
TechS Vidya e-Journal of Research
ISSN 2322 - 0791
Vol. 3 (2014-15) pp. 37-45.

Learning based face hallucination techniques: A survey

Premitha Premnath K

Department of Computer Science & Engineering
Vidya Academy of Science & Technology
Thrissur - 680501, Kerala, India
(email: premithakpnath@gmail.com)

Abstract: Face hallucination is the technique of inferring high resolution face image from low resolution face image. Face hallucination technique can be generally categorized into three types: interpolation based methods, reconstruction based methods and learning based methods. Recently learning based technique has gained a great deal of popularity. In this paper some of the learning based techniques - eigen transformation, neighbor embedding, discrete cosine transform, least square representation and sparse representation - are discussed. These methods are also compared based on the performance measures PSNR and SSIM.

Keywords: Super resolution, Face hallucination, Position patch, Learning based technique

1 Introduction

Super resolution is the process in which a high resolution (HR) image is inferred from a low resolution (LR) image. Face hallucination refers to developing a high resolution face image from a low resolution face image. It can have many applications in image enhancement, image compression and face recognition. A face super resolution method was first developed by Simon Baker [1] and proposed the term "*Face Hallucination*". In this method he used a Bayesian formulation and the high frequency detail of the image was extracted from the parental structure through training samples. A number of other methods for face hallucination have evolved ever since.

Face hallucination can be obtained from a single low resolution image or multiple frames of the same scene. The face hallucination algorithms can be generally classified into three types:

Published on 31 March 2015.

1. Interpolation based algorithms These methods have blurring problems especially when the input image size is very low.

2. Reconstruction based methods that model the relationship between the low resolution and high resolution images. The model will be based on the reconstruction constraints and smoothness constraints. These methods are limited by the number of input images and will not work well for a single image face hallucination.

3. Learning based method - In this method prior information is required. Learning based methods are implemented with the help of some training set from which the necessary informations are extracted. This method can provide better results and higher magnification than interpolation based methods and reconstruction based methods. The method used by Baker was a learning based algorithm that used a training set to develop the high resolution image.

2 Different learning based face hallucination techniques

Now-a-days, learning based techniques have gained more popularity. A number of learning based face hallucination techniques has been developed. Learning based face hallucination methods usually involve two steps: *global parametric modeling* and *local non-parametric modeling*. First step is to develop a global face image that models the main characteristics of a face image and the second step develops a residual image that adds the high frequency details to the previous result. The global parametric model and the local non-parametric model are integrated to obtain the final high resolution image. C. Liu and Freeman [2] proposed one such method in which the global model was developed using *principal component analysis* (PCA) and the local model was developed using *Markov random field* (MRF). Here two steps are required since some details are missed during the first step and so the second step was required to compensate that. This can be avoided and face hallucination can be done in a single step without loss of detailed information. One method for achieving this is to divide the image into different patches.

Neighbor embedding (NE) [4] is one such method which is an improved version of *locally linear embedding* (LLE) [19]. Here the local geometry of low resolution image patch space is mapped to the high resolution image patch space to generate the high resolution image patches. In NE, a fixed number of neighbors are selected for reconstruction. NE was extended by adaptively selecting the most relevant neighbor based on sparse coding. This was introduced by J Yang [8] which minimized the reconstruction error. It fails to make use of the prior knowledge and has limited subjective visual effects.

Position patch was evolved after that which gave importance to the position of the particular patch in the image. Position patches are the patches in all the training images that have the same position with certain patch in the face image input. Several methods had evolved that was based on the concept of position patch. *Hallucinating face by position patch* by Xiang Ma [7], is one such method using *least square representation* (LSR). LSR

is more efficient than other manifold learning based method as it is using the position information. The solution obtained through LSR becomes biased as the number of training samples is much larger than the dimension of the patch. Then *sparse representation* (SR) was developed by C Jung[9] combining both sparse coding and LSR. In this method more emphasis was given to sparsity rather than locality.

Some of the techniques like eigen transformation [3], neighbor embedding (NE) [4], DCT [15], least square representation (LSR) [7], and sparse representation (SR) [9], are discussed in the this section.

2.1 Eigen transformation

In this technique, face hallucination is considered as the transformation between two different image styles. Transform of face images from low-resolution to high-resolution is done based on mapping between two groups of training samples without deriving the transformation function. The hallucinated face image is rendered from the linear combination of training samples. Using a small training set, the method can produce satisfactory results. In this technique, PCA is used to represent the structural similarity of face images. In the PCA representation, different frequency components are uncorrelated. By selecting the number of eigen faces, the maximum amount of facial information can be extracted from the low-resolution face image and the noise can be removed. *Hallucinating face by eigen transformation* [3] has been deployed using eigen transformation.

The hallucination problem can be discussed under the framework of multiresolution analysis. A process of iterative smoothing and down sampling decomposes the face image into different bands, B0, . . .,BK. The low-frequency component is encoded in the down-sampled low-resolution image, and the difference between the original face image and the smoothed image contains the high-frequency detail. In this decomposition, different frequency bands are not independent. Some components of the high-frequency bands B0, . . . , BK can be inferred from the low-frequency band B0. This is a starting point for hallucination. A better way to address the dependency is using PCA, in which different frequency components are uncorrelated.

A face image can be reconstructed from eigenfaces in the PCA representation. Like the multiresolution analysis, PCA also decomposes the face image into different frequency components. The difference is that the PCA method utilizes the face distribution to decompose face structure into uncorrelated frequency components; thus, it can encode face information more concisely. In [3] PCA is implemented first to extract as much useful information as possible from a low-resolution face image, and then renders a high-resolution face image by eigen transformation.

PCA represents face images using a weighted combination of eigen faces. A set of eigen vectors, also called eigen faces, are computed from the eigen vectors. Eigenfaces with large eigen values are “face-like,” and characterize the low-frequency components. Eigen faces with small eigen values are “noise-like,” and characterize the high-frequency details. PCA is optimal for the face representation because the K-largest eigen faces account for most of the energy and are most informative for the face image set. The eigen facree number K

controls the detail level of the reconstructed face. As K increases, more details are added to the reconstructed face.

For eigen transformation, we use a training set containing low-resolution face images, and the corresponding high-resolution face images. Input low-resolution face image can be reconstructed from the optimal linear combination of the M low-resolution training face images. A weight is assigned to each training face image which represents its contribution to the reconstruction of the input image. The sample face that is more similar to the input face has a greater weight contribution.

When a low-resolution image x_l is input, it is approximated by a linear combination of the low-resolution images using the PCA method, and we get a set of coefficients $[c_1, c_2, \ldots, c_M]^T$ on the training set. Keeping the coefficients and replacing the low-resolution training images with the corresponding high-resolution ones, a new high-resolution face image can be synthesized. The synthesized face image is projected onto the high-resolution eigenfaces and reconstructed with constraints on the principal components. This transformation procedure is called *eigen transformation*, since it uses the eigenfaces to transform the input image to the output result. By adjusting K, the eigen transformation method can control noise distortion.

2.2 Neighbor embedding

In [4], neighbor embedding (NE) has been used. This has been inspired from *locally linear embedding* (LLE). It computes the low dimensional, neighbor preserving embedding of high dimensional inputs and recovers the global non-linear structure from locally linear fits. In LLE first a set of K nearest neighbors are computed and then the reconstruction weights of neighbors are calculated so as to minimize the reconstruction errors.

In this method, YIQ color model has been used where the Y channel represents luminance and I, Q channels represent chromaticity. Conversion between the RGB and YIQ color schemes can be done easily via a linear transformation. From this the luminance values are used to define features. For the low-resolution images, the features can be represented by considering the relative luminance changes within a patch. By this way a relatively small training set can be used. For the high-resolution images, the features are defined for each patch based only on the luminance values of the pixels in the patch.

While using LLE local geometry of high dimensional data is preserved in the embedded space. If the database is unevenly sampled, the K nearest neighbors - that are selected based on the Euclidean distance - will be from a single direction. This will result in information redundancy in that single direction and at the same time no information will be captured from other directions. LLE is also very sensitive to noise.

2.3 Discrete cosine transform

Hallucinating face in the DCT domain [15] is built in the discrete cosine transform (DCT) domain. This model is based upon the DCT coefficient which has two parts - DC coefficient estimation and AC coefficient inference. DC coefficients represent the average pixel

intensity of the target blocks. It can be estimated accurately by interpolation-based methods such as bilinear and cubic B-spline. AC coefficient contains the information of local features such as edges and corners around eyes, mouth of face image. It can not be estimated well by interpolation. In [15] only the local facial features embodied in AC coefficients are considered. By this way a more specific and efficient training set for AC coefficients can be built and used. As the DC coefficients are not taken into consideration, the learning process will be more robust since it is much less influenced by image illumination. In order to reduce the redundancy of the training set a compact block dictionary is built by a clustering-based training scheme.

Some of the advantages of deploying face hallucination in the DCT domain are quoted below:

- The DC coefficient which represents the average pixel intensity of a target block can be estimated accurately by a simple interpolation-based method such as cubic B-spline.
- It only needs to focus on building a specific learning-based inference model for low frequency AC coefficients which correspond to the local details of face image such as the edges, corners around eyes.
- A simplified learning-based inference model can be developed to infer the AC coefficients efficiently. This is based on the assumption that blocks of the prefiltered HRI built in the DCT domain are independent with each other.
- The data dimension of training and testing set can be reduced significantly.

The method used in [15] can be divided into two steps. First, the prefiltered HRI, is inferred in the DCT domain, which includes AC coefficient inference by learning and DC coefficient estimation by interpolation. Second, the final hallucinated result IH* is reconstructed from the prefiltered result IH by post filtering.

2.4 Least square representation

Least Square Representation (LSR) has been deployed in *hallucinating face by position patch* [7]. Usually learning based techniques involve two steps: First step produces the high level features and in the second step residue compensation is done. Different from the usual face hallucination, this technique does not incorporate dimensionality reduction methods into process and the residue compensation is no longer necessary because the non-feature information is reserved and the reconstruction coefficients contain both feature information and non-feature information. Patch position in the face image and image features are used to synthesize a high-resolution face image from a low resolution image. The image position-patches are used to hallucinate the high-resolution image. A one-step face hallucination based on position-patch instead of on a complicated probabilistic or manifold learning model has been used in this technique. Position-patches are defined as the patches in all training images that have the same position with certain patch in the face image input.

In this method residue compensation is not implemented and this is not affecting the clarity of the image because the position patches taken gives all the required non parametric informations. By this way the local non parametric modeling can be removed from the face hallucination. The residue compensation phase will become indispensable in the case where the whole image is considered as a single patch.

2.5 Sparse representation

The technique *sparse representation* (SR) has been derived from the idea of position patch based face hallucination technique. *Position patch based face hallucination using convex optimization* [9] has deployed this technique. In this method optimal weight is calculated in a different way from that used by X Ma et al. [7]. In [7] optimal weight vector is being calculated using the least square optimization where as in [9] sparse representation has been used for deriving the optimal weight vector. In [9] L^2 based optimization has been used where as in [7] L^1 based optimization is used, as each patch can be approximated with a smaller subset of patches than L^2-norm. L^2-norm provides nonzero weights for all patches. Thus, the SR [9] method can produce more stable face hallucination results when the number of the training position-patches is much larger than the dimension of the patch.

In LSR, dimension of the patch should be larger than the number of the training position patches. If this condition is not satisfied, least square estimation will produce biased solutions. Sparse representation has been developed in order to eliminate this problem. By compressed sensing theory, the training position-patch set can be regarded as an over complete dictionary with position-patches as its base elements. The over complete dictionary represents the testing LR patches. This representation is naturally sparse if the size of the training position-patches is quite large. By this way a more stable reconstruction weight $w(i, j)$ for face hallucination can be obtained. Jung et al. [9] obtained the optimal weight by solving convex optimization problem. The HR patch corresponding to the input LR patch is obtained by applying this optimal weight to all the corresponding position patches available in the training set. Finally the output HR face image is obtained by integrating all obtained HR patches.

3 Performance comparisons

The performances of different methods are measured using PSNR and SSIM. *Peak signal to noise ratio* (PSNR) is an expression for the ratio between the maximum possible value (power) and the power of distorting noise that affects the quality of its representation. It is used for comparing different image enhancement techniques on the same set of images, to identify whether a particular technique produces a better result. The higher the PSNR, the better degraded image has been reconstructed to match the original image and the better the reconstructive algorithm. The *structural similarity* (SSIM) index is a method for measuring the similarity between two images. The SSIM index can be viewed as a quality measure of one of the images provided the other image is regarded as of perfect quality.

Factors	**4**	**4**	**8**	**8**	**16**	**16**
Methods	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Eigen[3]	27.75	0.7582	26.05	0.7297	24.16	0.6954
NE[4]	31.23	0.8975	27.75	0.8088	24.60	0.7283
DCT[15]	30.55	0.9011	26.55	0.7881	21.79	0.6327
LSR[7]	31.90	0.9032	26.88	0.7813	23.94	0.7063
SR[9]	32.11	0.9048	26.88	0.7814	23.90	0.6993

Table 1: PSNR (dB) and SSIM comparison of different methods with different down sampling factors

Patch size	**Overlap**	**NE[4]**		**LSR[7]**		**SR[9]**	
		PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
4	0	28.80	0.8140	28.28	0.8042	27.91	0.7991
8	0	31.36	0.8838	30.58	0.8715	30.33	0.8661
	4	32.17	0.9044	31.18	0.8918	31.23	0.8912
12	0	30.03	0.8556	31.40	0.8897	31.45	0.8883
	4	31.23	0.8975	31.90	0.9032	32.11	0.9048
	8	31.47	0.9015	32.04	0.9083	32.50	0.9145
16	0	30.66	0.8906	31.79	0.8973	31.76	0.8934
	4	32.17	0.9037	32.15	0.9067	32.28	0.9061
	8	32.43	0.9099	32.27	0.9100	32.58	0.9138
	12	32.60	0.9146	32.36	0.9135	32.77	0.9191
24	0	31.18	0.8835	31.88	0.8970	31.50	0.8865
	8	31.58	0.8951	32.31	0.9086	32.05	0.9030
	16	31.73	0.8999	32.41	0.9122	32.46	0.9132
	20	31.78	0.9020	32.41	0.9134	32.52	0.9157

Table 2: PSNR (dB) and SSIM Comparison of different methods under different patch size and overlap pixels

References

[1] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 9, pp.1167–1183, 2002.

[2] C. Liu, H. Shum, and W. Freeman, "Face hallucination: Theory and practice," Int. J. Comput. Vision, vol. 7, no. 1, pp. 115–134, 2007.

[3] X. Wang and X. Tang, "Hallucinating face by eigen-transformation," IEEE Trans. Syst., Man, Cybern. C, vol. 35, no. 3, pp. 425–434, 2005.

[4] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2004, pp. 275–282.

[5] B. Li, H. Chang, S. Shan, and X. Chen, "Aligning coupled manifolds for face hallucination," IEEE Signal Process. Lett., vol. 16, no. 11, pp. 957–960, 2009.

[6] X. Ma, J. Zhang, and C. Qi, "Position-based face hallucination method," in Proc. IEEE Int. Conf. Multimedia and Expo. (ICME), 2009, pp. 290–293.

[7] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," Pattern Recognition., vol. 43, no. 6, pp. 3178–3194, 2010.

[8] J. Yang, H. Tang, Y. Ma, and T. Huang, "Face hallucination via sparse coding," in Proc. IEEE Int. Conf. Image Processing (ICIP), 2008, pp. 1264–1267.

[9] C. Jung, L. Jiao, B. Liu, and M. Gong, "Position-patch based face hallucination using convex optimization," IEEE Signal Process. Lett., vol. 18, no. 6, pp. 367–370, 2011.

[10] J. Park and S. Lee, "An example-based face hallucination method for single-frame, low-resolution facial images," IEEE Trans. Image Process., vol. 17, no. 10, pp. 1806–1816, 2008.

[11] K. Jia and S. Gong, "Generalized face super-resolution," IEEE Trans. Image Process., vol. 17, no. 6, pp. 873–886, 2008.

[12] E. Candes and J. Rombergt, "Magic: Recovery of sparse signals via convex programming," 2005 [Online]. Available: http://www.acm.caltech.edu/l1magic/

[13] S. Baker and T.Kanade, "Hallucinating faces," in Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition (FG), 2000, pp. 83–88.

[14] J. Jiang, R. Hu, Z. Han, T. Lu, and K. Huang, "Position-patch based face hallucination via locality-constrained representation," in Proc. IEEE Int. Conf. Multimedia and Expo (ICME), 2012, pp. 212–217.

[15] W. Zhang and W.-K. Cham, "Hallucinating face in the DCT domain," IEEE Trans. Image Process., vol. 20, no. 10, pp. 2769–2779, 2011.

[16] . Zhang and W.-K. Cham, "Learning-based face hallucination in DCT domain," in Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2009, pp. 1–8.

[17] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained Linear Coding for Image Classification," in Proc.IEEE Conf. Computer Vision and Pattern Recognition, pp.3360–3367, 2010.

[18] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding" in Proc. Advances in Neural Information Processing Systems (NIPS), 2009, pp. 2223–2231.

[19] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, vol. 290, no. 5500, pp. 2323–2326, 2000.