOU†(a), Nonmember, Fei ZHOU†(b), Member, and Qingmin LIAO†(c), Nonmember

SUMMARY In this letter, 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.

key words: face hallucination, face super-resolution, metric learning

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 for them. Face image SR can be roughly grouped into two categories: multi-image reconstruction based face image SR [1] and single-image learning based face image SR [2]. In this letter, we focus on the latter one, which is also known as Face hallucination (FH).

FH is an ill-posed inverse problem. To address this problem, prior knowledge on face structure and regularizations for the reconstruction have been investigated. In [3], the HR face was constructed by the Principal Component Analysis (PCA) coefficients estimated from the LR subspace. However, because of the limitation of global linear subspace representation, the global based method in [3] tend to produce smooth results with less details. To overcome this problem, local patch-based methods are proposed. Face image is cut into small local patches and then these patches are individually handled. In [4], the patch to be hallucinated was only constructed by the training patches coming from the same position in the face images. Local patch based methods focus more on the local details of face image, and thus plausible results are produced.

FH 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 spaces [5]. 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 [6]. If we use the same distance metric (such as Euclidean distance) to measure the local geometry in each space, the desired representation coefficients in these two spaces will not identical.

Recently, metric learning, which learns a specific distance metric for a particular task, has been proved to be very useful in many machine learning tasks [7]. Motivated by the idea of metric learning, we propose to learn a new distance metric in 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 turns to be more consistent.

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

2. Related Works

Inspired by the locally linear embedding (LLE) in manifold learning, it was assumed in [5] that the LR patches and their corresponding HR counterparts could be represented by a linear combination of k-nearest neighbors in their own spaces with the same weights. Based on [5], in [4], position prior of face was employed, the patch to be hallucinated was only constructed by the training patches coming from the same position in face images. In [8], a locality constraint term was introduced to regularize the reconstruction weights. However, these works relay on the consistency between source and target space, which does not hold in practice. To address this inconsistency problem, in [6], 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.
space. Recently, a multi-layer iterative neighbor embedding method was proposed in [9], 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 consistent with the target space.

Supervised metric learning aims to learn a distance metric for a specific task from side supervision information [7]. In [10], a metric was learnt to make the Mahalanobis distance which is widely used in metric learning for a specific task from side supervision information. And then the Mahalanobis distance space, but in a new distance metric space. The loss function with respect to the iteration 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 \( I'_{\ell} = \{x'_1, x'_2, \ldots, x'_K\} \) and \( I'_{\ell} = \{y'_1, y'_2, \ldots, y'_K\} \) to learn the distance metric \( M \) at position index \((i, j)\). \( K \) is the size of training set, \( x'_1, x'_2, \ldots, x'_K \) and \( y'_1, y'_2, \ldots, y'_K \) are the LR and corresponding HR patch features. For each \( y'_i \) \( (i = 1, 2, \ldots, K) \) in \( I'_{\ell} \), the optimal reconstruction weights \( \omega'^{\dagger}_i = [\omega'^{\dagger}_i, \omega'^{\dagger}_i, \ldots, \omega'^{\dagger}_i]^T \) can be obtained:

\[
\omega'^{\dagger}_i = \arg \min_{\omega'} \left\{ \sum_{j=1}^{N} (y'_i - \sum_{j=1}^{N} \omega'_{ij}y_j)^T (y'_i - \sum_{j=1}^{N} \omega'_{ij}y_j) \right\} + \lambda_1 \sum_{j=1}^{N} (y'_i - y_j)^T (y'_i - y_j)(\omega'_{ij})^2; \quad \text{s.t.} \quad \sum_{j=1}^{N} \omega'_{ij} = 1. \quad (2)
\]

The first term in (2) is the reconstruction error, and the second term is the locality-constraint regularization for \( \omega' \). (2) has an analytical solution:

\[
\omega'^{\dagger}_i = \omega'^{\dagger}_i / (1^T \omega'^{\dagger}_i), \quad (3)
\]

where \( \omega'^{\dagger}_i = (C + \lambda_1 \cdot \text{diag}(d))^{-1} \), \( C = (Y - y'_1)^T (Y - y'_1) \), \( Y = [y'_1, y'_2, \ldots, y'_K] \), \( I \) is an \( N \times 1 \) vector with all ones, and \( d = ([y'_1 - y_1]^T, [y'_2 - y_1]^T, [y'_3 - y_2]^T, \ldots, [y'_K - y_K]^T) \).

\( \omega'^{\dagger}_i \) equals to the optimal reconstruction weights in the source space. And then \( M \) can be estimated in \( I_{\ell} \):

\[
M = \frac{1}{K} \arg \min_{M} \sum_{i=1}^{K} f(M, \omega'^{\dagger}_i)). \quad (4)
\]

\[
M = \frac{1}{K} \sum_{i=1}^{K} M_i, \quad \text{and} \quad M_i = \frac{1}{N} \left( \sum_{i=1}^{N} \omega'^{\dagger}_i x_i \right) - y'_i, \quad (5)
\]

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 estimated:

\[
\omega = \arg \min_{\omega} f(M, \omega), \quad \text{s.t.} \quad \sum_{i=1}^{N} \omega_i = 1. \quad (5)
\]
Let \( \frac{\partial g}{\partial a_0} = 0 \), and \( \omega \) has an analytical solution:

\[
\omega = (Q^{-1} \cdot 1) / (1^T \cdot Q^{-1} \cdot 1),
\]

(6)

\[
Q = (x \cdot 1^T - X)^T M (x \cdot 1^T - X) + \lambda I, \quad \Lambda = \text{diag}(x - x_1)^T M (x - x_1),
\]

(7)

(8)

\[ \omega = \sum_{i=1}^{N} \omega_i, \quad \text{and} \quad I = \text{an } N \times 1 \text{ vector with all ones}. \]

The hallucinated HR patch feature \( y \) is represented as:

\[ y = Y \omega. \]

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 [6] and [9]. In [6], 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 [9], the source space is gradually updated, to make it more consistent with the target space. However, the result of the iteration cannot be guaranteed.

4. Experiments

In this section we describe the experiments conducted on the FERET [15] and FEI dataset [16]. All the face images are manually aligned. We compare our method with some state-of-the-art FH and image SR methods, including the Bicubic interpolation (BI), the multilayer iterative neighbor embedding (MLNE) in [9], and the Adjusted Anchored Neighborhood Regression (A+) in [14]. All the LR images are formed by smoothing (an average filter of size 4 \( \times \) 4) and down-sampling (the down-sampling factor is 4) corresponding HR images. The LR patch size is 3 \( \times \) 3 and the overlap of the LR patches is 1. All the experiments are performed using Matlab2014a on a 3.30 GHz Intel i5-2500 CPU with 8GB RAM. We tune the parameters of other methods to get their best results. The objective evaluation criterion is the average PSNR (dB) and SSIM of the hallucinated images.

4.1 Experiments in the FERET Database

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

4.2 Experiments in the FEI Database

In this experiment, we use all the 400 images at frontal pose from the 200 subjects, and randomly choose 360 images (180 subjects) as the training set, and the rest 40 images (20 subjects) as the test 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 Sect. 4.1. It takes 16.41 seconds to hallucinate all the 40 test images. Table 1 shows the numerical results in the testing set, and some examples are shown in Fig. 2. We can see that the proposed method also achieves the best numerical results, and recovers the most facial details.

4.3 Learned Distance Metric vs Euclidean Distance

In this experiment, we describe the experiments conducted on the FERET [15] and FEI dataset [16]. All the face images are manually aligned. We compare our method with some state-of-the-art FH and image SR methods, including the Bicubic interpolation (BI), the multilayer iterative neighbor embedding (MLNE) in [9], and the Adjusted Anchored Neighborhood Regression (A+) in [14]. All the LR images are formed by smoothing (an average filter of size 4 \( \times \) 4) and down-sampling (the down-sampling factor is 4) corresponding HR images. The LR patch size is 3 \( \times \) 3 and the overlap of the LR patches is 1. All the experiments are performed using Matlab2014a on a 3.30 GHz Intel i5-2500 CPU with 8GB RAM. We tune the parameters of other methods to get their best results. The objective evaluation criterion is the average PSNR (dB) and SSIM of the hallucinated images.

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 source space update, and thus the update can not be guaranteed to the target space. NEFC heavily relays 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 letter, a novel approach for FH 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 new 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 FH and image SR methods.

Acknowledgements

The authors would like to thank the reviewers for their comments. This work was supported by the National Natural Science Foundation of China under Grant 61271393 and 61301183, by the China Post-Doctoral Science Foundation under Grant 2013M540947 and 2014T70083, and by the Special Foundation for the Development of Strategic Emerging Industries of Shenzhen under Grant JCYJ20140107155840272, JCYJ20150331151358138 and ZDSYS20140509172959974.

References

[1] Y. Jin and C.-S. Bouganis, “Robust multi-image based blind face hallucination,” CVPR, pp.5252–5260, 2015.
[2] Z. Wang, R. Hu, S. Wang, and J. Jiang, “Face hallucination via weighted adaptive sparse regularization,” IEEE Trans. Circuits Syst. Video Technol., vol.24, no.5, pp.802–813, May 2014.
[3] X. Wang and X. Tang, “Hallucinating face by eigentransformation,” IEEE Trans. Syst., Man, Cybern., vol.35, no.3, pp.425–434, Aug. 2005.
[4] X. Ma, J. Zhang, and C. Qi, “Hallucinating face by position-patch,” PR, vol.43, no.6, pp.2224–2236, 2010.
[5] H. Chang, D.-Y. Yeung, and Y. Xiong, “Super-resolution through neighbor embedding,” CVPR, pp.275–282, 2004.
[6] Z. Zhang, C. Qi, and Y. Hao, “Locality preserving partial least squares for neighbor embedding-based face hallucination,” ICIP, pp.409–413, 2015.
[7] B. Kulis, “Metric learning: A survey,” Foundations and Trends in Machine Learning, vol.5, no.4, pp.287–364, 2013.
[8] J. Jiang, R. Hu, Z. Wang, and Z. Han, “Noise robust face hallucination via locality-constrained representation,” IEEE Trans. Multimedia, vol.16, no.5, pp.1268–1281, Aug 2014.
[9] J. Jiang, R. Hu, Z. Wang, and Z. Han, “Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning,” IEEE Trans. Image Process., vol.23, no.10, pp.4220–4231, Oct. 2014.
[10] K.Q. Weinberger and L.K. Saul, “Distance metric learning for large margin nearest neighbor classification,” JMLR, vol.10, pp.207–244, June 2009.
[11] D. Dai, T. Kroeger, R. Timofte, and L.V. Gool, “Metric imitation by manifold transfer for efficient vision applications,” CVPR, pp.3527–3536, 2015.
[12] J. Wang, J. Yang, K. Xu, F. Lv, T. Huang, and Y. Gong, “Locality-constrained linear coding for image classification,” CVPR, pp.3360–3367, 2010.
[13] Y. Li, J. Liu, W. Yang, and Z. Guo, “Multi-pose face hallucination via neighbor embedding for facial components,” ICIP, pp.4037–4041, 2015.
[14] R. Timofte, V.D. Smet, and L.V. Gool, “A+: adjusted anchored neighborhood regression for fast super-resolution,” ACCV, vol.9006, pp.111–126, 2014.
[15] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, “The feret evaluation methodology for face-recognition algorithms,” IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.10, pp.1090–1104, Oct. 2000.
[16] C.E. Thomaz and G.A. Giraldi, “A new ranking method for principal components analysis and its application to face image analysis,” IVC, vol.28, no.6, pp.902–913, 2010.