# INCORPORATING PRIMAL SKETCH BASED LEARNING INTO LOW BIT-RATE IMAGE COMPRESSION

*Yang Li[1]*, Xiaoyan Sun[2], Hongkai Xiong[1], Feng Wu[2]*

[1]Department of Electrical Engineering, Shanghai Jiao Tong University, P.R.China 200240
[2]Microsoft Research Asia, No.49 Zhichun Road, Beijing, China 100080

## ABSTRACT

This paper proposes an image compression approach, in which we incorporate primal sketch based learning into the mainstream image compression framework. The key idea of our approach is to use primal sketch information to enhance the quality of distorted images. With this method, we only encode the down-sampled image and use the primal sketch based learning to recover the high frequency information which has been removed by down-sampling. Experimental results demonstrate that our scheme achieves better objective visual quality as well as subjective quality compared with JPEG2000 at the same bit-rates.

***Index Terms***—image compression, primal sketch, low bit-rate coding

## 1. INTRODUCTION

Low bit-rate image coding is important in applications such as storage in low memory devices or streaming data over wireless network. However, the state-of-the-art image coding schemes have inherent visible artifacts at low bit-rate, such as block effect introduced by DCT transform and ringing effect introduced by wavelet transform. It makes the low bit-rate-coded images difficult to be clearly displayed on a high resolution screen or magnified to check specific regions.

Attempts have been made to compress images at lower resolution for bits-saving at low bit-rate [1]. They lead to good quality of reconstructed low resolution images but their up-sampled versions still need to be enhanced. Developments on up-sampling have been reported in various super/high resolution schemes by taking advantage of statistical characteristics of images in transform domain [2][3]. Furthermore, an image reconstruction method from its multi-scale edge representation is also presented in [4].

Recent researches in computer vision enlighten us on decomposing an image into visual primitives such as contours and textures to construct a high quality image. Bayesian approach based image hallucination [5] is one of these schemes on constructing super resolution images. It demonstrates that the primitives of an image can be learned from the primitives of other generic images. For the non-distorted images, the hallucinated results are very encouraging.

The progresses in the vision techniques have opened a new vista in image coding. In this paper, we present an image coding scheme with primal sketch based learning. The key contribution lies in the incorporation of learning based mapping on primal sketch into the current standard image compression framework. The high frequency information discarded at the encoder will be recovered at the decoder by primal sketch based patch mapping. Both subjective and objective visual qualities are enhanced.

The rest of this paper is organized as follows. In Section 2, we give an overview of our whole framework. Detailed training and mapping processes are introduced in Section 3 and 4. Section 5 presents the experiment results. Conclusion will be drawn at Section 6.

## 2. OVERVIEW OF OUR APPROACH

An overview of our approach is shown in *Figure* 1. There are three steps within this compression framework. In the first step, an original image $I_o$ is down-sampled into $1/\alpha$ size by a low pass filter. In the second step, standard JPEG2000 is utilized to gain a compression ratio $\beta$ on the down-sampled image. Therefore the overall compression ratio is $1/(\alpha^2\beta)$. Obviously, high frequency information is lost after the low pass filtering. Thus, in the third step, we adopt learning based mapping to infer the dropped information.

* This work was done during Yang Li's internship in Microsoft Research Asia. This work was partly supported by the Grants NSFC No. 60632040 from the National Science Foundation of China.

Mathematically, the whole process can be formulated as (1.1) and (1.2):

$$I_d = (I_o * G_d) \downarrow \alpha \quad (1.1)$$

$$I_u = (I'_d * G_u) \uparrow \alpha \quad (1.2)$$

Where $I_o$ denotes the original high resolution image, $G_d$ and $G_u$ are two low pass filters for down-sampling and up-sampling respectively. $I_d$ is the down-sampled low resolution image, $I'_d$ is the distorted down-sampled image, and $I_u$ is the up-sampled low resolution image. "*" denotes the convolution operator.

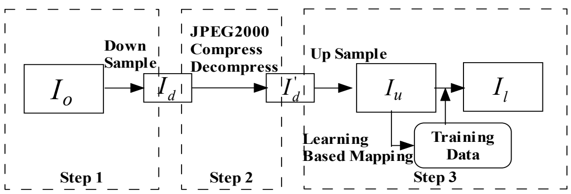



*Figure* 1 Overview of our approach.

The most important part of our work is to recover the missing information discarded at the encoder side in order to eliminate the unpleasant artifacts brought by low bit-rate coding. Sun et al. have claimed in [5] that for high quality images, patches extracted along the edges of an up-sampled image are effective for mapping because they are in low dimensionality. Therefore, in our low bit-rate image coding scheme, we also focus on edge regions to recover the high frequency information. The effectiveness of this approach will be analyzed in Section 4.

## 3. PRIMAL SKETCH BASED LEARNING AND MAPPING

The essential part of our proposed scheme is the Learning Based Mapping process as shown in *Figure* 1. We employ a training set to assist the mapping process. The training set is obtained from a set of generic images offline, which contains pairs of patches as shown in *Figure* 2.

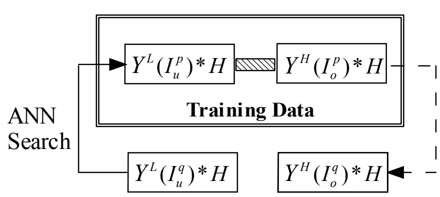



*Figure* 2 Locate the missing high frequency information.

In this figure, $H$ is a high pass filter. $p$ and $q$ denote two different images respectively. $I_o$ is the original version of an image, while $I_u$ is its up-sampled low resolution version of an image by our scheme. $Y^H(I_o^p)$ is a high resolution patch from the original image $I_o^p$, and $Y^L(I_u^p)$ is the relative low resolution patch from the corresponding image $I_u^p$ which is up-sampled after distortion. For another image $I_o^q$, we want to use the similarity between $Y^L(I_u^q)*H$ and $Y^L(I_u^p)*H$ to infer the missing $Y^H(I_o^q)*H$.

A crucial aspect of the patch mapping process is the definition of similarity. This similarity should be able to measure the relationship between the original patch $Y^H(I_o^p)*H$ and its up-sampled version $Y^L(I_u^p)*H$. Meanwhile, it should also be able to measure the relationship between patches $Y^L(I_u^p)*H$ and $Y^L(I_u^q)*H$. This issue is challenging, because we want to find a metric that is able to preserve recognizable features between the original patch $Y^H(I_o^p)*H$ and its up-sampled version $Y^L(I_u^p)*H$, while at the same time these features are also able to be applied to a different target image $I_o^q$.

Our approach is to apply a linear map between the high resolution and low resolution patches on the primal sketch. More specifically, let $Y^L(I_u^p)*H$ denote the patch of a low resolution image $I_u^p$ after high pass filtering, and $Y^L(I_u^q)*H$ denote the low resolution patch after the same process of another image $I_u^q$. We observe that, on the primal sketch, if $(Y^L(I_u^q)*H)/\sigma_q$ is similar to $(Y^L(I_u^p)*H)/\sigma_p$, $(Y^H(I_o^p)*H)/\sigma_p$ is also very similar to $(Y^H(I_o^q)*H)/\sigma_q$. Where $\sigma_p$ and $\sigma_q$ are the standard deviations of the luminance after high pass filtering with respect to luminance distributions in $I_u^p*H$ and $I_u^q*H$, respectively. Therefore, we can deduce the missing $(Y^H(I_o^q)*H)/\sigma_q$ from $(Y^H(I_o^p)*H)/\sigma_p$ by the mapping of (3.1) and (3.2), the latter of which can be found in the training data.

$$(Y^L(I_u^q)*H)/\sigma_q \leftrightarrow (Y^L(I_u^p)*H)/\sigma_p \quad (3.1)$$

$$(Y^L(I_u^p)*H)/\sigma_p \leftrightarrow (Y^H(I_o^p)*H)/\sigma_p \quad (3.2)$$

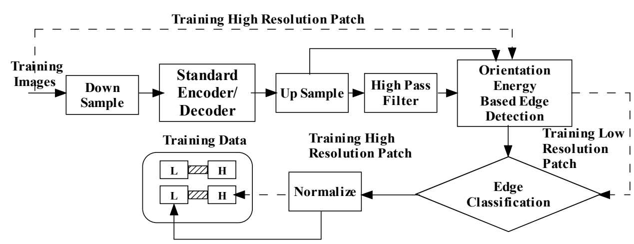



*Figure* 3 Learning Process.

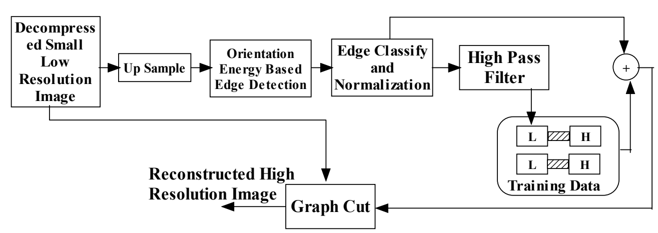



*Figure* 4 Mapping Process.

*Figure* 3 shows the detailed learning process. The orientation energy based edge detection [6] is used here to determine the primal sketch region. Edge Classifier is adopted to categorize the high frequency parts of low resolution and high resolution images respectively into several categories according to their edge types and orientations. Dashed lines in *Figure* 3 indicate the process of learning high resolution patches, while solid lines are for the process of learning low resolution ones. A low resolution patch and a high resolution patch on the same position are stored as a pair in the training set. We use ANN search [7] to locate a potential candidate.

*Figure* 4 presents a more specific procedure of the mapping process shown in *Figure* 1. It is almost the same as the learning process. To make the mapped primal sketch region much smoother, patches are partly covered. However, the decision of pixel values in overlapped regions is a problem to be handled. To reduce annoying artifacts in these regions, graph cut [8] is adopted here as a post process.

An advantage of patch mapping is that we provide a very natural means of specifying image transformations. Rather than select from different filters, we can simply supply an appropriate exemplar without caring too much mid process.

## 4. PROCESSING ON PRIMAL SKETCH

Why do we perform learning and mapping process on primal sketch? Sun et al. have demonstrated in [5], for non-distorted images, patch primitives in contour regions are in low dimensionality. Here, we adopt the similar metric to verify whether this attribute still remains in our scheme. Let $x$ denote an original patch, and $x'$ be its most similar patch in terms of pixel values. We use the metric $e(x) = \|x - x'\| / \|x\|$ to evaluate the effectiveness of our patch mapping process. For a given match error $e$, the hit rate $h$ represents the percentage of test data whose match errors are less than $e$. Receiver Operating Characteristic (ROC) curves (*Figure* 5) show the relationship between $e$ and $h$. Obviously, at a given match error, the higher the hit rate represents the better generalization of the training data, which means the primitives in that regions are in low dimensionality.

To test the effectiveness of our scheme, $9 \times 9$ patches are extracted along contours to form primal sketch regions. Remaining part is named as non primal sketch regions. The two ROC curves shown in *Figure* 5 indicate that patches in primal sketch regions result in a higher hit rate under the same match error and same training set size, comparing with patches outside primal sketch regions. For example, when the training set contains about $5 \times 10^6$ pairs of patches, the match error in the primal sketch region is less than 0.2 for 90% patches. However, for patches in non primal sketch regions, match error is about 0.5 for 70% patches under the same training set size. Therefore, the primitives in primal sketch regions are in low dimensionality. It is reasonable to perform the mapping process only in primal sketch regions. Moreover, primal sketch regions contain more high frequency information. If these parts can be recovered better, visual quality will thus be improved.

## 5. EXPERIMENT RESULTS

We have tested our approach on a set of generic images. Daubechies 9/7 wavelet filters are used here for down-sampling ( $1/2 \times 1/2$ ) and up-sampling ( $2 \times 2$ ). The JPEG2000 coding method [10] is utilized to compress the down-sampled low resolution images. The training set consists of 24 images [9], and about 4,000,000 patch pairs are extracted from the test images to create the training set. All patch pairs are classified into 48 classes (3 types * 16 orientations) according to the type and orientation information given by the orientation energy detection scheme [6]. To ensure satisfactory results, same distortion process on the training images must be guaranteed. Then we use ANN tree search algorithm [7] to map the potential candidate.

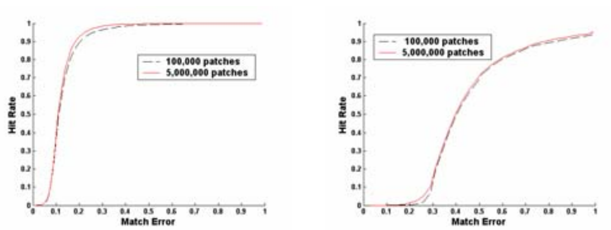



*Figure* 5 ROC curves for primal sketch region mapping (left) and non primal sketch region mapping (right).

We compare our approach with JPEG2000 [10]. *Figure* 7 shows the subjective result of *Lena* ( $512 \times 512$ ) at 0.1bpp. We also compare our scheme with classical Bi-cubic up-sampling. *Figure* 8 presents more results of various content images. It is shown that our scheme achieves better overall subjective visual quality, especially for the parts around detailed regions. We have encouraging results in Lena's hair, the contour of baby's face, the fringe of the tulips, and border of the books. Besides, *Figure* 6 demonstrates that our scheme also outperforms JPEG2000 in terms of PSNR at low bit-rate, where the image *Lena* is tested.

## 6. CONCLUSION

In this paper, a primal sketch based image coding is presented. We use patches on the contours to recover the high frequency information which has been removed. Knowledge learned from other generic images is utilized to enhance visual quality of a target image. Experimental results show that our scheme outperforms JPEG2000 in terms of visual quality and PSNR at low bit-rate.

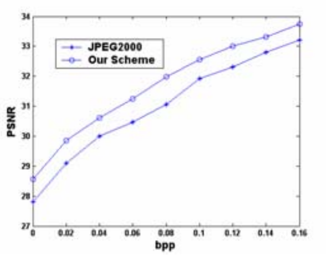



*Figure* 6 PSNR comparisons.

## ACKNOWLEDGEMENTS

We are grateful to Jian Sun for sharing the software. The discussion among us is really helpful. Also, we thank Chunbo Zhu for the generous assistance on revising this paper.

## REFERENCES

[1] W. Lin, L.Dong, "Adaptive downsampling to improve image compression at low bit rates", *IEEE Trans. on Image Processing*. Vol.15, Issue.9, Sept. 2006.

[2] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Super-Resolution Reconstruction of Compressed Video Using Transform-domain Statistics", *IEEE Trans. on Image Processing*, vol. 13, No. 1, Jan. 2004.

[3] S. C. Park, M. G. Kang, C. A. Segall, and A. K. Katsaggelos, "Spatially Adaptive High Resolution Image Reconstruction of DCT-Based Compressed Images", *IEEE Trans. on Image Processing*, vol. 13, No. 4, Apr. 2004.

[4] S. G. Mallat, and S. Zhong. "Characterization of Signals From Multiscale Edges". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.14, No. 7, pp. 710-732, July 1992.

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

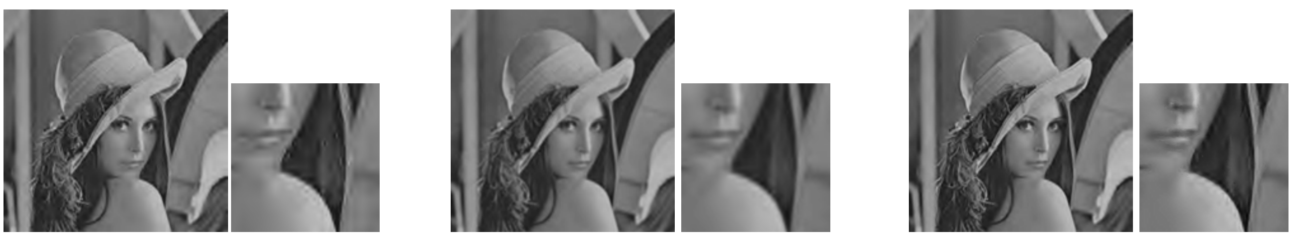
[6] P. Perona, and J. Malik, "Detecting and localizing edges composed of steps, peaks and roofs". *Computer Vision, 1990. Proceedings, Third International Conference on Publication*, pp. 52-57. Dec. 1990.

[7] D. Mount, and S. A. Ann: Library for approximation nearest neighbor searching. http://wwww.cs.umd.edu/mount/ANN.

[8] V. Kwatra, A. Schodl, I. Essa, G. Turk, and A. Bobick, "Graphcut textures: image and video synthesis using graph cuts," *ACM SIGGRAPH* 2003, pp. 277–286.

[9] http://r0k.us/graphics/kodak/

[10] D. S. Taubman, M. W. Marcellin, "JPEG2000 Image Compression Fundamentals, Standards and Practice". *Kluwer Academic Publishers*. 2001.



(a) (b) (c)

*Figure* 7 *Lena* at 0.1 bpp. (a) JPEG2000 (PSNR: 30.46dB). (b) Bicubic Up-sampling (PSNR: 28.75dB) (c) Our scheme (PSNR: 31.25dB).

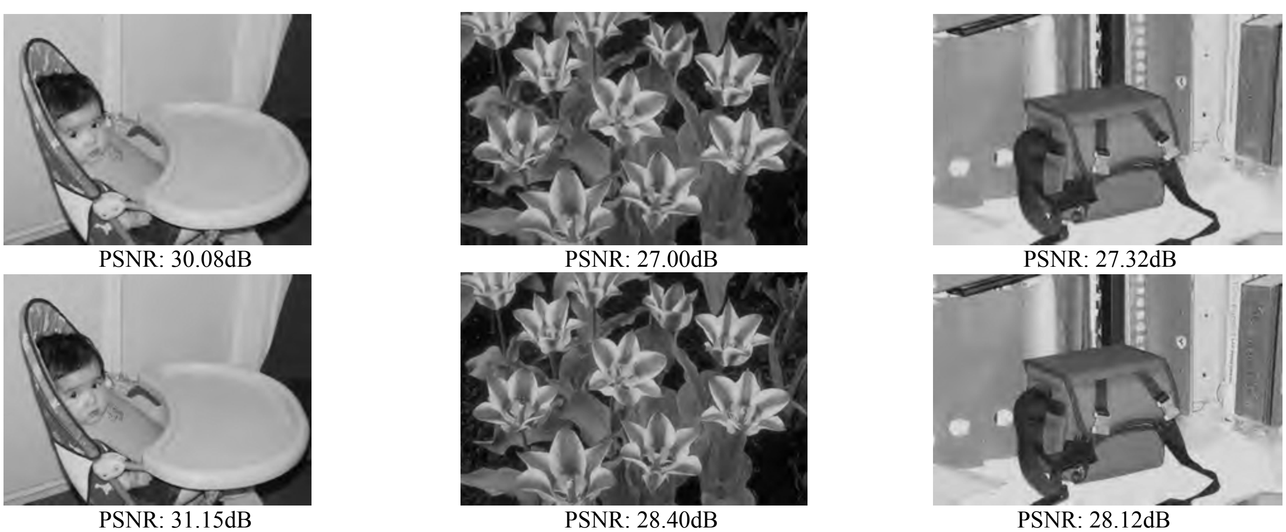



*Figure* 8 Top row: JPEG2000 (0.1bpp). Bottom row: our scheme (0.1bpp).