

Structured Face Hallucination

Chih-Yuan Yang Sifei Liu Ming-Hsuan Yang
Electrical Engineering and Computer Science
University of California at Merced
{cyang35, sliu32, mhyang}@ucmerced.edu

Abstract

The goal of face hallucination is to generate high-resolution images with fidelity from low-resolution ones. In contrast to existing methods based on patch similarity or holistic constraints in the image space, we propose to exploit local image structures for face hallucination. Each face image is represented in terms of facial components, contours and smooth regions. The image structure is maintained via matching gradients in the reconstructed high-resolution output. For facial components, we align input images to generate accurate exemplars and transfer the high-frequency details for preserving structural consistency. For contours, we learn statistical priors to generate salient structures in the high-resolution images. A patch matching method is utilized on the smooth regions where the image gradients are preserved. Experimental results demonstrate that the proposed algorithm generates hallucinated face images with favorable quality and adaptability.

1. Introduction

Face hallucination is a domain-specific super-resolution problem with the goal to generate high-resolution (HR) images from low-resolution (LR) inputs, which finds numerous vision applications. Since a LR image can be modeled from a HR image by a linear convolution process with downsampling, the hallucination problem can be viewed as an inverse task to reconstruct the high-frequency details. While recent work focuses on the generic super-resolution problem, considerable less attention is paid to face hallucination. In this paper, we propose a face hallucination algorithm that exploits domain-specific image structures to generate HR results with high fidelity.

The structure of face image is described in terms of facial components, contours and smooth regions. A landmark detection algorithm is utilized to locate facial components and contours, and process facial alignment in both frontal faces and those at different poses. In this work, the exemplar face dataset consists both LR face images and the corresponding HR ones. The landmark points of each HR exemplar face

image are extracted with two labels (pose information and glasses indicator). From the set of HR exemplar images, the corresponding LR images with landmarks and labels are generated.

Given a test LR image, the pose and landmark points are extracted from an intermediate HR image via bicubic interpolation. Based on the pose and landmark points, the aligned facial components of the input images are compared with those of the training LR images. The LR exemplar images with most similar components are selected, and their gradients are preserved in reconstructing the output HR image. To preserve the structure of edges, we generate HR edges through an anisotropic interpolation and restore the sharpness of edges via statistical priors. For other remaining smooth region, we generate the HR details through a patch match method. The image gradients of these three sources are integrated and used to generate the HR image via the back-projection algorithm. Extensive experiments with comparisons to the state-of-the-art methods show that high-quality images with richer details can be generated by the proposed algorithm without assuming faces are well aligned, at fixed pose and without facial expression change.

2. Related Work

In contrast to generic super-resolution algorithms, recent work in face hallucination aims to learn the mapping between HR and LR patches from a set of exemplar images to recover the missing details of an input frame. In [1], the relationship between LR and HR image patches are modeled in a probabilistic framework such that high-frequency details can be transferred from exemplar images for face hallucination. For every query patch cropped from an input image, the most similar LR patch is retrieved from an exemplar set and the corresponding HR patch is transferred in terms of the first and second order derivatives. The generated face images contain significantly richer details than those by bicubic interpolation, but some artifacts also can be introduced as the transferred HR patches are not structurally consistent although their LR patches are similar to the LR test patches. In other words, a pure patch-based approach

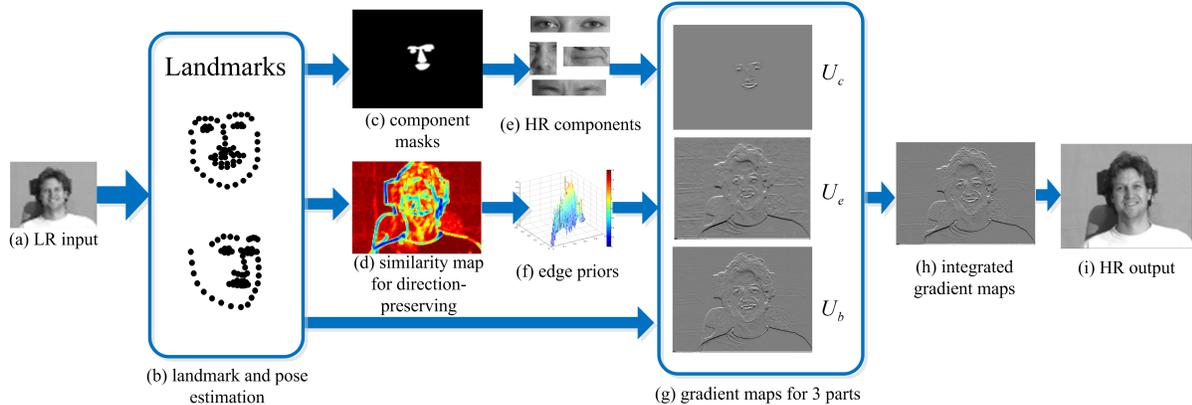


Figure 1. Main steps of the proposed algorithm. (a) A LR input image. (b) Landmarks are estimated to align components and detect pose. (c) Component masks are generated from the test face images. (d) A direction-preserving upsampling method is used to generate HR image. (e) Based on the component masks, the corresponding HR components are found and the set of gradient maps U_c are generated. (f) Priors are used to restore edge sharpness and generate the set of gradient maps U_e . (g) Three sets of gradient maps based on components, edges and smooth regions are generated. (h) Three sets of gradient maps are combined into one. (i) A HR reconstructed image is generated through Eq. 1.

without exploiting structural information is not effective in resolving ambiguities between HR and LR patches.

In [9], a face hallucination method is proposed which enforces linear constraints for HR face images using a subspace learned from a set of training images via Principal Component Analysis (PCA). To restore the high-frequency details not included in the PCA subspace, a patch-based Markov Random Field model is used to reconstruct the residues. Due to the limitation of linear subspace representations, this method performs well only when the images are precisely aligned at fixed poses and expressions. In other cases, the results usually contain ghostly effects due to the adopted PCA-based holistic appearance model.

Instead of enforcing constraints only in reconstructing HR images, a method using constraints in both LR and HR image space is proposed [14]. However, the HR images usually contain ghostly artifacts as a result of using a subspace representation. Furthermore, this method is hard to be extended to handle faces at different poses and expressions effectively. The global linear constraints of subspace representations are replaced by multiple local constraints learned from exemplar patches [10]. When the exemplar and test images are precisely aligned with similar appearance, the adopted local linear constraints are effective as the mapping between HR and LR local patches can be modeled via manifold learning [3]. However, the hallucinated images contain blocky artifacts. On the other hand, the bases can also be learned from all training patches [16] using sparse representation. However, the resulting HR images may contain significant noisy artifacts along contours since the number of training patches collected along edges are relatively less than that of smooth regions and thus the sparse representation dictionary is not effective in reconstructing these re-

gions.

Most recently an algorithm is introduced [13] to handle faces with various poses and expressions by aligning images based on optical flow of SIFT features. This method performs well when training faces are highly similar to the test face in terms of the identity, pose, and expression. However, the global structure of faces is not preserved with this approach as the segments are matched based on local features, thereby generating significant distortion of HR images especially when the identity is not known.

3. Proposed Algorithm

Given a LR test image I_l , we generate a set of HR gradient maps U from exemplar images such that we can generate a HR image I_h based on matching HR gradients and LR intensity values by

$$I_h = \underset{I}{\operatorname{argmin}} \|\nabla I - U\|^2 \quad \text{s.t.} \quad (I \otimes G) \downarrow = I_l, \quad (1)$$

where ∇ is a gradient operator, \otimes denotes convolution, G is σ -width Gaussian kernel, and \downarrow represents a downsampling operator. We group image structures of a face into three categories including facial components, edges, and smooth regions, whose gradients are generated by specific methods to produce the best visual quality. The gradients of facial components are transferred from the corresponding components of exemplar images to preserve the consistency of high-frequency details. The pair of two eyes or two eyebrows is considered as one component rather than two to better preserve consistency, as illustrated in Figure 1(e). On the other hand, local edge properties can be well modeled to learn the statistical relationship between LR and HR to produce stable high-quality results. For smooth regions (e.g., hair

and background), the gradients are reconstructed by small patches to prevent over-smooth artifacts. In addition, we exploit the similarity between the test image and the training images to drive an efficient patch matching algorithm to reduce the computational load of retrieving exemplar patches. Figure 1 shows the main steps of the proposed algorithm.

3.1. Gradient Maps for Facial Components

In order to generate effective gradients of facial components, we prepare a dataset in which every face image is associated with a set of landmark points and two label sets indicating the pose of the face and the existence of glasses on the face. The landmark points are used to generate an aligned image while the pose and glasses labels restrict the search domains.

Aligning exemplar images. Given a LR test image I_l , we generate an intermediate HR image I_b by bicubic interpolation, localize its landmark points and estimate the pose of the test image using [17]. We use the estimated results to select a set of exemplar images in the dataset which have the same pose as I_b . Each face is annotated by several landmark points such that all the facial components and contours are known (Figure 1(b)). Suppose a facial component is annotated by n landmark points denoted as $\{x_i^b, y_i^b\}_{i=1}^n$ of I_b and $\{x_i^e, y_i^e\}_{i=1}^n$ of an exemplar image. We determine the optimal parameters of rotation, scaling and in-plane shift (θ , λ , Δx , and Δy) from

$$\operatorname{argmin}_{\theta, \lambda, \Delta x, \Delta y} \sum_{i=1}^n \|T(x_i^e, y_i^e) - (x_i^b, y_i^b)\|^2, \quad (2)$$

where T is a non-reflective symmetric transformation. We use the estimated parameters to generate an aligned exemplar image, denoted by H . Note that the alignment is carried out for each facial component individually, which is different from existing methods [9, 16, 10] in which faces are aligned based on eyes locations. The proposed alignment approach is more flexible for dealing with face images containing various expressions and shapes because they cannot be effectively aligned by eye positions only.

Determining exemplar images. Suppose $\{H_j\}$ is a set of aligned HR exemplar images for a specific component c , e.g., a nose, a mouth, a pair of eyes, or a pair of eyebrows. We generate the corresponding LR exemplar image

$$L_j = (H_j \otimes G) \downarrow, \quad (3)$$

and compare L_j and the I_l to determine the best exemplar image for the component c . Since every H_j is aligned to the intermediate image I_b at component c , we focus on comparisons of pixels belonging to component c and disregard the other pixels. Based on the landmark points belonging to component c estimated from I_b , we create a HR mask map

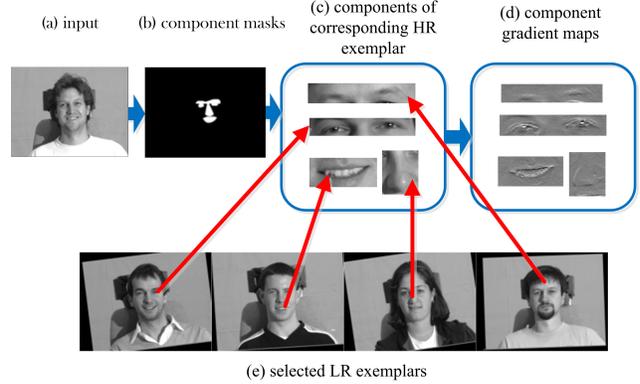


Figure 2. Transferring structural information of each facial component. (a) Input image. (b) Facial components are extracted based on component masks. (c) Components of HR images corresponding to the best matched LR exemplars. (d) Gradient maps U_c are transferred. (e) Selected LR exemplars.

M_h whose pixel values are 1 if the pixel is in c , and otherwise are 0, as shown in Figure 1(b). We generate a LR mask map M_l by downsampling M_h through Eq. 3 and determine the best exemplar by

$$j^* = \operatorname{argmin}_{j \in S} \sum_{p \in I_l} M_l(p) \cdot \|V(p) - V_j(p)\|^2, \quad (4)$$

where V is the gradient maps of I_l as well as V_j of L_j , and S is a set of index j . The index set S is determined by the component c and the labels of glasses associated with the exemplar images $\{H_j\}$. Due to low-pass filtering and subsampling for generating LR images in Eq. 3, some structures of HR images may be missing. For example, the details of glasses frames of a HR image may be missing in the downsampled LR images. As we determine the best exemplar by comparing features in LR, to prevent artifacts caused by selecting an incorrect HR exemplar image, we utilize the labels of glasses to exclude gradient maps V_j from index set S if the component c may be covered by glasses. On the contrary, if the component c is irrelevant to glasses such as a mouth, all V_j are included in S . Figure 2 shows an example of four matched images based on different facial components (the dark boundary of each LR exemplar is the result of alignment). Note that the best matched facial components are matched from images of different subjects.

Transferring gradients of components. Once the best LR exemplar image L_{j^*} is determined for a component, we transfer the gradients of the corresponding source HR image H_{j^*} for the pixels whose values in the mask M_h are 1 as the gradients in the set of gradient maps U_c . The same process is carried out for each component to generate the most effective image gradients, and together they form the gradient map U_c .

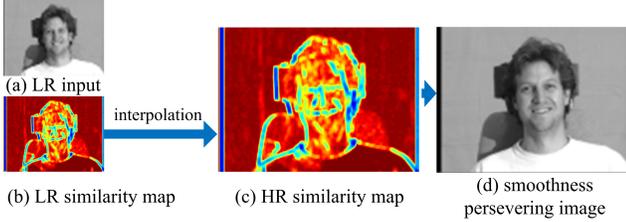


Figure 3. Direction-preserving upsampling. (a) Input image. (b) A set of LR similarity maps is generated. (c) The set of LR similarity maps are upsampled to a set of HR maps through bilinear interpolation to preserve the directions of edges. (d) The HR image is generated by the set of HR similarity maps (c) and the input image (a) through Eq. 6.

3.2. Gradient Maps for Facial Contours

Contours of faces are often represented by image edges which are usually a few pixels in width. Several methods that exploit edge statistics of natural images for super-resolution have been proposed [5, 12]. Although the generated edges are visually pleasing, the HR image may contain significant artifacts (especially along sharp edges) as they are generated by enhancing the contrast of edges from a bicubic interpolated image where edges are jaggy. In this work, we propose to preserve the structure of edges and restore their sharpness through learned statistical priors. Rather than generating sharp edges based on interpolated images, we develop a direction-preserving upsampling function that eliminates the artifacts for prior learning.

Direction-preserving upsampling. For each pixel p in I_l , we compute the directional similarity for each direction k

$$f_k(p) = \exp(-\|P - Q_k\|/\sigma), \quad k = 1, \dots, K, \quad (5)$$

where σ is a parameter, P is a patch centered at pixel p , and Q_k is a neighboring patches of pixel p at direction k . We use patches rather than pixels to compute the directional similarity because patches are less sensitive to noise than pixels. Let $\{S_k\}$ be the K maps of directional similarity computed from I_l . We individually upsample a LR map S_k to a HR map T_k through bilinear interpolation, which does not change the relative magnitude from the original sequence $\{S_1(p), \dots, S_K(p)\}$ to the new sequence $\{T_1(p), \dots, T_K(p)\}$, thereby preserving the directionality similarity. We use the upsampled directional similarity maps $\{T_k\}$ to regularize an under-constrained optimization problem

$$I_d = \operatorname{argmin}_I \sum_k \|f_k(I) - T_k\|^2 \quad \text{s.t.} \quad (I \otimes G) \downarrow = I_l, \quad (6)$$

where $f_k(I)$ is a matrix where each position has the value of $f_k(p)$. Figure 3 shows the process of direction-preserving upsampling. The algorithmic details for solving Eq. 6 are described in the supplementary material.

Learning statistical priors for edge sharpness. Edges in the upsampled image I_d are clear and smooth but the not sharp enough because the sharpness is not modeled in the regularization term of Eq. 6. We learn a non-parametric mapping function to restore the sharpness of edge in I_d . Given a HR training image I_h , we generate a LR image through Eq. 3 and upsample it through Eq. 6 to generate a direction-preserving image I_d . Since the structure of edges are highly symmetric with greatest magnitude of gradients along the center, we label the pixels at edge centers (Figure 4(d)) in I_d by

$$C(p) = \begin{cases} 1 & \text{if } m_p > m_q \text{ and } m_p > m_r \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where m is the magnitude of gradients, and q and r are two neighboring pixels of p at the two opposite normal directions (the direction with largest magnitude changes, as shown in Figure 4(e)). Suppose m' is the magnitude of gradients computed from the training image I_h . For every pixel p in I_d , we extract the (m'_p, m_p, m_c, t) as the features of a statistic sample, where c is the closest edge center pixel of p at distance t in I_d . We collect millions of samples from images of the exemplar dataset, and separate the domain of (m_p, m_c, t) into thousands of bins. We partition the collected samples into each bin, and compute the average of m'_p values of the partitioned samples in individual bins to produce a lookup table (Figure 4(f)) mapping (m_p, m_c, t) to \bar{m}'_p . More details regarding this non-parametric mapping function are described in the supplementary material.

Generating gradient maps. Given a LR test image I_l , we obtain its direction-preserving image I_d through Eq. 6 and label the pixels of edge center by Eq. 7. For very pixel p in I_d , we extract features (m_p, m_c, d) and use the learned look up table to retrieve an expected magnitude value \bar{m}'_p . Suppose U_d are the gradient maps of I_d , we generate the Suppose U_d is the set of the gradient maps of I_d , we generate the set of gradient maps for facial contours U_e by

$$U_e(p) = \frac{\bar{m}'_p}{m_p} \cdot U_d(p). \quad (8)$$

According to the definition of magnitude of gradients $m_p = \|U_d(p)\|$, and statistically \bar{m}'_p presents the magnitude of gradients of p in the training image. Thus the sharpness of edges is stored in U_e as well as the directions are preserved through U_d from I_d .

3.3. Gradient Maps for Smooth Regions

For upsampling regions without distinct structures such as the background, we use small patches to account for large variation of appearance. We generate LR exemplar images from the matched dataset using Eq. 3 and utilize the Patch-Match algorithm [2] to reduce the computational load of retrieving the most similar LR exemplar patches. Suppose the

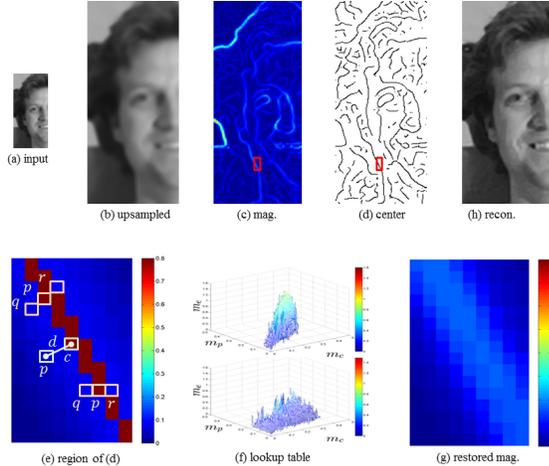


Figure 4. Sharpness restoration (best viewed on a color display). (a) An input LR image. (b) The upsampled image of (a) through Eq. 6. (c) The magnitude of gradients of (b). (d) The map of edge center labeled computed from (c) through Eq. 7. (e) An enlarged region of (c,d). (f) The averaged m_e values in different bins where t is 1 in the top figure and $\sqrt{5}$ in the bottom figure. (g) The map of restored magnitude of gradients from (e) through (f). Note the values are increased over (d). (h) An image generated by the restored gradients to show the effectiveness of the restored edge sharpness.

scaling factor is s , every pixel in the LR image is upsampled to $s \times s$ pixels in HR. Since every retrieved LR patch is generated from a HR patch, we use the central $s \times s$ pixels of the HR patch to generate a HR image and then apply the back projection algorithm [7] to adjust HR image based on the input image I_l to ensure the HR image satisfies Eq. 3. We extract the gradients of the back-projected HR image as the gradients of smooth regions, denoted by U_b (Figure 1(g)).

3.4. Integrating Gradient Maps

In order to generate the required gradient map set U for producing the output HR image, we generate two weight maps w_c and w_e . We set map w_c as the summation of all HR mask maps M_h (Figure 1(b)), and set $w_e(p) = \min\{1, \alpha \bar{m}_e\}$, where \bar{m}_e is the gradient magnitude in Eq. 8, and α is a parameter to control the weight. In the map w_e , the weight values are large when the pixels are close to edges. We integrate the three sets of gradient maps U_c, U_e, U_b by

$$U = w_c U_c + (1 - w_c)(w_e U_e + (1 - w_e) U_b), \quad (9)$$

for every pixel, and generate the output HR image using U through Eq. 1. Figure 5 shows HR images generated by different gradient maps. It is clear that the use of each gradient map facilitates generating better results for different facial structures.

4. Experimental Results

We use the Multi-PIE dataset [6] to evaluate the proposed algorithm as it consists of subjects with variation of



(a) without U_c (b) with U_c (c) without U_e (d) with U_e Figure 5. Effectiveness of the integrated gradient maps. (a)(b) The generated gradient maps U_c ensure consistency of high-frequency details at components. (c)(d) The generated gradient maps U_e ensure clear and sharp edges.

age, pose, and expression. As the main goal of this paper is on face hallucination, we use the set of images taken under the same lighting conditions. The pose labels and landmarks of each image are given in the dataset, and we manually generate the glasses labels for training images. One set with 2,184 320×240 images at upright frontal pose of 289 individuals is used as the training dataset for face hallucination experiments (Figure 6, Figure 7 and Figure 8). Another set of 283 images with pose at 30 degrees of 60 subjects is also used for experiments (Figure 9 and Figure 10). The test set consists of LR images consist of 351 images of frontal and different poses where there the subject identities do not overlap with those of the training set. We generate the input LR images by downsampling the original HR test images through Eq. 3 with a scaling factor 4 and the Gaussian kernel width 1.6, commonly used in the literature [9]. The ground truth HR images in the test set are used for comparisons with the generated hallucination results. Implemented in MATLAB, it takes 1 minute to process one 60×80 LR image on a machine with 2.8 GHz Quad Core CPU. All the MATLAB code and dataset will be made available to the public.

For color images, we apply the proposed algorithm on grayscale channel and the color channels are upsampled by bicubic interpolation to make fair comparisons with existing methods [9, 16, 10]. We quantitatively evaluate the generated images using PSNR, structural similarity (SSIM) [15] and DIIVINE index [11]. The PSNR and SSIM values indicate the fidelity of the generated images to the original image, while DIIVINE index measures the image quality based on regressed scores collected from user study.

To label facial landmarks, we use the algorithm of [17] which produces the landmarks as the active appearance model [4] with 68 points as shown in Figure 1(b). For eye and mouth components, the masks are generated by the convex polygons formed by adjacent landmarks. As the landmarks of eyebrows and nose do not form a close polygons, we mask eyebrows as the rectangles where the landmarks are the center vertical segments. The nose mask is defined by a triangle covering the top and bottom landmark points (Figure 1(c)).

We implement several state-of-the-art face hallucination

algorithms [9, 16, 10] for comparisons. More qualitative and quantitative results with comparisons to other generic super resolution methods are available in the supplementary material. Figures 6, 7 and 8 show hallucinated faces of frontal pose where the input images are enlarged by nearest neighbor interpolation for illustration purpose. The method of [9] performs well but it generates noisy and blocky effects (e.g., hair) and over-smooth details (e.g., mouth). In addition, the hallucinated results are inconsistent (e.g., one eye of Figure 6(e) contains glasses and the other one does not) with significant ghostly effects. While the algorithms based on sparse coding [16] and position-patch [10] generate high-frequency textures, the results do not contain fine facial details such as contours and hair (See Figure 6). This can be attributed to that most training patches are composed of textures without sharp edges. The proposed algorithm reconstruct fine details of facial components such as the spots and moles in Figure 7(f) and individual tooth in Figure 6(f). Note the details in the transferred components are different from the ones in the ground truth images (e.g., the width of pupils and the arrangements of teeth are different) as the subjects of training and test sets do not overlap. However, they are visually correct and pleasing because downsampling is a many-to-one mapping and the ground truth image is just one of many feasible HR images. We also note that the classic back-projection method [7] produces the largest PSNR and SSIM values in many cases but the results contain jaggy edges (e.g., hair, nose and contour of Figure 6(b), Figure 7(b), and Figure 8(b)) and fewer details (eyes and mouth of Figure 6(b), Figure 7(b), and Figure 8(b)).

In addition, the proposed algorithm also generates better details due to the glasses label and the component-level alignment. Take glasses for example, while other methods [16, 9, 10] easily generate false glasses contours (Figure 6(c),(e) and Figure 7(c),(e)), no glasses at all (Figure 8(d) and Figure 9(e)), or ghostly effects (Figure 9(d)), the hallucinated results by the proposed algorithm are generally correct with more details.

We also compare the hallucination results on faces at different pose. The training process for [9, 10, 16] are carried out on a training set containing faces taken at 30% pose. As shown in Figure 9 and Figure 10, the subspace-based methods [10] and [9] do not perform well as both the subspace learning and the patch reconstruction at fix positions require precise face alignment. However, it is more difficult to align face images at different pose such that PCA subspace can be well constructed for hallucination. In spite of pose variation, the back-projection [7] method and the proposed algorithm perform consistently well. However, our algorithm generates more fine details and sharper edges. We also evaluate the performance of different algorithms using exemplars from the PubFig dataset [8] where face images are not taken in a well controlled lab environment. Fig-

ure 11 shows the proposed algorithm performs well whereas the other algorithms fail to generate high quality results.

5. Conclusion

A novel approach that exploits image structures for face hallucination is proposed in this paper. The image structures of a face are grouped into three categories including facial components, edges, and smooth regions. Their gradient maps are generated and integrated to produce HR results with the best visual quality. Experimental results show that the proposed algorithm generates hallucinated face images with fine and consistent details over state-of-the-art algorithms.

6. Acknowledgments

This work is supported in part by the NSF CAREER Grant #1149783 and NSF IIS Grant #1152576.

References

- [1] S. Baker and T. Kanade. Hallucinating faces. In *FG*, 2000.
- [2] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein. The generalized patchmatch correspondence algorithm. In *ECCV*, 2010.
- [3] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *CVPR*, 2004.
- [4] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *PAMI*, 23(6):681–685, 2001.
- [5] R. Fattal. Image upsampling via imposed edge statistics. In *SIGGRAPH*, 2007.
- [6] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *FG*, 2008.
- [7] M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP*, 53(3):231–239, 1991.
- [8] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and Simile Classifiers for Face Verification. In *ICCV*, 2009.
- [9] C. Liu, H.-Y. Shum, and W. T. Freeman. Face hallucination: Theory and practice. *IJCV*, 75(1):115–134, 2007.
- [10] X. Ma, J. Zhang, and C. Qi. Hallucinating face by position-patch. *PR*, 43(6):2224–2236, 2010.
- [11] A. Moorthy and A. Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *TIP*, 20(12):3350–3364, 2011.
- [12] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum. Image super-resolution using gradient profile prior. In *CVPR*, 2008.
- [13] M. F. Tappen and C. Liu. A Bayesian approach to alignment-based image hallucination. In *ECCV*, 2012.
- [14] X. Wang and X. Tang. Hallucinating face by eigentransformation. *SMC*, 35(3):425–434, 2005.
- [15] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *TIP*, 13(4):600–612, 2004.
- [16] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *TIP*, 2010.
- [17] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *CVPR*, 2012.

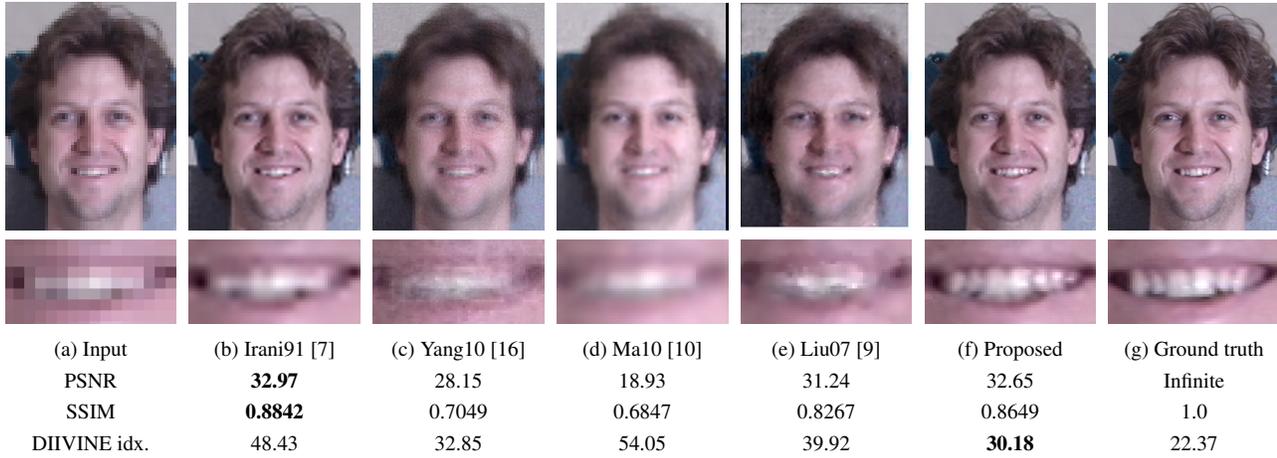


Figure 6. Qualitative comparison for 4 times upsampled upright frontal faces (results best viewed on a high-resolution display).

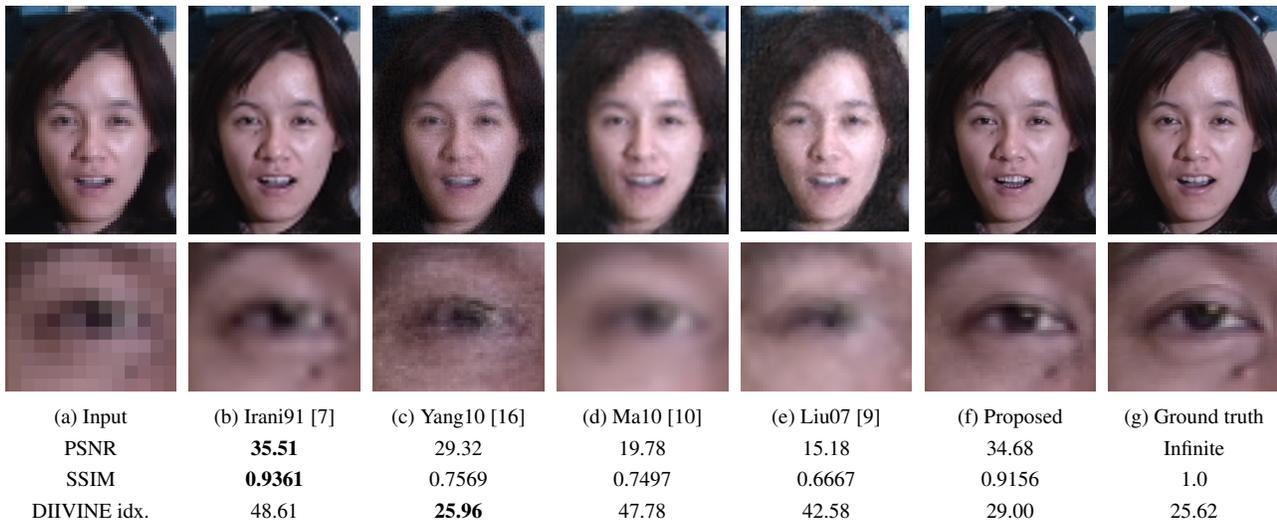


Figure 7. Qualitative comparison for 4 times upsampled upright frontal faces (results best viewed on a high-resolution display).

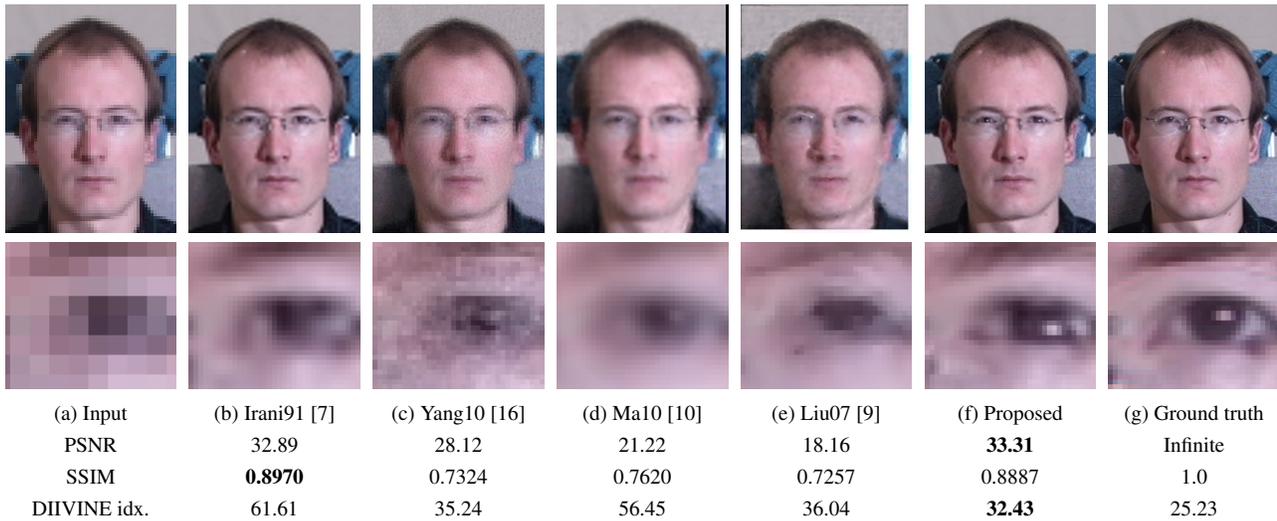
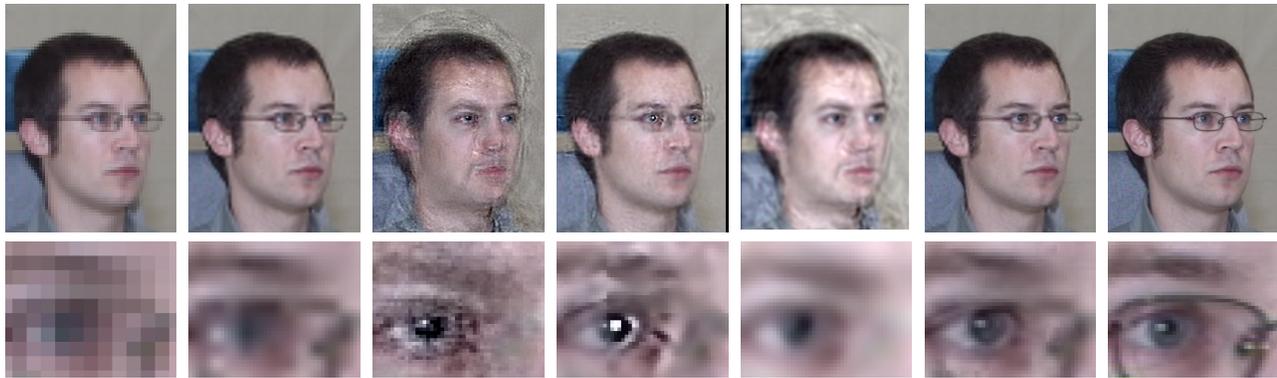


Figure 8. Qualitative comparison for 4 times upsampled upright frontal faces (results best viewed on a high-resolution display).



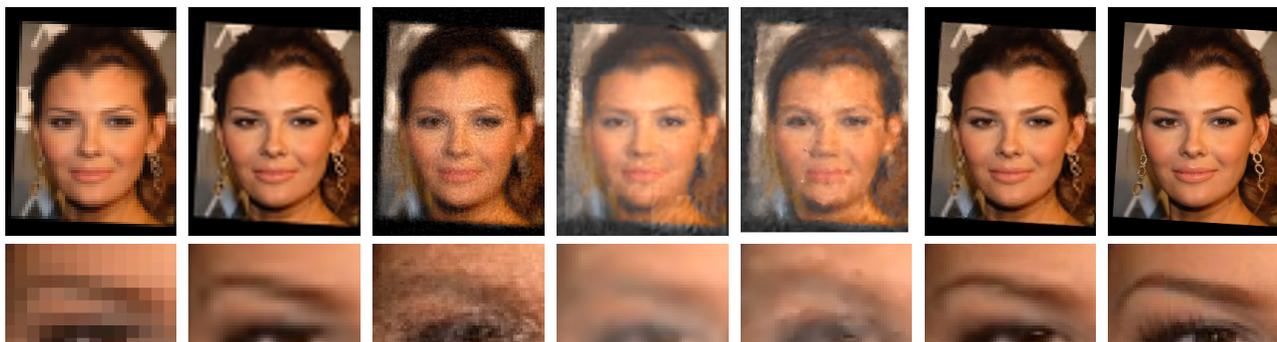
(a) Input	(b) Irani91 [7]	(c) Yang10 [16]	(d) Ma10 [10]	(e) Liu07 [9]	(f) Proposed	(g) Ground truth
PSNR	33.87	23.17	20.89	16.53	33.23	Infinite
SSIM	0.9126	0.4943	0.7968	0.6663	0.8873	1.0
DIIVINE idx.	57.94	29.57	36.23	50.96	32.82	30.05

Figure 9. Qualitative comparison for 4 times upsampled non-frontal faces (results best viewed on a high-resolution display).



(a) Input	(b) Irani91 [7]	(c) Yang10 [16]	(d) Ma10 [10]	(e) Liu07 [9]	(f) Proposed	(g) Ground truth
PSNR	35.11	23.20	21.78	16.12	34.22	Infinite
SSIM	0.9028	0.4733	0.7595	0.6332	0.8711	1.0
DIIVINE idx.	48.48	27.06	40.52	49.71	30.36	22.21

Figure 10. Qualitative comparison for 4 times upsampled non-frontal faces (results best viewed on a high-resolution display).



(a) Input	(b) Irani91 [7]	(c) Yang10 [16]	(d) Ma10 [10]	(e) Liu07 [9]	(f) Proposed	(g) Ground truth
PSNR	29.33	24.16	15.23	13.05	30.04	Infinite
SSIM	0.8338	0.5474	0.5230	0.4948	0.8798	1.0
DIIVINE idx.	50.05	23.14	47.25	38.91	33.92	29.34

Figure 11. Qualitative comparison for 4 times upsampled upright frontal faces (results best viewed on a high-resolution display).