THREE-TIERED NETWORK MODEL FOR IMAGE HALLUCINATION

Lin Ma[1,], Yonghua Zhang[2], Yan Lu[2], Feng Wu[2], Debin Zhao[1]*

[1]Harbin Institute of Technology, Harbin, 150001, China
[2]Microsoft Research Asia, Beijing, 100080, China

ABSTRACT

In this paper, we propose a novel three-tiered network model for image hallucination based on the learnt knowledge composed of image patches relating low and high resolution. A common problem of previous hallucination methods is that irregularities are usually introduced into the constructed high-resolution images. We remove the irregularities in three steps. First, the hallucination with primal sketch priors is performed to construct a coarse high-frequency component. Second, enhancement is implemented to enforce local compatibility between the patches in the constructed component. Third, a Markov network is utilized to refine the enhanced high-frequency component. Experiments demonstrate that our model can hallucinate higher-quality images than existing methods.

Index Terms— Image hallucination, three-tiered network model, Markov network.

1. INTRODUCTION

Image hallucination as a low-level vision problem has attracted a lot of attentions in computer vision since it was first proposed by Freeman et al. in 2000 [1]. Typically, for one input low-resolution (LR) image, one corresponding high-resolution (HR) image can be hallucinated with the help of the knowledge learnt from a training set of images. The knowledge usually consists of pairs of patches relating low and high resolution, which are small features of low frequency (LF) and high frequency (HF) in image.

The most fundamental issue in image hallucination is whether an input LF patch can be used as the index to find the proper HF patch in the learnt knowledge. LF patch is often of low-dimension in image hallucination, whereas HF patch of high-dimension. In theory, it is an ill-posed problem, namely, one low-dimension patch may be mapped to more than one high-dimension patches.

Several approaches have been proposed to solve this problem. In [1], Freeman et al. model the relationships of LF and HF patches by a Markov network and use Bayesian belief propagation to find a local maximum of the posterior probability. Liu et al. propose a two-step statistical modeling approach that integrates both a global parametric model and a local non-parametric model [2]. Sun et al. apply a Markov-chain based inference algorithm to hallucinate primal sketch priors instead of low-level features, which achieves better visual quality in generic images [3]. The locally linear embedding (LLE) approach is adopted in [4] and [5] to consider local geometry during the mapping.

(a) (b) (c)

Figure 1: An example to check the mapping problem in image hallucination: (a) original image; (b) hallucinated result by [3]; (c) expected hallucination.

Recently, the similar ideas to image hallucination have been also proposed for image compression and vector quantization. Li et al. propose a low bit-rate image compression scheme by incorporating primal sketch based learning [6]. Wu et al. propose a new vector quantization by introducing visual patterns on designing codebooks [7]. Both demonstrate good visual quality at low bit-rate compression.

However, is the problem of mapping from LF patch to HF patch really solved? Let us see a simple example here. *Figure 1* (a) is a part of original image Lena. *Figure 1* (b) is the hallucinated result by [3]. *Figure 1* (c) is the expected hallucinated result, where we use HF patches from Lena, although which are not available in practical applications, to find the most similar HF patches in the learnt knowledge. The same knowledge is used in *Figure 1* (b) and (c). *Figure 1* (c) verifies that all requested HF patches already exist in the knowledge. One can observe that there are still several visible artifacts in *Figure 1* (b). It is caused by improper HF patches blending.

In order to solve the above problem, we propose a novel three-tiered network model for image hallucination. First, a coarse high-frequency component is obtained by hallucination with primal sketch priors [3]. Second, it is enhanced to preserve local consistencies between adjacent HF patches. Third, inspired by the above expected hallucinated result, we propose a Markov network to model the relationship between the existent HF patch in the learnt knowledge and the enhanced one. Finally, hallucination result is obtained by

*This work was done while Lin Ma was with Microsoft Research Asia as a Research Intern.

blending the target high-frequency component inferred with our model into the image interpolated from the LR image.

The rest of this paper is organized as follows. In Section 2, three-tiered network model for image hallucination is introduced. In Section 3, we give a brief description of our multi-phase image hallucination framework. The experimental results are presented in Section 4. Finally, Section 5 concludes the paper.

2. THREE-TIERED NETWORK MODEL

The proposed model for image hallucination, as shown in *Figure 2*, consists of three tiers: *Hallucination*, *Enhancement* and *Realistic Mapping*. In *Figure 2*, circles represent the network nodes and lines indicate statistical relationships between the nodes. L_s^i is the LF patch and $H_s^{k\_i}$ is the HF patch to be inferred in tier k ($k = 1,2,3$), with i as the node number.

2.1. Hallucination

In the first tier, hallucination can be formulated as follows. Given a low-frequency image L interpolated from the LR image, a high-frequency component $H_{C,1}$ is hallucinated with the knowledge learnt from a training set of images. The generation of each HF patch $H_s^{1\_i}$ should preserve both the mapping accuracy from L_s^i to $H_s^{1\_i}$ and the compatibility between adjacent HF patches.

With Markov-chain based inference algorithm in [3], the MAP of high-frequency component $H_{C,1}$ is obtained by maximizing posterior probability $p(H|L)$, with $p(H)$ as the prior learnt knowledge:

$$H_{C,1} = argmax\big(p(H|L)\big). \tag{1}$$

Considering mapping accuracy and compatibility,

$$H_{C,1} = argmax\left(\prod_{(i\,j)} \psi_{ij}\left(H_s^{1\_i}, H_s^{1\_j}\right) \prod_i \phi_i\left(L_s^i, H_s^{1\_i}\right)\right) \tag{2}$$

where $(i\ j)$ is the adjacent node pair, $\psi_{ij}\left(H_s^{1\_i}, H_s^{1\_j}\right)$ is the compatibility function, and $\phi_i\left(L_s^i, H_s^{1\_i}\right)$ is the accuracy function of mapping from L_s^i to $H_s^{1\_i}$.

However, the above method cannot preserve the compatibility and smoothness between adjacent HF patches, as shown in *Figure 1* (b). It is caused by the inaccurate mapping, which roots in the ill-posed problem. In order to solve the problem and enhance the consistency between adjacent HF patches, we propose to enforce $H_{C,1}$ in the *Enhancement* tier.

2.2. Enhancement

In this tier, compatibility is enhanced by combining H_E with $H_{C,1}$ generated in the *Hallucination* tier. Thus, a high-frequency component $H_{C,2}$ with higher compatibility is

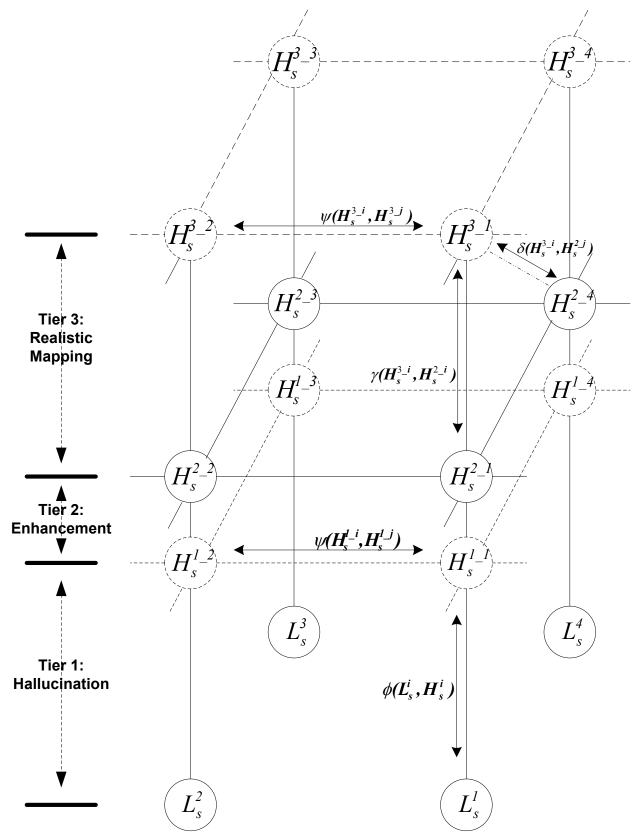


Figure 2: Three-tiered network model for image hallucination.

obtained. Since each combined HF patch in $H_{C,2}$ is not as realistic as the one in the learnt knowledge, the visual quality of the resulting image is unsatisfactory when $H_{C,2}$ is blended into L. Therefore, in the *Realistic Mapping* tier, we propose a model to replace the enhanced HF patch with the existent one in the learnt knowledge.

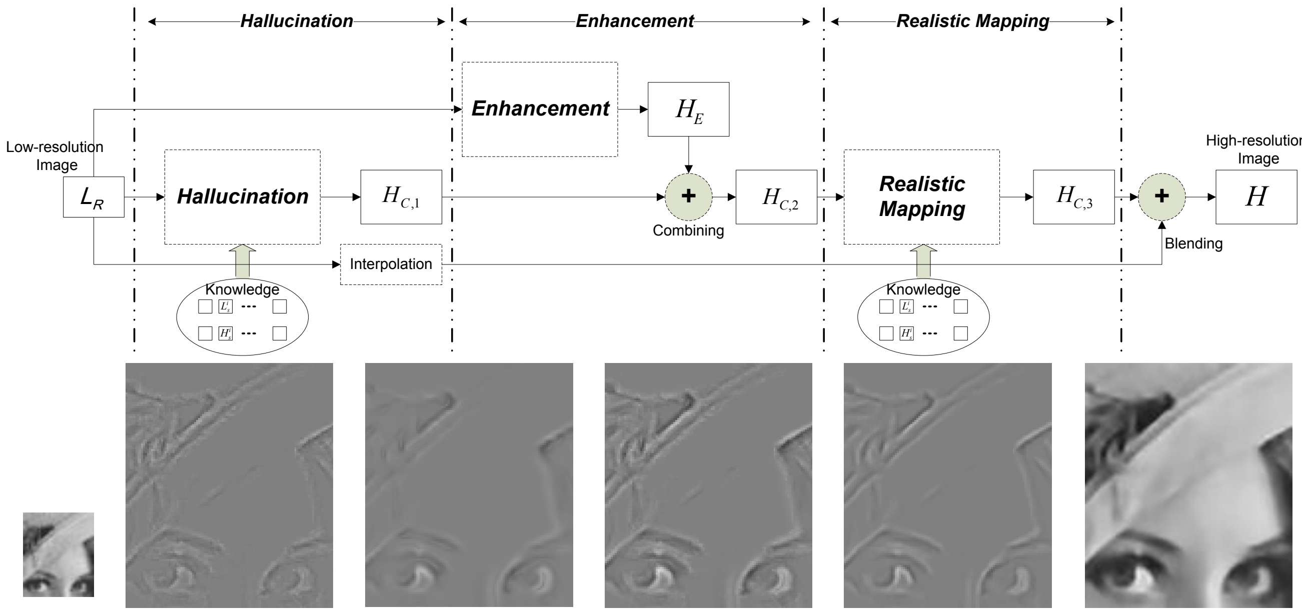
2.3. Realistic Mapping

We utilize a Markov network to model the relationship between enhanced and existent HF patch as well as the relationship between neighboring HF patches. The MAP of component $H_{C,3}$ is obtained by:

$$H_{C,3} = argmax\left(p(H|H_{C,2})\right)$$

$$= argmax\left(\prod_i \gamma_i\left(H_s^{3\_i}, H_s^{2\_i}\right) \prod_{(i\,j)} \psi_{ij}\left(H_s^{3\_i}, H_s^{3\_j}\right) \prod_{(i\,j)} \delta_{ij}\left(H_s^{3\_i}, H_s^{2\_j}\right)\right), \tag{3}$$

where $\gamma_i\left(H_s^{3\_i}, H_s^{2\_i}\right)$ indicates the mapping accuracy from the enhanced patch to the existent one in the learnt knowledge, $\varphi_{ij}\left(H_s^{3\_i}, H_s^{3\_j}\right)$ is the compatibility function between neighboring patches in the same tier and $\delta_{ij}\left(H_s^{3\_i}, H_s^{2\_j}\right)$ indicates compatibility of neighboring patches in adjacent tiers. As the influence of δ has been considered by mapping



(a)L_R (b)$H_{C,1}$ (c)H_E (d)$H_{C,2}$ (e)$H_{C,3}$ (f) H

Figure 3: Our image hallucination framework: (a) the input low-resolution image L_R; (b) high-frequency component $H_{C,1}$ hallucinated by [3]; (c) high-frequency component H_E generated in Enhancement tier; (d) enhanced high-frequency nent $H_{C,2}$; (e) target high-resolution component $H_{C,3}$ with our model.

accuracy ϕ and compatibility ψ, we can simplify the posterior probability $p(H|H_{C,2})$ as:

$$H_{C,3} = argmax\left(p(H|H_{C,2})\right)$$
$$\approx argmax\left(\prod_i \gamma_i\left(H_s^{3\_i}, H_s^{2\_i}\right)\prod_{(i\,j)} \psi_{ij}\left(H_s^{3\_i}, H_s^{3\_j}\right)\right). \quad (4)$$

3. MULTI-PHASE IMAGE HALLUCINATION

Based on the proposed three-tiered network model, we further develop a multi-phase image hallucination framework. *Figure 3* depicts the three processing phases.

In the phase of *Hallucination*, image hallucination with primal sketch priors as that in [3] is utilized to generate the coarse high-frequency component $H_{C,1}$, with the assistance of the learnt knowledge.

In the phase of *Enhancement*, a high compatibility component H_E inferred is combined with $H_{C,1}$ to produce the high-frequency component $H_{C,2}$ with higher compatibility. The proposed enhancement method is based on Laplacian pyramid [8], which has been widely used in image compression and enhancement [9] [10]. Similar to [10], a control function [11] calculated from the Gaussian image is performed on the interpolated Laplacian image to generate H_E. Therefore, the compatibility in primal sketch is preserved, and meanwhile other regions remain unaffected, which can be observed in *Figure 3* (c).

In the phase of *Realistic Mapping*, the replacement of enhanced HF patch should consider both the mapping accuracy and compatibility of neighboring patches. In our implementation, we utilize Euclidean distance to define the mapping accuracy ϕ from the existent patches to the enhanced ones. As for the compatibility function ψ, Sum Squared Difference (SSD) of the overlapped region between adjacent HF patches is calculated to evaluate the compatibility [3]. However, since $H_{C,2}$ is with higher compatibility, we just use a simple method to enforce the relationship between adjacent patches by averaging the pixel values in overlapped regions.

In the end, the high-frequency component $H_{C,3}$ is generated by combining all HF patches derived from the learnt knowledge. By blending it into L, the HR image H is obtained.

4. EXPERIMENTAL RESULTS

We build the knowledge for our image hallucination from band- and high-pass pairs, corresponding to the low- and high-frequency patches. These pairs are extracted from a training set of 24 Kodak images [12]. Since the missing HF patches to be inferred are densely distributed over the regions of image primitives, we extract all the center-aligned patch pairs in these regions. In our implementation, the size of the primitive patch in the knowledge is 7×7. Details of image primitive patch extraction can be referred to [3]. The total number of the extracted patch pairs in the knowledge is about 50,000.

Then, we test our approach on a set of generic images. The input LR image is obtained by blurring and downsampling the HR image. We compare our approach with bicubic interpolation, backprojection [13] and image hallucination

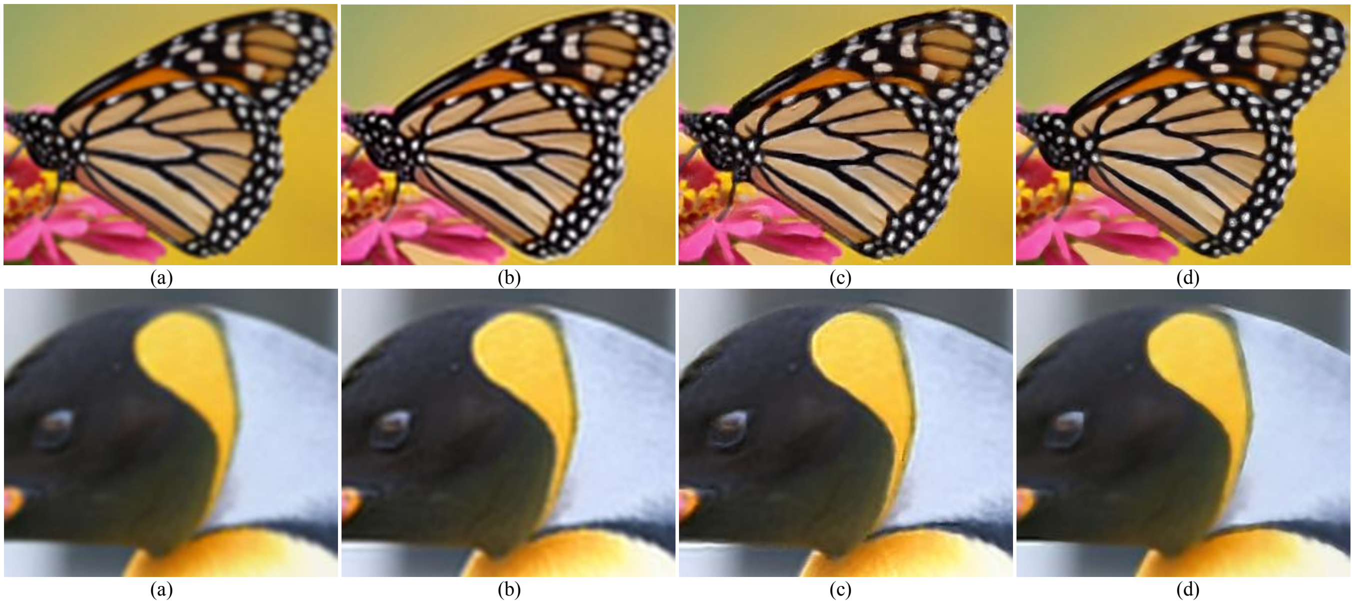


Figure 4: Test images magnified by three times using (a) bicubic interpolation; (b) backprojection; (c) hallucination by [3]; (d) our approach.

proposed by Sun et al.[3], all with a magnification factor of 3. The results are shown in *Figure 4*. On one hand, it is clear to see that our approach sharpens the edges, compared with the "blurring" artifacts introduced by bicubic interpolation and the "ringing" artifacts brought in by backprojection [13]. On the other hand, due to the three-tiered model, our approach removes the irregularities in the synthesized high-resolution image, compared with the hallucination results by [3].

5. CONCLUSION

In this paper, we have proposed a novel three-tiered network model for image hallucination, which can help to enforce the accuracy of mapping from low-frequency patches to high-frequency ones, and meanwhile preserve the compatibility between adjacent high-frequency patches. Based on the proposed model, a multi-phase framework of image hallucination is built. Experimental results have shown that our approach synthesizes higher-quality images compared with other schemes.

6. REFERENCES

[1] W. T. Freeman, E. Pasztor, O. Carmichael, "Learning low-level vision", *International Journal of Computer Vision*, pp. 25-47, 2000.

[2] C. Liu, H. Y. Shum, C. Zhang, "A two-step approach to hallucinating faces: global parametric model and local nonparametric model", *International Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 192-198, 2001.

[3] J. Sun, N. N Zheng, H. Tao, H. Y. Shum, "Image Hallucination with Primitive Sketch Priors", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003.

[4] H. Chang, D. Y. Yeung, Y. Xiong, "Super-resolution Through Neighbor Embedding", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 275-222, 2004.

[5] W. Fan, D. Y. Yeung, "Image Hallucination Using Neighbor Embedding over Visual Primitive Manifolds", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 18-23, 2007.

[6] Y. Li, X. Sun, H. Xiong, F. Wu, "Incorporating Primal Sketch based Learning into Low Bit-Rate Image Compression", *International Conference on Image Processing*, vol. 3, pp. 173-176, 2007.

[7] F. Wu, X. Sun, "Image compression by visual pattern vector quantization (VPVQ)", *Data Compression Conference*, 2008.

[8] P.J. Burt, E.A. Adelson, "The Laplacian Pyramid as a Compact Image Code", *IEEE Transactions on Communications*, vol. 31, pp. 523-540, 1983.

[9] H. Greenspan, C.H. Anderson, S. Akber, "Image Enhancement By Nonlinear Extrapolation in Frequency Space", *IEEE Transactions on Image Processing*, vol. 9, no. 6, pp. 1035-1048, 2000.

[10] B.W. Jeon, R.H. Park, S. Yang, "Resolution Enhancement by Prediction of the High-Frequency Image Based on the Laplacian Pyramid", *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID. 72520, 2006.

[11] G. Ramponi, A. Polesel, "A Rational Unsharp Masking Technique", *Journal of Electronic Imaging*, vol. 7, pp. 333-338, 1998.

[12] Kodak Lossless True Color Image Suite: http://r0k.us/graphics/kodak.

[13] M. Irani, S. Peleg, "Motion Analysis for Image Enhancement: Resolution, Occlusion, and Transparency", *Journal of Visual Communication and Image Representation*, vol. 4, pp. 324-335, 1993.