FACE HALLUCINATION BASED ON NONPARAMETRIC BAYESIAN LEARNING

Minqi Li, Richard Yi Da Xu, Xiangjian He

Faculty of Engineering and Information Technology
University of Technology, Sydney

ABSTRACT

In this paper, we propose a novel example-based face hallucination method through nonparametric Bayesian learning based on the assumption that human faces have similar local pixel structure. We cluster the low resolution (LR) face image patches by nonparametric method distance dependent Chinese Restaurant process (ddCRP) and calculate the centres of the clusters (i.e., subspaces). Then, we learn the mapping coefficients from the LR patches to high resolution (HR) patches in each subspace. Finally, the HR patches of input low resolution face image can be efficiently generated by a simple linear regression. The spatial distance constraint is employed to aid the learning of subspace centers so that every subspace will better reflect the detailed information of image patches. Experimental results show our method is efficient and promising for face hallucination.

Index Terms— Face hallucination, ddCRP, nonparametric Bayesian

1. INTRODUCTION

Face image-based application has been well developed and investigated in recent years. It is widely used in many areas such as face recognition, video surveillance, facial expression estimation, image enhancement, image compression and so on. However, due to the limitations of environment and capturing systems, people often obtain low resolution human face images. The quality of face images will further affect the performance of many computer vision and pattern recognition applications. To solve the problem, it is significant to render a high resolution face image from the low resolution one. These techniques are named face hallucination or face super-resolution [1, 2].

The difference between face hallucination and general super-resolution problem is that the face images have regular structures and textures. Compare with general super-resolution problem, face hallucination is challenging because people are so sensitive to the changes of the look of a human face. Small deviations might significantly affect human perception, whereas for super-resolution of generic images such as, buildings, plants, the errors can be more tolerant [3]. Another challenge of hallucinating faces is from the complex application conditions such as variances in illuminations, poses, or views and difficulty of aligning faces at low resolution images [1, 3].

Different methods have been researched to face hallucination. One simple way is through an interpolation using a base function or through interpolating a kernel function to obtain a much higher density of pixels in a processed image [4]. Because of the simplicity of interpolation, it is only applied in these applications with low requirements. However, using this parametric method it is often difficult to interpolate details well within texture and corner-like local regions of intensities [5, 3].

Example-based super-resolution (SR) schemes have proven to be able to reconstruct significantly finer details of a low resolution image compared to interpolation-based schemes [5]. The general idea of example-based approach is to construct a database, learn the statistical correlation between the low resolution and high resolution images and apply it to the input image [3].

Different methods have been studied to learn the mapping relationship between LR and HR images such as [1, 6],

1) Sparse representation-based approaches;

2) Subspace learning approaches: a) locally linear embedding, and b) Linear Subspace Learning-based Approaches;

3) Bayesian inference method: learning priors from numerous feature vectors to generate a function mapping features from LR images to HR images.

Markov random fields can be used to reduce the ambiguity problem between LR and HR patches by minimizing the difference of overlapping HR patches. However, all methods require computationally expensive processes in extracting complex features or searching exemplars [1]. In [7], the authors propose a divide-and-conquer algorithm using k-means to learn K clusters. Each cluster can be viewed as a set of anchor points to represent the feature space of natural image patches for super resolution. For different dataset or applications, K should be set differently. The training can be done off-line and the reconstruction is built by linear regression which is efficient.

Inspired by [7], we develop a similar frame work by involving the similar structure of face images applied in face hallucination. We novelly adapt a nonparametric Bayesian method ddCRP to model the spatial dependencies among

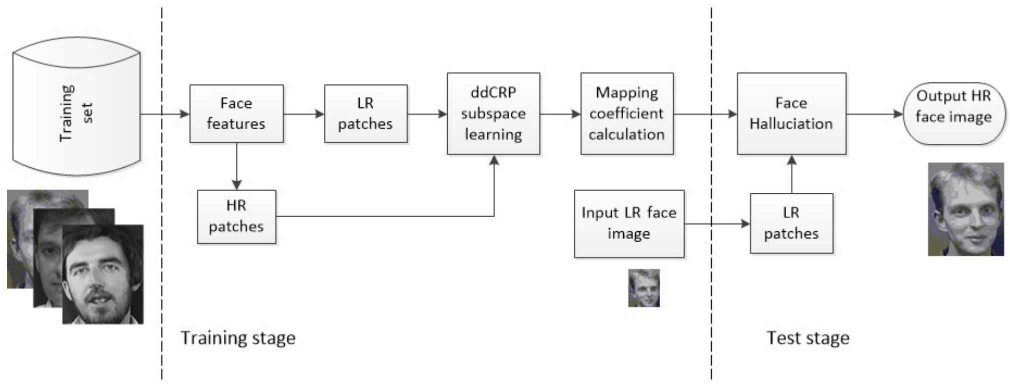

Fig. 1. The framework of our proposed method.

the image patches and inference the subspace of human face patches. The spacial constraint CRP gives us more details in the final reconstructed HR image as face images characterized similar spatial structures.

The remainder of this paper is organized as follows. Section 2 introduces the background of ddCRP and the details of our proposed method. In section 3, experimental results demonstrate the performance of our method. In section 4, we draw a conclusion and discuss the relationship between the constraint of distance dependence in ddCRP to MRF.

2. PROPOSED ALGORITHM

2.1. Background of ddCRP

Dirichlet process (DP) as nonparametric Bayesian method provides a valuable suite of flexible clustering algorithms for high dimensional data analysis. It has been extensively used in computer vision and pattern recognition areas such as image segmentation, text modelling, computational biology and so on [8, 9, 10, 11]. DP mixtures can be described via the Chinese restaurant process (CRP). The probability of a customer sitting at a table is computed from the number of other customers already sitting at that table. Despite the success of the traditional CRP, it ignores the spatial distance between data elements. Distance dependent Chinese restaurant process (ddCRP) which is the random seating assignment of the customers depends on the distances between the data elements. It provides a new tool for flexible clustering of non-exchangeable data.

The ddCRP alters the CRP by modeling customer links not to tables, but to other customers. Given a decay function f, sequential distances D, scaling parameter α, the link can be independently draws the customer assignments conditioned on the distance measurements by the distribution [8]

$$p(c_j = j) = \begin{cases} f(d_{ij}) & \text{if } i \geq j \\ \alpha & \text{otherwise} \end{cases} \tag{1}$$

In our application, d_{ij} is an externally specified distance between image patches i and j. α determines the probability that a patch links to themselves rather than another patch. The decay function f decides how the distance between two patches affects their probability of connecting to each other.

Details of the ddCRP can be found in [8].

2.2. Proposed method

Figure 1 shows the framework of the proposed face hallucination method. Firstly, we collect a large set of LR and HR paired patches from the training set of human face images. The intensity values in the set of LR patches minus their means are used as features for learning subspace centers using the ddCRP method. We learn the mapping functions from the LR patches to corresponding HR patches in each cluster. Here, we refer each cluster as a subspace of face image patches. After the mapping coefficients of each cluster are learned, in the test stage, HR patches of input low resolution face image can be generated by simple linear regression. Finally, the HR face image is reconstructed from the HR patches added by the LR mean.

At the training stage, given a set of high resolution face images, the corresponding low resolution image is generated by a Gaussian kernel and a down scaling factor [12, 7]. We randomly extract a large set of HR and LR patches from the HR and LR image pairs.

We expect that the patches in the same cluster have similar distribution and are constrained by their positions in the images. In ddCRP, two patches are placed in the same table if one can reach the other by traversing the customer assignments. The full generative process for the observed patches $x_{l_{1:N}}$ is described as follows:

1 For each customer/patch, $i \in \{1, ..., N\}$, sample the assignments $c_i \sim ddCRP$. This determines the table/cluster assignments z.

2 For each table/cluster, $k \in \{1, ...\}$, sample parameter $\phi \sim G_0$.

3 For each customer/patch, $i \in \{1, ..., N\}$, assign the patch $\phi_{z(c_i)}$.

We place a conjugate normal-inverse-Wishart prior on the patches distribution G_0.

The fully conjugate prior density is

$$\begin{aligned} p(\mu, \mathbf{\Sigma}) =& NIW(\mu, \mathbf{\Sigma}|\mathbf{m_0}, \kappa_0, \nu_0, \mathbf{S_0}) \\ =& \mathcal{N}(\mu|\mathbf{m_0}, \frac{1}{\kappa_0}\mathbf{\Sigma}) \cdot IW(\mathbf{\Sigma}|\mathbf{S_0}, \nu_0) \end{aligned} \tag{2}$$

where μ and Σ are the mean and covariance matrix of a multivariate Gaussian, $\mathbf{m_0}, \kappa_0, \nu_0, \mathbf{S_0}$ are hyper-parameters.

The assignment of each patch can be described as

$$p(c_i|\mathbf{c_{-i}}, \mathbf{x}, D, f, \alpha, \mu, \mathbf{\Sigma}) \propto \begin{cases} p(c_i|\mathbf{x}, D, f, \alpha)\Delta(\mathbf{x}, z, \mu, \mathbf{\Sigma}) \\ \quad \text{if } c_i \text{ links } k_1 \text{ and } k_2 \\ p(c_i|D, f, \alpha) \quad \text{otherwise} \end{cases} \tag{3}$$

where

$$\Delta(\mathbf{x}, z, \mu, \mathbf{\Sigma}) = \frac{p(x_{z(c_1:N)=k}, \mu, \Sigma)}{p(x_{z(c_1:N)=k_1}, \mu, \Sigma)p(x_{z(c_1:N)=k_2}, \mu, \Sigma)} \tag{4}$$

Here, k_1 and k_2 are patches in $z(c_{-i})$, and $k = k_1 \bigcup k_2$.

This is the posterior of the spatial ddCRP defined by formula (1). We approximate this posterior using a Gibbs sampler, which iteratively draws c_i from the conditional distribution.

The marginal likelihood of data is [13].

$$\begin{aligned} p(\mathbf{x}) =& \int_\mu \int_\Sigma p(\mathbf{x}|\mu, \mathbf{\Sigma})p(\mu, \mathbf{\Sigma})d\mu d\Sigma \\ =& \frac{(2\pi)^{-ND/2}}{Z_{NIW}(D, \kappa_0, \nu_0, \mathbf{S}_0)}|\Sigma|^{-\frac{\nu_0+N+D+2}{2}} \\ & (2\pi)^{-ND/2}\frac{Z_{NIW}(D, \kappa_N, \nu_N, \mathbf{S}_N)}{Z_{NIW}(D, \kappa_0, \nu_0, \mathbf{S}_0)} \\ =& (\pi)^{-ND/2}\frac{\kappa_0^{D/2}|S_0|^{\nu_0/2}}{\kappa_N^{D/2}|S_N|^{\nu_N/2}}\prod_{i=1}^{D}\frac{\Gamma(\frac{\nu_N+1-i}{2})}{\Gamma(\frac{\nu_0+1-i}{2})} \end{aligned} \tag{5}$$

where

$$\begin{aligned} Z_{NIW}(D, \kappa_0, \nu_0, \mathbf{S}_0) =& 2^{\frac{(\nu_0+1)D}{2}}\pi^{D(D+1)/4}\kappa_0^{-D/2}|S_0|^{-\nu_0/2} \\ & \prod_{i=1}^{D}\Gamma(\frac{\nu_0+1-i}{2}) \end{aligned}$$

After the inference of ddCRP, K cluster centres can be calculated by the means of patches in the same clusters which viewed as subspaces of the training patches. We discard the clusters with a small number of patches as these clusters present very rare features in the subspace of image patches, and learn regression coefficients to predict HR patches from LR patches. We learn a set of linear regression functions to individually predict the n feature values in HR.

Suppose there are l_i LR patches in the same cluster centre K_i. Let $L_i \in R^{m \times l_i}$ and $H_i \in R^{n \times l_i}$, $(i = 1, , l_i)$ be vectorized features of the LR patches and central region of HR patches. m and n are dimensions of LR and HR patches.

The regression coefficients are computed by solving the linear least square problem represented by [7]:

$$\widetilde{C} = \arg\min_C \left\| H_i - C\begin{pmatrix} L_i \\ 1 \end{pmatrix} \right\|^2 \tag{6}$$

At the test stage, given a LR image, we crop each LR patch to compute the LR features and search for the closest cluster centre. The predicted HR patch is then reconstructed through linear regression of the learned mapping coefficients and by adding the LR patch mean to the HR features.

$$H_i = \widetilde{C}\begin{pmatrix} L_i \\ 1 \end{pmatrix} \tag{7}$$

Finally, we get the high resolution face image by averaging the summary of the high patches.

3. EXPERIMENT

3.1. Implementation

In our experiment, the training set contains 200 face images which are selected from the ORL database [14]. The remaining images are used for testing. Without loss of generality, we will magnify the input face image by a factor of 3. The images collected from ORL database are human faces of 40 people with different pose. 10000 patches are generated in the LR and HR training set. As suggested by [7], the LR patch size is 7*7 pixels and HR patch is set to contain 11*11 pixels.

The hyperparameters that regularize our ddCRP prior can be specified based on the properties of the face image patches and training set. It is sensitive to the results. We use Euclidean distance for matrix D shown in equation (2). The decay function f is set to 1 when the distance is larger than 1.5 pixels. The self-connection parameter $\alpha = 10^{-6}$. For the hyper-parameter of the normal-inverse-Wishart prior $NIW(\mu, \mathbf{\Sigma}|\mathbf{m_0}, \kappa_0, \nu_0, \mathbf{S_0})$, we set the degree freedom $\nu_0 = 50$, a value which makes the prior variance nearly as large as possible while ensuring that the mean remains finite. In the experiment, we ran the sampler for 100 times and get 16 clusters with more than 100 patches finally. The converging time depends on the distance defined previously and the size of the training set.

3.2. Performance

To estimate the results, we compare our method with some typical face hallucination and super-resolution methods. We use PSNR as the objective measurement of image quality.

Figure 2 presents the face hallucination results of 4 face images by different methods. The PSNR results are showed in Table 1. The second column presents the results of Bicubic interpolation. Column (c) shows Chang's results using locally linear embedding (LLE) method [6] and Yang's method [15] by sparse representation is presented in column (d). We use

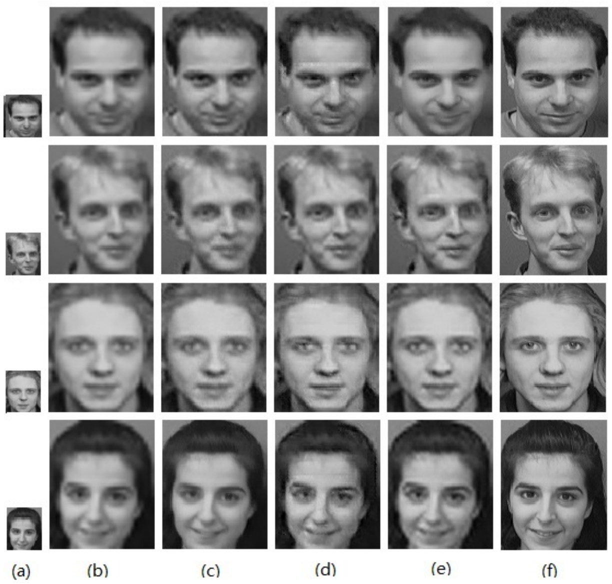

Fig. 2. Face hallucination results by different methods: (a) the input LR faces, (b) the bicubic interpolation, (c) Chang's method, (d) Yang's method, (e) our proposed method, (f) the original HR faces.

Images	Bicubic	Chang's method	Yang's method	proposed method
1	28.5468	29.0900	27.2338	29.3940
2	28.3883	28.1526	28.5304	29.3444
3	28.0830	28.1530	28.2230	28.8568
4	27.3925	28.8511	28.0983	28.8065

Table 1. Performance of different algorithms: PSNR(dB)

10000 patches to generate 128 codesize dictionary by Yang's method and the number of nearest neighbors is 5 for Chang's method.

4. CONCLUSION AND DISCUSSION

In this paper, we present a novel example-based face hallucination method based on nonparametric subspace learning. According to the assumption that face images have similar local pixel structures, we cluster the LR face patches by dd-CRP and learning the mapping coefficients for each cluster. We use the LR input face to search a database for the example faces that are most similar to the input. The HR image is reconstructed from a linear regression. Experimental results show that the proposed method performs well in terms of both reconstruction error and visual quality.

The spatial distance constraint is employed to aid the learning of subspace centers. This constraint here between the patches can be considered as an extension of Markov random fields constraint. MRF is usually applied on the neighborhood of patches in the same single image. However, spatial distance constraint can be applied on the patches from different images. Therefore, this is a more flexible constraint to describe the neighborhood of face image pixels.

5. REFERENCES

[1] Nannan Wang, Dacheng Tao, Xinbo Gao, Xuelong Li, and Jie Li, "A comprehensive survey to face hallucination," *International Journal of Computer Vision*, vol. 106, no. 1, pp. 9–30, 2014.

[2] Xiaoguang Li, Qing Xia, Li Zhuo, and Kin Man Lam, "A face hallucination algorithm via kpls-eigentransformation model," in *Signal Processing, Communication and Computing (ICSPCC), 2012 IEEE International Conference on*. IEEE, 2012, pp. 462–467.

[3] Ce Liu, Heung-Yeung Shum, and William T Freeman, "Face hallucination: Theory and practice," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 115–134, 2007.

[4] Kaibing Zhang, Xinbo Gao, Xuelong Li, and Dacheng Tao, "Partially supervised neighbor embedding for example-based image super-resolution," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 2, pp. 230–239, 2011.

[5] Chih-Chung Hsu, Chia-Wen Lin, Chiou-Ting Hsu, H-YM Liao, and Jen-Yu Yu, "Face hallucination using bayesian global estimation and local basis selection," in *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*. IEEE, 2010, pp. 449–453.

[6] Hong Chang, Dit-Yan Yeung, and Yimin Xiong, "Super-resolution through neighbor embedding," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. IEEE, 2004, vol. 1, pp. I–I.

[7] Chih-Yuan Yang and Ming-Hsuan Yang, "Fast direct super-resolution by simple functions," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 561–568.

[8] David M Blei and Peter I Frazier, "Distance dependent chinese restaurant processes," *The Journal of Machine Learning Research*, vol. 12, pp. 2461–2488, 2011.

[9] Soumya Ghosh, Matthew Loper, Erik B Sudderth, and Michael J Black, "From deformations to parts: Motion-based segmentation of 3d objects," in *Advances in Neural Information Processing Systems*, 2012, pp. 1997–2005.

[10] Soumya Ghosh, Andrei B Ungureanu, Erik B Sudderth, and David M Blei, "Spatial distance dependent chinese restaurant processes for image segmentation," in *Advances in Neural Information Processing Systems*, 2011, pp. 1476–1484.

[11] Peter Orbanz and Joachim M Buhmann, "Nonparametric bayesian image segmentation," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 25–45, 2008.

[12] Michael Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*, Springer, 2010.

[13] Herman Kamper, "Gibbs sampling for fitting finite and infinite gaussian mixture models," 2013.

[14] Ferdinando S Samaria and Andy C Harter, "Parameterisation of a stochastic model for human face identification," in *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*. IEEE, 1994, pp. 138–142.

[15] Jianchao Yang, Hao Tang, Yi Ma, and Thomas Huang, "Face hallucination via sparse coding," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 1264–1267.