SUPER-RESOLUTION FOR LOW QUALITY THUMBNAIL IMAGES

Zhiwei Xiong[1], Xiaoyan Sun[2], Feng Wu[2]*

[1]Dept. of Electronic Science and Technology, Univ. of Science and Technology of China, Hefei, China
[2]Internet Media Group, Microsoft Research Asia, Beijing, China
Email: xzw@mail.ustc.edu.cn, xysun@microsoft.com, fengwu@microsoft.com

ABSTRACT

This paper proposes a single-image super-resolution scheme for enlarging low quality thumbnail images widely distributed on the web, which are often generated by downsampling plus compression. To obtain visually pleasurable high-resolution versions for this kind of low-resolution images, we first adopt a PDE-based image regularization technique to alleviate the compression noise in the distorted thumbnails, and then use learning-based pair matching to further enhance the high-frequency details in the upsampled images. Experimental results show that our solution achieves better visual quality for both offline and online test images, compared with traditional methods.

Index Terms— super-resolution, compression, image regularization, pair matching, thumbnail

1. INTRODUCTION

Super-resolution (SR) is the technique achieving high-resolution (HR) enlargements of pixel-based low-resolution (LR) images. There are two kinds of SR, according to the amount of LR images utilized: multi-image SR, which requires several LR images of the same scene to be aligned in sub-pixel accuracy, and single-image SR, which generates a HR image from a unique source. In this paper, we mainly focus on the single-image SR.

Single-image SR has various applications in real world. For example, many images on the web are in their LR forms (often called "thumbnails"); so as to reduce the response time of opening these web pages. This process is especially important for image search engines, since a large number of corresponding images need to be displayed simultaneously once a query is entered. In most cases, users need to click on the thumbnail to get its original HR version. Nevertheless, sometimes it is rather frustrating that the source image is removed or the server is not available. Single-image SR, however, can save users from the bother of linking to every source image if only an enlarged preview is desired. In other words, users can obtain an approximate HR image through SR only from the thumbnail. Then, they can make their own decision whether to go for the original HR version.

Previous work on single-image SR can be categorized into three classes: interpolation-based [1], [2], reconstruction-based [3], [4] and learning-based [5]-[8]. Despite of great diversity in implementation, these methods have a common premise, that the LR images are only downsampled from the original HR images. However, this is not always the truth in practical applications. To further reduce the thumbnail size on the web page while still retaining its resolution at a certain level, compression is often adopted after downsampling as a supplement. For image search engines, compression can reduce the total size of thumbnail by up to 50% without obvious loss of perceptual quality in the LR form, which greatly shorten the response time. However, once SR (any of the above) is performed in this case, compression artifacts will be magnified out and the visual quality of resulting HR images are poor.

To solve the above problem, we propose a novel SR scheme for low quality thumbnail images, mainly those suffering from compression distortion. In our solution, the compression quality parameter (QP) is first obtained from the thumbnail, followed by adaptive image regularization for denoising; with learning-based pair matching on the upsampled image we get the final HR version. In the whole process, we aim at eliminating the undesired high-frequency noise and enhancing the desired high-frequency details.

The rest of this paper is organized as follows. In Sec. 2 we formulate the SR problem in the compression scenario. The framework of our solution is described in Sec. 3, along with the details of each procedure. Experimental results are presented in Sec. 4, and Sec. 5 concludes the paper.

2. SUPER-RESOLUTION UNDER COMPRESSION

The single-image SR problem can be viewed as restoring a HR image Y from a LR input X. For learning-based approaches, a set of training data Z obtained from one or more HR images and their LR versions is also utilized. As pointed out in [6], these approaches should work since a collection of image pixels are special signals that have much less va-

[*] This work has been done during Z. Xiong's internship at Microsoft Research Asia.

riability than a set of completely random variables, and high-frequency details that do not exist in X can thus be “stolen” from Z by exploiting certain regularities. Generally we can formulate the learning-based SR as:

$$Y(X) = Y^*(X) + M(Y^*, Z), \quad Y^*(X) = (X * G_u) \uparrow \alpha \qquad (1)$$

The final HR image Y has two main components: an upsampled version of X as Y^* and a high-frequency enhancing mask generated by M, the pair matching process searching corresponding high-frequency details of Y^* from Z. (Some reconstruction constraints may be further enforced on Y, which we do not formulate apparently here.) G_u denotes an upsampling filter and α is the upsampling rate. “$*$” stands for the convolution operator.

It is obvious that the resulting quality of Y can be improved given more carefully designed G_u and M, as well as the selection of Z. Accordingly, there are literatures of various adaptive interpolation, different pair matching and components learning. However, it has rarely been investigated that what influence the quality of X can have on Y.

Suppose X is a distorted version of original thumbnail X_0. Taking DCT-based compression for example, there is:

$$X = X_0 + E \qquad (2)$$

E represents the quantization error in the spatial domain introduced by compression, which can be approximated as a zero-mean Gaussian random vector with spatially varying and correlated covariance matrix C_E [9]:

$$E \sim N(0, C_E) \qquad (3)$$

$N(\mu, C)$ denotes a normal distribution with mean vector μ and covariance matrix C. Due to the presence of E, the SR result $Y(X)$ must have a difference compared with $Y(X_0)$, which behaves as distortion:

$$D = Y(X) - Y(X_0) = (Y^* + M(Y^*, Z)) - (Y_0^* + M(Y_0^*, Z)) \qquad (4)$$

where $Y^* = Y^*(X)$, $Y_0^* = Y^*(X_0)$. Assume the upsampling filter G_u is linear, then

$$D = (E * G_u) \uparrow \alpha + M(Y^*, Z) - M(Y_0^*, Z) \qquad (5)$$

Figure 1(a) shows a two-step result of learning-based SR under JPEG compression (QP = 60, α = 3). We can see that both upsampling and pair matching suffer from artifacts, leading to poor quality of the HR images.

To alleviate the influence of the compression noise E, a denoised version of X is required. Here we adopt a PDE-based image regularization method as proposed in [10]. Suppose $X : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^n$ is a vector-valued image, the PDE updating velocity can be locally estimated by applying a spatially varying Gaussian smooth mask $\mathbf{G}^{(\mathbf{T})}$ over X:

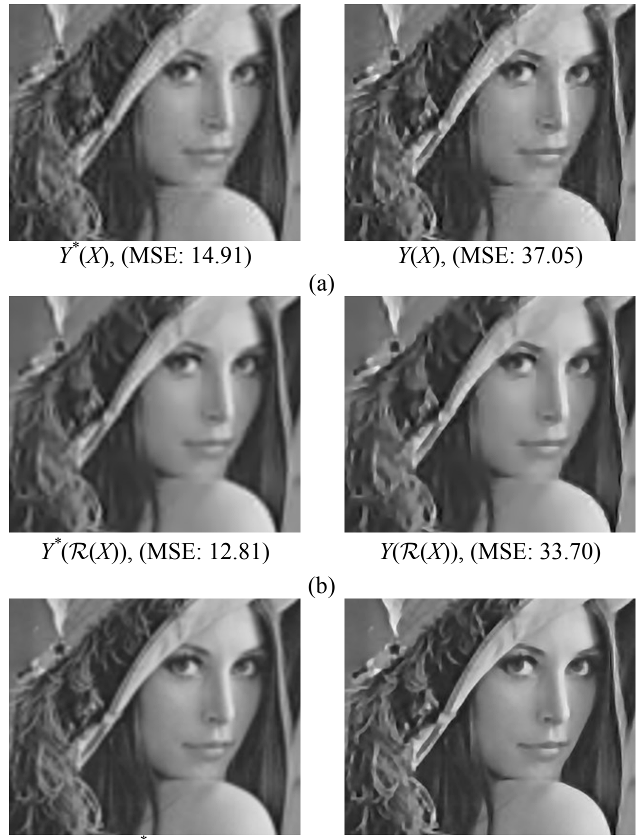
$$\Delta X_k = \sum\nolimits_{i,j=-1}^{1} \mathbf{G}^{(\mathbf{T})}(i,j) X(x-i, y-j), \quad k = 1, \cdots, n \qquad (6)$$

$\mathbf{G}^{(\mathbf{T})}$ is defined by:

$$\mathbf{G}^{(\mathbf{T})}(\mathbf{x}) = \frac{1}{4\pi} \exp\left(-\frac{\mathbf{x}^T \mathbf{T}^{-1} \mathbf{x}}{4}\right) \quad with \quad \mathbf{x} = (x, y)^T \qquad (7)$$

$\mathbf{T}$ is a 2×2 tensor field defined pointwise as:

$$\mathbf{T} = f_+\left(\sqrt{\lambda_+^* + \lambda_-^*}\right) \theta_-^* \theta_-^{*T} + f_-\left(\sqrt{\lambda_+^* + \lambda_-^*}\right) \theta_+^* \theta_+^{*T} \qquad (8)$$



$Y^*(X)$, (MSE: 14.91) $Y(X)$, (MSE: 37.05)

(a)

$Y^*(\mathcal{R}(X))$, (MSE: 12.81) $Y(\mathcal{R}(X))$, (MSE: 33.70)

(b)

$Y^*(X_0)$ $Y(X_0)$

(c)

Figure 1. Super-resolution results under compression. From left to right: upsampling via bicubic interpolation, pair matching via image hallucination [7]; (a) compressed, without regularization; (b) compressed, regularized; (c) anchors without compression.

$\lambda_\pm^*$ are the two eigenvalues of the Gaussian smoothed structure tensor $\mathbf{G}_\sigma = \sum\nolimits_{k=1}^{n} \nabla X_k \nabla X_k^T * G_\sigma$, and $\theta_\pm^*$ are its corresponding eigenvectors (θ_+^* stands for the direction of local maximum variation). $f_\pm$ are two weighting functions chosen as $f_+ = 1/(1+s^2)$ and $f_- = 1/\sqrt{1+s^2}$. The regularization process can be finally formulated by:

$$\mathcal{R}(X) = X + \frac{t_0}{\max(|\Delta X|)} \Delta X, \quad X = (X_1, \cdots, X_n) \qquad (9)$$

where t_0 is a positive constant representing the updating step.

Figure 1(b) shows the corresponding result with regularized SR. It can be easily observed that artifacts after upsampling and pair matching are effectively reduced. Meanwhile, we calculate the two-step distortion D in the MSE form, compared with the result under no compression (as shown in Figure 1(c)). The lower MSE in (b) indicates that the objective image quality is also improved if regularization is performed on the thumbnail first. (Note that the MSE of upsampling and pair matching are calculated with respective

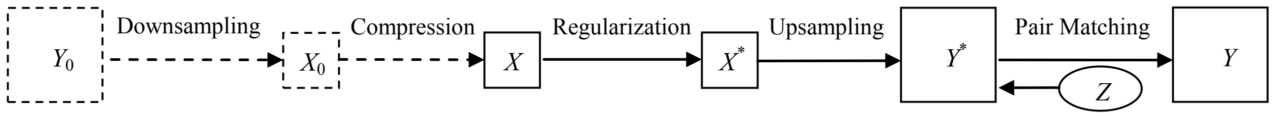


Figure 2. Framework of super-resolution under compression

anchors and the comparison between them is meaningless.)

Also, we can see the edges are observably enhanced in the results after pair matching. We choose learning-based pair matching since it "borrows" high-frequency details form a training database without compression noise. Other SR methods, such as sharpened interpolation and reconstruction-based approaches, are totally dependent on the thumbnail. They may magnify the remaining noise after regularization, which counteracts the effect of denoising.

3. FRAMEWORK OF OUR SOLUTION

The framework of our SR scheme is shown in Figure 2. An original image Y_0 is first downsampled by a low-pass filter to form a LR version X_0. Compression is then performed on X_0, resulting in a distorted thumbnail X, which is the input of our SR system. The system consists of three main modules: adaptive image regularization, upsampling and pair matching. Regularization is performed on X to get an artifacts-relieved LR image X^*. X^* is then upsampled via bicubic interpolation, resulting in Y^*, the intermediate HR result. The final HR image Y is obtained with pair matching from Y^* and a training database Z.

Since most images on the web are in JPEG format, we mainly discuss the DCT-based compression in this paper. It works the same way for other image formats and compression styles. Given a low quality thumbnail (here "quality" refers to the QP of JPEG), we first parse the quantization table from the JPEG file header. It is then compared with the standard table to get an estimate of QP.

3.1. Adaptive image regularization

The PDE-based image regularization mentioned above performs selective smoothing along the direction of large variations and thus has the advantage of preserving global features (such as edges) in images while denoising. However, it is iterative and time-consuming. That's why we put regularization before upsampling, where the image resolution is still low. Moreover, we find out an empiristic formula to calculate the least iteration times under a certain QP, so as to minimize the computational complexity.

Given a thumbnail X with QP = q ($0 \le q \le 100$), the visual quality of the final HR image Y will no longer have obvious improvement after m iterations:

$$m = \begin{cases} \lceil 9(100-q)/40 \rceil & 60 \le q \le 100 \\ 9, & 0 \le q < 60 \end{cases} \tag{10}$$

3.2. Learning-based pair matching

Bicubic interpolation resizes the regularized thumbnail X^* to the desired resolution. Still, high-frequency details are missing and the upsampled image Y^* appears to be hazy. To further enhance its visual quality, we use pair matching in image hallucination [7] to restore some "primal sketches".

The pair matching step goes as follows: first an edge map is detected from Y^*. Then Y^* is high-pass filtered and patches along the edges are extracted. These patches (actually mid-frequency) are used to search high-frequency details from the database Z via nearest neighbor matching. Z consists of a number of patch pairs; each pair comprises a mid-frequency data and a corresponding high-frequency data as defined in [7].

4. EXPERIMENTAL RESULTS

We test our SR scheme on both offline images generated via downsampling followed by compression, and real distorted thumbnails on the web. For the former, we use Gaussian filtering for downsampling ($1/3 \times 1/3$) and JPEG for compression (QP = 60); for the latter, the downsampling method and ratio are unknown, while compression is still JPEG with different QP. The PDE updating step in regularization is 5.0, and the upsampling rate is 3×3. A 5M database is used in pair matching, which is trained on the 24 *Kodak* images from [11].

A portion of offline example *Lena* (512×512) is already shown in Figure 1, from which we can have a comparison between learning-based SR with and without regularization, for both the subjective visual quality and objective MSE metric. In Figure 3 we give some results of distorted thumbnails on the web from *Live image search* [12]. They are compared with image hallucination, bicubic and regularized bicubic interpolation, respectively. The first two results in Figure 3 exhibit the effectiveness of adaptive regularization for compressed thumbnails. Our approach greatly reduces the compression noise while the high-frequency details are still well preserved and enhanced. The last two results demonstrate the necessity of the learning-based pair matching process in our solution, as the single interpolation, whether regularized or not, fails to give as clear edges as we do.

In addition, the overall complexity of the proposed SR scheme is low enough for real time application on the web. We have an online test on a large number of thumbnails from [12] (where the thumbnail size is around 160×160), and the average SR time for a single image is less than 1s.

Hallucination | Bicubic | Regularized Bicubic

Our approach | Our approach | Our approach

Figure 3. Super-resolution results of distorted thumbnails on the web

5. CONCLUSION

In this paper, we present a single-image SR scheme in the compression scenario, which integrates image regularization and learning-based pair matching together to generate HR enlargements for low quality thumbnail images. Image regularization alleviates the compression artifacts and pair matching restores the high-frequency details. Experimental results prove the effectiveness of this combination. Due to its low complexity, our approach provides a practical solution for the enlarged preview of compressed LR images on the web, especially the numerous thumbnails provided by image search engines.

The performance of our solution can be further improved if taking into consideration of adaptive interpolation as well as other image enhancing techniques. We would like to study on these issues later.

REFERENCES

[1] J. Allebach and P. W. Wong, "Edge-directed interpolation," in *Proc. Int. Conf. Image Processing*, 1996, pp. 707-710.

[2] L. Xin and M. T. Orchard, "New edge-directed interpolation," in *Proc. Int. Conf. Image Processing*, 2000, pp. 311-314.

[3] M. Irani and S. Peleg, "Motion analysis for image enhancement: resolution, occlusion and transparency," *Journal of Visual Communication and Image Representation*, vol. 4, pp. 324-335, 1993.

[4] B. S. Morse and D. Schwartzwald, "Image magnification using level-set reconstruction," in *Proc. Computer Vision and Pattern Recognition*, 2001, pp. 333-340.

[5] W. T. Freeman and E. C. Pasztor, "Learning low-level vision," in *Proc. Int. Conf. Computer Vision*, 1999, pp. 1182 – 1189.

[6] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, March-April 2002.

[7] J. Sun, N. Zheng, H. Tao, and H. Shum, "Image hallucination with primal sketch priors," in *Proc. Computer Vision and Pattern Recognition*, 2003, pp. 729–736.

[8] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. Computer Vision and Pattern Recognition*, 2004, pp. 275–282.

[9] M.A. Robertson and R.M. Stevenson, "DCT quantization noise in compressed images," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 27–38, Jan. 2005.

[10] D. Tschumperlé and R. Deriche, "Vector-valued image regularization with PDE's: a common framework for different applications", in *Proc. Computer Vision and Pattern Recognition*, 2003, pp. 651-656.

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

[12] http://www.live.com/?&scope=images