Volume 0 (1981), Number 0 pp. 1–14 COMPUTER GRAPHICS *forum*

Visual-Quality Optimizing Super Resolution

Feng Liu[1] Jinjun Wang[2] Shenghuo Zhu[2] Michael Gleicher[1] Yihong Gong[2]

[1]Department of Computer Sciences, University of Wisconsin-Madison
[2]NEC Laboratories America, Inc.

Abstract

In this paper, we propose a robust image super-resolution algorithm, which aims to maximize the overall visual quality of super-resolution results. We consider a good super-resolution algorithm to be fidelity preserving, image detail enhancing and smooth. Accordingly, we define perception-based measures for these visual qualities. Based on these quality measures, we formulate image super-resolution as an optimization problem aiming to maximize the overall quality. Since the quality measures are quadratic, the optimization can be solved efficiently. Experiments on a large image set and subjective user study demonstrate the effectiveness of the perception-based quality measures and the robustness and efficiency of the presented method.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation Display algorithms; I.4.3 [Image Processing and Computer Vision]: Enhancement Sharpening and deblurring

1. Introduction

In this paper, we consider creating an image of higher resolution than the provided input image. In this *single image super-resolution* problem, we aim to create the high resolution result such that it has better quality than a straightforward upsampling of the source image. For many applications, such as television and computer graphics, the ultimate measure of quality is visual. Therefore, in this paper we introduce methods that optimize the visual quality of the high resolution results.

The challenge of single image super-resolution results from estimating more pixel values than the given. This requires reconstructing the image such that more samples can be taken. Sampling theory suggests an "ideal" reconstruction that avoids adding information beyond the source data. However, such a reconstruction is ideal only in an information theoretic sense: upsampled images lack detail and appear blurry. For us, an ideal image is one that is perceived to be high-quality by a viewer, even if the details are not an accurate reconstruction of the original scene.

The goal of a visually appealing result suggests a different approach for creating the high resolution result. We define mathematically the visual quality of the resulting image. A *visually ideal* reconstruction is one that optimizes these metrics. We show how such an optimization can be posed and solved efficiently and robustly to provide a practical and effective method for upsampling images that provides better visual quality than previous approaches. Because our ultimate goal is the subjective visual quality, we mainly assess our results empirically, providing a user study that confirms the effectiveness of our approach.

Our approach, like other single image super-resolution methods, relies on additional information beyond the pixel samples of the source image. However, whereas previous approaches relied on either sets of example images or assumptions about the imaging process (§2), our approach adds information based on a model of what is visually appealing. This means that our approach is not only more likely to achieve our stated goal of visual quality, but also is less prone to artifacts from inappropriate or insufficient example sets, or from incorrect imaging assumptions.

The central contribution of this paper is a new approach to single image super-resolution that explicitly considers perceived visual quality. In §3.1-§3.3 we detail three perceptually inspired metrics for reconstruction quality, and in §3.4 we describe how these metrics can be optimized efficiently to create super-resolution images. Results of our prototype implementation are evaluated empirically in §4, where we also compare our method to many other representative methods.

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

2. Related Work

Image super-resolution (SR) methods create high resolution results (HR) from lower resolution inputs (LR). Methods consider two distinct problems: synthesizing the HR from a single LR source image, or assembling an HR from multiple LR source images. Our approach addresses the single-image problem. The multi-image problem, requires very different methods, see [BS98b, FREM04, PPK03] for surveys.

Single image super-resolution must upsample the provided image, effectively reconstructing the underlying image and sampling it at a greater frequency. An ideal reconstruction (in the signal processing sense) avoids aliasing, the addition of high frequencies that could not be represented in the LR source. Ideal reconstructions can be approximately effective using polynomial interpolation or truncated reconstruction kernels. Upsampling based on such methods (especially bicubic and Lanczos) is ubiquitous and extensively studied (c.f. [UAE95]). However, because the approximation to the ideal reconstruction aims to avoid aliasing, it cannot create sharp features in the result, and may exhibit ringing (due to the Gibbs phenomena). Methods to sharpen image features have been presented. For example, Polesel et al. presented an adaptive unsharp masking method to enhance the image contrast [PRM00]. Many super-resolution methods can produce sharp features, even though such details are not fully resolved in the source.

To create details in the HR result beyond those resolved in the input, super-resolution methods rely on additional information beyond the input. When that additional information is inadequate, the methods fail. For example, edge directed interpolation methods guide interpolation along edges [AW96, JA95, LO01, DHX*07], inferring the details based on heuristics of edge localization and shape. However, poor edge identification can lead to artifacts. Alternatively, back projection methods, introduced by Irani and Peleg [IP91, IP93], assume the Point Spread Function (PSF) is known and reverse it by iterative projection. However, poor approximation to the PSF can lead to serious ringing artifacts. In contrast, our approach rarely creates objectionable artifacts.

Many recent *example-based* super-resolution techniques rely on data from other images to inform the addition of details. Variants have learned the co-occurrence prior between HR and LR image patches or coefficients [STS03, JJC04, JS06, CYX04], or image feature statistics [Fat07], and process the LR input along with an appropriate smoothness constraint [FPC00] to generate the HR image. Examples of example-based methods include: Baker et al. [BS98a] develop recognition-based super-resolution algorithm where the cost function includes the results of a set of recognition decisions to enforce the condition that the gradient in the HR image should be equal to the gradient in the best matching training image. Capel et al. [CZ01] proposed a super-resolution technique from multiple views using principal component analysis to learn image models either to directly constrain the maximum likelihood (ML) estimate or as a prior for a maximum a posteriori (MAP) estimate. Freeman et al. introduced a parametric Markov network to learn the statistics between the "scene" and the "image" as a framework for handling low-level vision tasks such as super-resolution [FPC00]. Wang et al. extended Freeman's framework to a Conditional Random Field [WTS05]. Learning on coefficient domain is also reported, such as wavelet coefficients [JJC04]. Jiji et al. [JS06] introduced an edge-based super-resolution method to learn the correlation of the image contourlet between HR/LR images and argued that no global dependencies need to be considered if the contourlet feature is used.

Example-based super-resolution is limited by the training set. If the set is too small, examples cannot be found. If the set is too large, the algorithm may use an inappropriate example. For this reason, example-based methods are effective in constrained domains, such as faces [LSZ01, GBA*03, WT04, LLT05, DKA04] and fingerprints [JC04], where a domain specific training set can be used. In contrast, our approach is general and does not rely on a training set.

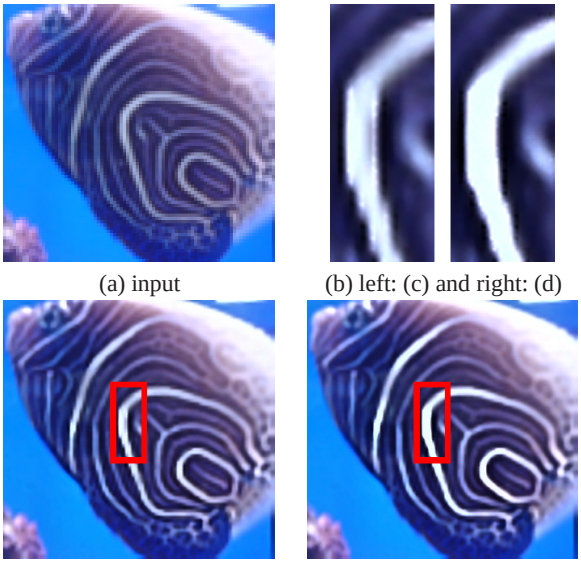
3. Our approach

Our goal is to create super-resolution results that have good perceived visual quality. To achieve this, we must quantify the subjective notion of visual quality as metrics that can be computed on the images for the specific case of image super-resolution. We have identified three criteria that correspond to perceived qualities of super-resolution image results:

- Fidelity preserving. The result should have the same general appearance as the input.
- Detail enhancing. The result should have sharp features where they are expected.
- Smoothness. The result should have continuity where it is expected, and avoid unnatural high-frequency artifacts.

The first criterion ensures that the result looks like a higher resolution version of the same image. The latter criteria are inspired by the observation that people are very sensitive to high-frequency contrast [Lam91, IKN98, IK01, Not00]. They expect to see crisp edges, but also expect smoothness between the discontinuities.

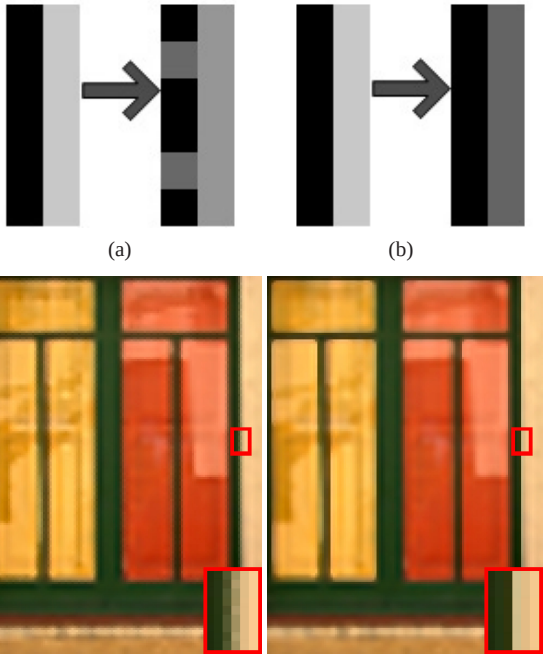
We encode violations against each of the above quality aspects as a cost, and formulate the super-resolution as an optimization problem which aims to find a required high-resolution image minimizing the total cost. The key to success is to define effective quality measures. Since visual quality assessment is a subjective task, we propose perception-based measures as described in the following subsections.



(a) input (b) left: (c) and right: (d)

(c) neighborhood: 3×3 (d) neighborhood: 5×5

Figure 1: *Neighborhood affect. A small neighborhood sometimes can lead to zigzag artifacts (c). Increasing the neighborhood size can relieve this problem (d).*



(a) (b)

(c) Ignore masking effect (d) Consider masking effect

Figure 2: *Masking effect. (Please refer to the electronic version to appreciate the difference between the left and the right.)*

3.1. Fidelity-preserving

The goal of fidelity-preserving is to ensure that the super-resolution result has a similar general appearance to the input. A straightforward way is to encourage the result to interpolate the input pixel values. Denoting the resulting high-resolution image as I^h and low-resolution image as I^l, we define the following fidelity measure:

$$E_{fd1} = \sum_{p(x,y) \in I^l} \|I^h(x \times scale, y \times scale) - I^l(x,y)\|_2^2 \quad (1)$$

where $p(x,y)$ is a pixel in the low-resolution input I^l, and $scale$ is the magnification rate.

The above fidelity measure alone is not sufficient to guarantee preserved fidelity. According to research from visual perception and neuro-science, the human visual system is more sensitive to local contrast than to pixel values [Lam91, IKN98, IK01, Not00]. Local contrast in images can be approximated by their gradient fields. So besides the measure of Equation 1, we encourage the gradient field of the high-resolution image to be close to that of the real world high-resolution image. The challenge is to obtain the gradient field of the real world high-resolution image. Since bicubic up-sampling result provides a close approximation to the real world high-resolution image at least perceptually, we use its gradient field as the reference. The second fidelity measure we use is defined as follows:

$$E_{fd2} = \sum_{p \in I^h, \theta} \|G^h(p,\theta) - G^{app}(p,\theta)\|_2^2 \quad (2)$$

where $G^h(p,\theta)$ is the high-resolution image's gradient value at pixel p in direction θ, and $G^{app}(p,\theta)$ is that of the approximation of the real world gradient field. In practice, the orientation is quantized into 8 bins corresponding to each pixel's 8-connected neighborhood. Occasionally, this small neighborhood can lead to zigzag artifacts. Finer quantization can be used to relieve this problem at the expense of more computation as illustrated in Figure 1.

This uniform data-independent measure is not effective however. According to research in perception, change in the high gradient value direction is less obvious than the same amount of change in the low gradient value one. This phenomenon is called the "masking effect" [KK95]. For example, as illustrated in Figure 2, the amount of change from the left image to the right in (a) is the same as that in (b) according the measure of Equation 2. However, the change in (a) is more obvious perceptually.

Concerning the contrast masking, a good approximation of the human visual system response on contrast can be modeled by a transducer function(c.f. [Wil80]). We adopt an approximation of the Weber's law [VV90] to account for the masking effect by weighting the gradient change in inverse proportion to the gradient magnitude as follows. The advantage of this approximation is that it leads to a simple

quadratic optimization problem.

$$E_{fd2} = \sum_{p \in I^h, \theta} \frac{\|G^h(p,\theta) - G^{app}(p,\theta)\|_2^2}{\|G^{app}(p,\theta)\|_2^2 + \varepsilon} \quad (3)$$

where ε is a constant for the sake of numerical stability, set as 1.0 in our work. The effect of considering the masking effect is illustrated in Figure 2(c) and (d).

3.2. Detail-enhancing

Preserving and/or enhancing image details is one of the major focuses of image super-resolution methods. Many existing methods enhance image details by sharpening image edges [AW96, JA95, LO01, TTT06, DHX*07, Fat07]. Edge directed methods [AW96, JA95, TTT06] estimate high-resolution edges from low-resolution input, and use the edge information to guide super-resolution operations, such as interpolation and image reconstruction. The performance of these methods are subject to the quality of high-resolution edge estimation, which is hard. A recent method [Fat07] learns edge statistics to guide super-resolution. A fundamental problem with edge guided methods is that edges are not always enough to represent image details. For example, details in image regions with rich fine textures are hard for edges to describe.

From the view of visual perception, details manifest themselves to the low-level human visual system by local contrast. Hence, we propose preserving/enhancing image details by enhancing local contrast instead of edges. The textureness criteria defined by Bae et al. [BPD06] can be a good local contrast measurement. We calculate the local contrast in each image patch as the sum of difference between every two pixels since this popular definition leads a quadratic term in the following Equation 4, which is easy to optimize. We define the following measure to enhance the detail:

$$E_{dt} = -\sum_{patch_k \in I^h} w_k \sum_{p_i, p_j \in patch_k} \|I^h(p_i) - I^h(p_j)\|_2^2 \quad (4)$$

where $patch_k$ denotes an image patch in high-resolution image I^h. w_k is a weight. As illustrated in the top row of Figure 3(b) and (c), in our system, each block is of size $(scale+1) \times (scale+1)$, with four corners corresponding to the 4 neighboring pixels in the low-resolution input. This neighborhood is chosen to bound the interaction between pixels locally. In practice, a larger neighborhood is easier to reduce the fidelity of the super-resolution result. There are several options to set the weights. For example, w_k can be a binary variable, set as 1 when there is an edge passing through the patch. The edge can be estimated from the bi-cubic up-sampling result using the Canny algorithm [Can86]. To set w_k, no accurate information about the edge location inside the patch is required, which improves the tolerance of our method against error in edge estimation. The advantage of this strategy is robustness to noise. When the input image quality is good, another way is to set w_k as the contrast inside the low resolution counterpart of $patch_k$.

3.3. Smoothness

Smooth results are usually favored by the human visual system. A popular way to guarantee smoothness is to encourage neighboring pixels to have similar values as follows:

$$E_{sm} = \sum_{(x_i,y_i) \in I^h} \sum_{(x_j,y_j) \in N_8(x_i,y_i)} \|I(x_i,y_i) - I(x_j,y_j)\|_2^2 \quad (5)$$

where $N_8(x_i,y_i)$ is the 8-connection neighborhood of pixel (x_i,y_i). In the scenario of super-resolution, however, using the smoothness constraint in Equation 5 is dangerous. It can blur real image edges and diminish details. We use a similar approach to the Tikhonov regularization commonly used in reconstruction based super-resolution methods [KS93]. Since bi-cubic up-sampling result already provides an over-smoothed version of a real world scene, we encourage the super-resolution result to be close to the bi-cubic up-sampling one as follows:

$$E_{sm} = \sum_{(x,y) \in I^h} \|I^h(x,y) - I^b(x,y)\|_2^2 \quad (6)$$

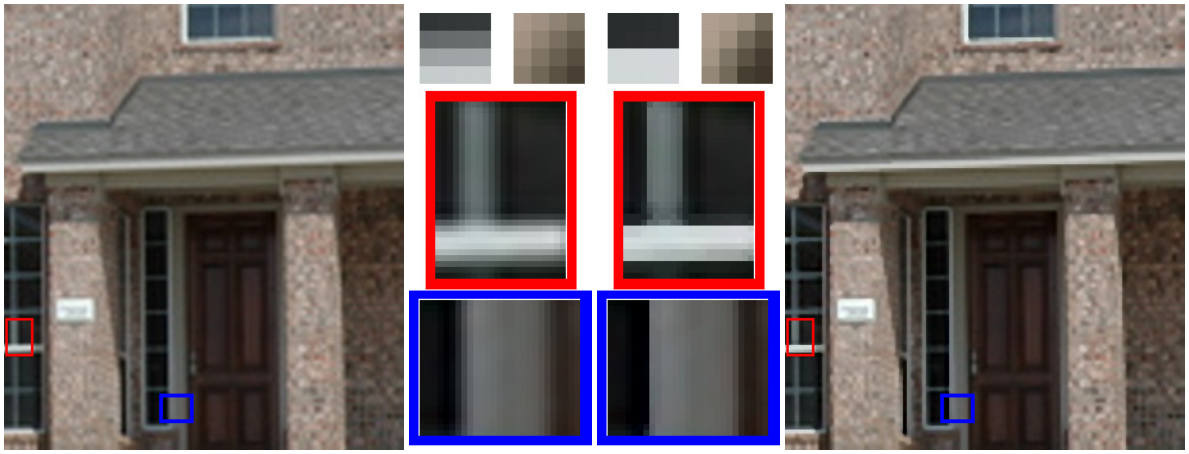
where I^b is the bi-cubic up-sampling result. The above scheme provides necessary smoothness while avoiding drastic over-smoothness. This specific smoothness term contains one of the fidelity-preserving terms defined in Equation 1, which encourages the high-resolution image close to the low-resolution input. These two terms are used for different purposes. Equation 1 is used to preserve fidelity, while Equation 6 is used to achieve smooth results. Practically, although using Equation 6 can also help achieve preserving fidelity, it affects all the pixels, which is too strong a constraint. Equation 1 only constrains on grid points, allowing pixels inside each grid to vary and achieve enhancing details. Also, other smoothness terms can be used. For example, as shown in our user study described in Section 4, although the majority of users do not like over-smooth results, some do. To provide more smoothness while keeping the trade-off between smoothness and sharpness manageable, we propose another alternative smoothness term, which aims to minimize the Laplacian of the high-resolution image as follows:

$$E_{sm} = \sum_{(x,y) \in I^h} \| \frac{\partial^2 I^h}{\partial x^2} + \frac{\partial^2 I^h}{\partial y^2} \|_2^2 \quad (7)$$

3.4. System solver

Based on the quality measures defined in the above subsections, we formulate image super-resolution as an optimization problem by linearly combining all the measures as follows:

$$E = \lambda_{fid1} E_{fid1} + \lambda_{fid2} E_{fid2} + \lambda_{dt} E_{dt} + \lambda_{sm} E_{sm} \quad (8)$$



(a) Bicubic ×3 (b) Bicubic zoom in (c) Ours zoom in (d) Ours×3

Figure 3: *Increasing local contrast for detail preserving/enhancing.*

where $\lambda_?$ is the weight for each term. Currently we set these weights empirically, and the default settings are $\lambda_{fid1} = 1.0$, $\lambda_{fid2} = 100.0^{\dagger}$, $\lambda_{dt} = 1.0$ and $\lambda_{sm} = 1.0$. Since the meaning of each weight is intuitive and directly related to a high-level super-resolution property, users can also personalize the super-resolution operation by changing the default parameters. (It should be noted that giving a very large weight to E_{dt} will induce a non-positive definite matrix from Equation 8 since E_{dt} is negative when solving the system as described in the following paragraph. In practice, we do not find it a problem during experiments, since a very big weight to this term leads to unattractive results.)

We calculate image gradients and Laplacian using finite difference methods. Since all the measures are at most quadratic, the above problem is a quadratic minimization problem. We can re-formulate Equation 8 into the following linear system:

$$A_{n\times n} I^h_{vn} = b_n \tag{9}$$

where n is the size of the high-resolution image I^h, I^h_v is the vector representation of I^h, A is a $n \times n$ matrix, and b is a n dimensional vector. For a 500 × 500 image, A can be 0.25M by 0.25M, which makes solving the linear system potentially expensive. However, in the above measures, since each pixel only interacts with its local neighborhood, A is a narrow banded sparse diagonal matrix. Exploiting this sparsity provides efficient solving.

We solve the system using a sparse implementation of the pre-conditioned conjugate gradient method [BBC*94], where the pre-conditioner is obtained by sparse incomplete Cholesky factorization [Saa96]. It takes about 4 seconds to create a 500 × 500 image. The solution can be out of the range of [0, 255]. We find from experiments that simply clamping the solution will not hurt the result perceptually. Otherwise, the quadratic system of Equation 8 with the bound constraint can be efficiently solved using a linear constrained least squares solver (LCLS) [LBAD*06].

4. Evaluation

Evaluating the visual quality of super-resolution results is difficult. Objective measures, such as the root mean square error (RMSE) and signal-to-noise ratio (SNR), are popular in evaluating super-resolution results. However, they have difficulty measuring visual quality [LO01]. Although there have been many efforts on image quality assessment, no gold standard method has been proposed [DVKG*00, WSB02, TLZZ04, WBSS04, WS05, WWS*06]. The eye-ball scheme is often adopted too to appreciate the super-resolution results. In this paper, we employ all these methods to evaluate our method. Moreover, since image visual quality assessment is a subjective task, we designed a web-based user study for further evaluation.

4.1. Subjective evaluation

We designed a web-based user study to evaluate the performance of our algorithm. The goal of this study was to determine how our algorithm compares to others in creating high-quality images. In this study, the results of the presented algorithm were compared to those of the other algorithms. To make the user study tractable, the number of total trials assigned to each participant should be reasonable. This requires us to select a small number of algorithms to compare with on a small number of images. Meanwhile, to guarantee

† λ_{fid2} is significantly big because E_{fid2} has a big denominator.

Figure 4: *Screen copy of the user study.*

the effectiveness of this study, we need to select representative methods and representative images.

In this study, besides the presented algorithm, we selected another 4 representative algorithms to compare. They are nearest neighbor interpolation (NN), bi-cubic interpolation (Bicubic), back projection (BP) [IP91, IP93] and the soft-edge algorithm (SEDGE) [DHX*07]. NN is selected to test if a super-resolution algorithm performs better than nothing. The bi-cubic method is selected because it is the most popular method used in practice. It can create smooth results, however it often blurs the true image edges. The BP algorithm is selected for its representativeness of methods based on some assumption of PSF functions. It can create visually faithful results, however it can introduce ringing artifacts. SEDGE is one of the most recently published methods. It is also a representative of edge directed methods. SEDGE algorithm can create crisp edges as well as smooth edge profiles, however it is dependent on the performance of edge detection and the assumption that edge profiles shall be smooth. Inaccurate edge detection or invalid assumptions can lead to such artifacts as overly smooth edge profiles and flat small image regions.

We experiment with the above 5 algorithms on an image set that contains 600 images and covers a large variety of image categories, including animals, trees, cars, planes, buildings, human faces, etc. In our experiments, we use all the five algorithms to create for each image in the image set a 3x and 4x high-resolution result. In this way, each algorithm creates 1200 results. Some of the results are illustrated in Figures 8, 9 and 11. (**Please refer to the electronic version for better visual quality.**)

In this study, the results of the presented algorithm were compared to those of the other 4 algorithms. For each pair of comparisons, we randomly selected 10 images from the image set. So each participant did 40 trials. In each trial, he/she is shown 2 super-resolution images created by two different algorithms with the same input. Which algorithm's result was shown on the left or right is randomized, as was the order of trials. Each participant was asked to select "*the image they think has better quality*" by simply clicking the image. A screen copy of the user study website is shown in Figure 4.

ours v.s.	images	mean	std. error.	p value
NN	10	9.82	0.07	$\ll$ 1.0e-4
Bicubic	10	7.18	0.39	$\ll$ 1.0e-4
BP	10	8.84	0.23	$\ll$ 1.0e-4
SEDGE	10	9.10	0.24	$\ll$ 1.0e-4

Table 1: *User study result.*

In total, 49 participants participated in the study. These participants consisted of graduate students in varying majors, and employees in varying companies.

To assess the presented algorithm, we counted the number of trials where the super-resolution result from our algorithm was selected. On average, participants chose our result over a NN result 9.82 out of 10 times(98.2%), over a bi-cubic result 7.18 out of 10 times (71.8%), over a BP result over 8.84 out of 10 times (88.4%), and over a SEDGE result over 9.10 out of 10 times (91.0%). These results suggest that our algorithm is preferred to the other methods. We computed the significance using a t-test(n = 49), and found all the p-values to be smaller than 1.0e-4, which shows the preference of our algorithm over NN, Bicubic, BP, and SEDGE. The user study result is reported in Table 1.

We looked into each comparison and found from the comparison between bi-cubic and our algorithm that 16.3% of participants consistently prefer the bi-cubic up-sampling results. We sent to these participants the questionnaire "*why do you prefer the results on the left to the right*" together with the link which displays the images they selected on the left and the corresponding unselected one on the right. Most of the responses we received show that they prefer smooth results rather than those with sharp edges. This also partially explains why our algorithm is preferred to SEDGE, which creates sharper edges than ours.

Overall, this subjective user study suggests that our algorithm creates higher visual-quality super-resolution results than others. An important finding we obtain from this study is that users' preferences vary over individuals. A non-trivial percent of human participants prefer smooth results rather than sharp ones. This finding again supports our proposal to use subjective user study to evaluate super-resolution algorithms.

4.1.1. Robustness test

We manually examined all the results to check the robustness of our algorithm. We consider a result to be a failure if it has obvious artifacts. The failure rate for our algorithm, BP and SEDGE are 2.5% and 21.2% and 32.0% respectively. The typical artifacts of each algorithm are shown in Figure 11. Although this evaluation is biased, the significantly smaller failure rate of our algorithm suggests its relative robustness. This result actually is consistent with the subjective user study.

4.2. Objective evaluation

It is well recognized that objective metrics such as the root mean square error (RMSE) and signal-to-noise ratio(SNR) have difficulty measuring visual quality [LO01]. If we approximate the visual quality of a super-solution result with its similarity to an ideal up-sampling result, these objective metrics can still be used as a rough quality measurement. This partially justifies why they are still the popular measurements used in practice. In this study, besides the RMSE, we use the Structural SIMilarity (SSIM) index for image quality assessment [WBSS04]. The SSIM is based on the assumption that human visual perception is highly adapted for extracting structural information from a scene. It measures the degradation of structural information during transforming one image to another, and is a popular method in measuring image distortion.

We calculate the RMSE and SSIM values of all the 1200 images created by the 5 super-resolution algorithms studied in the previous section. The result is reported in Table 2. According to these objective measures, our method performs better than the NN, BP and SEDGE methods. However, there is no significant difference between Bicubic and ours. Meanwhile, we notice that NN is significantly better than BP and SEDGE, which seems to conflict with our subject impression that BP and SEDGE perform better than NN in most cases. Typical examples are shown in Figure 8. This confirms that current objective metrics have difficulty measuring visual quality [LO01].

	NN	Bicubic	BP	SEDGE	Ours
RMSE	16.96	14.52	18.96	17.53	14.75
SSIM	0.688	0.734	0.685	0.668	0.736

Table 2: *Objective metrics. By RMSE, smaller values indicate better performances; For SSIM, larger values indicates better performances.*

4.3. More comparison

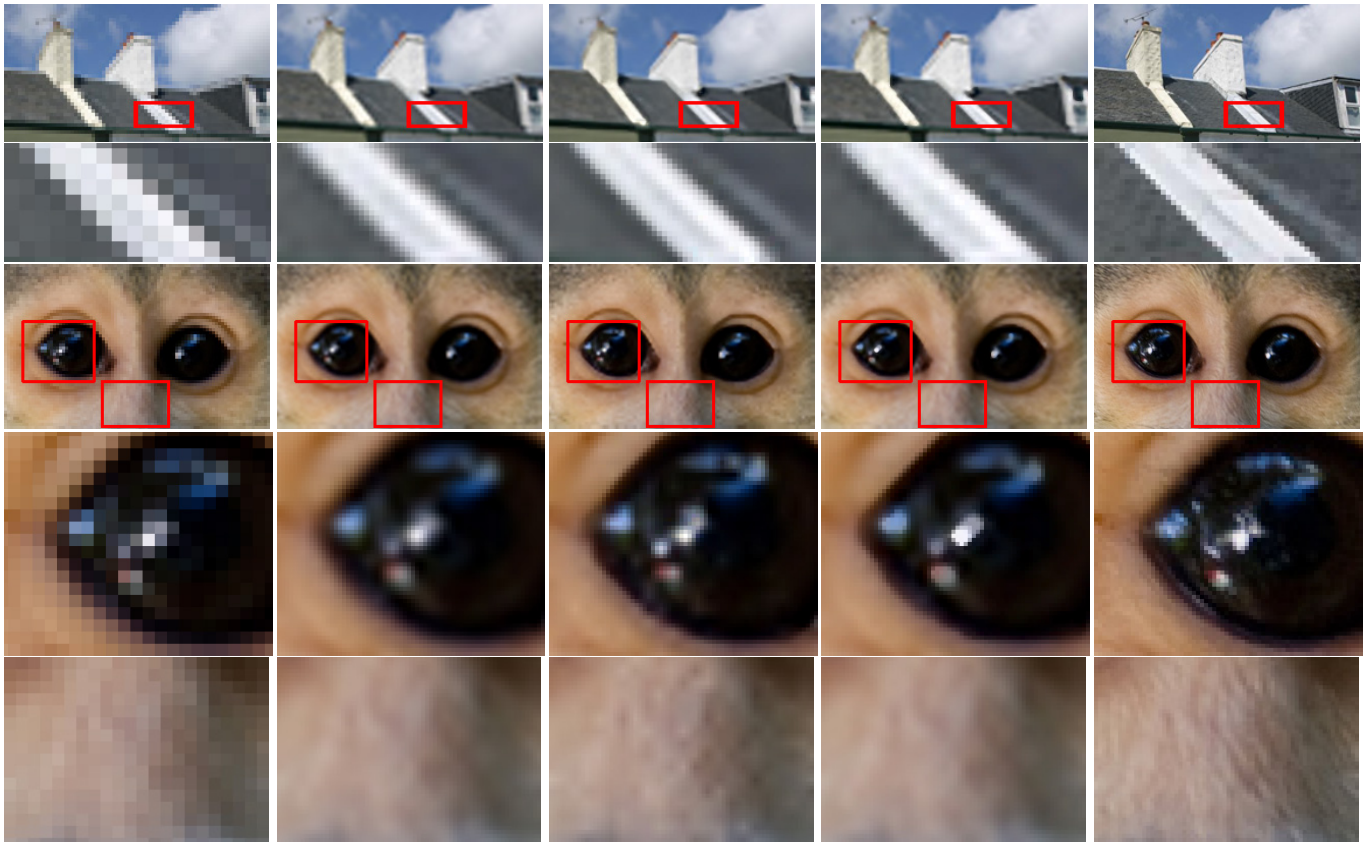
To further appreciate the performance of the presented method, we compare it to the Sharp Bicubic method [Ado] and more recent methods, including *new edge-directed interpolation* (NEDI) [LO01], *image hallucination* (IH) [STS03], *image upsampling via imposed edge statistics* (IES) [Fat07], *nonlinear enhancement algorithm* (NLE) [GAA00] and *multiple regressors algorithm* (MR) [TRF04]. These comparisons are illustrated in Figures 5, 6 and 7.

These examples demonstrate that the presented method can achieve up-sampling results comparable to all these advanced methods. Compared to these methods, one particular advantage of our method is its flexibility for users. As confirmed in our user study, users' preferences vary over individuals. Because the quality measures in our algorithm directly correspond to user requirements, our algorithm supports a convenient user interface for them to personalize their super-resolution operations. For example, increasing the weight of the detail enhancing term can lead to sharp results as illustrated in Figure 10.

5. Conclusion

In this paper, we present a robust perception-based image super-resolution algorithm. The main contributions are the visual-quality maximizing framework for super-resolution and the design of perception-based image visual quality measures in the scenario of super-resolution. Besides being effective, these quadratic and local quality measures enable efficient processing. Our subjective user study confirms that the presented algorithm can create visually appealing results. The user study also shows that users' preference to the super resolution results is diverse. Since each quality measure is an intuitive property description of a super resolution result, this presented algorithm supports users to personalize the super-resolution operation by changing the default parameter setting. Moreover, new measures can be easily added into the proposed framework.

A major concern of the presented method is its lack of



(a) NN ×3 (b) Bicubic ×3 (c) IH ×3 (d) Ours ×3 (e) Ground truth

Figure 5: *Super-resolution examples. These examples compare our algorithm to the image hallucination(IH) [STS03].*

control on the global edge profile smoothness, since it does not use a mid-level vision representation, such as edges. However, as we argued in previous sections, using mid-level or high-level information automatically obtained from analysis on low-resolution input is a double-edged sword. For example, automatic control of the edge smoothness is difficult. Fine details along edges could be the manifest of true image characteristics, and they could also be rasterization/sampling artifacts. Our experiments confirm that in practice, although advanced edge smoothing algorithms, such as [DHX*07], can create good results, they can also create objectionable artifacts. Also, current methods for extracting mid-level vision representation are not reliable enough. The presented method achieves stable results over a wide variety of images at the expense of control of the global edge smoothness. Therefore, we consider the presented algorithm a candidate for general image editing software.

Although we argued that using mid-level/high-level analysis of the low-resolution input could be risky, providing such an option for users could still be helpful. So an important extension to the proposed algorithm is a component which considers the inter-patch interaction and encourages smooth edges. A promising way is to extend the covariance based edge adaptation scheme [LO01] by encouraging high-resolution covariance to be consistent with its low-resolution counterpart.

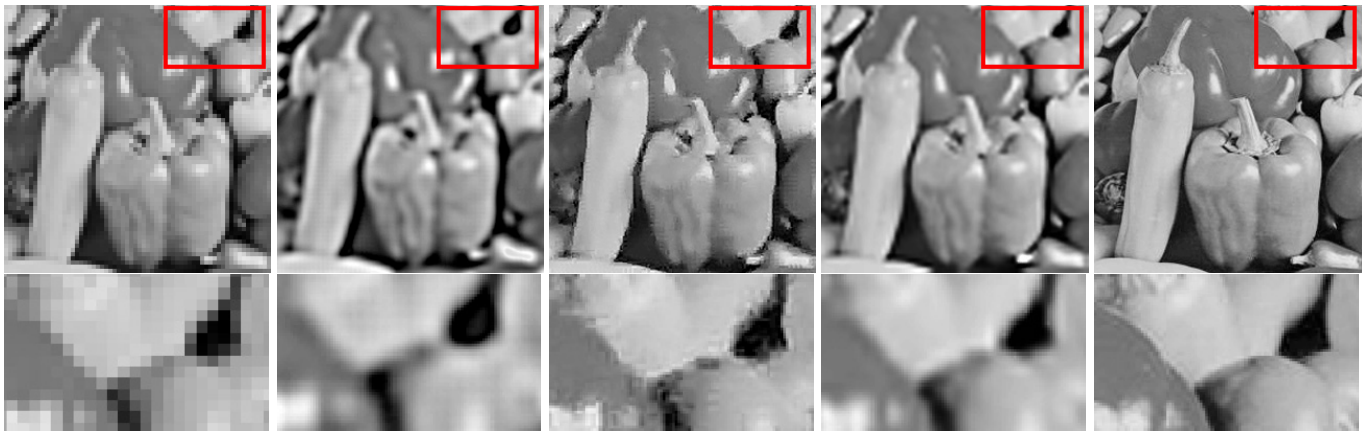
Acknowledgements We would like to thank reviewers for their constructive suggestions and Dr. Kai Yu for helpful discussions on this project. We also thank Dr. Jian Sun for his help with the experiments.

References

[Ado] ADOBE SYSTEM INC.: Adobe Photoshop CS3.

[AW96] ALLEBACH J., WONG P. W.: Edge-directed interpolation. In *IEEE ICIP* (1996), pp. 707–710.

[BBC*94] BARRETT R., BERRY M., CHAN T. F., DEMMEL J., DONATO J., DONGARRA J., EIJKHOUT V., POZO R., ROMINE C., DER VORST H. V.: *Templates for the Solution of Linear Systems: Building Blocks for It-*



(a) NN ×4 (b) NLE ×4 from [TRF04] (c) MR ×4 from [TRF04] (d) Ours ×4 (e) Ground truth

Figure 6: *Super-resolution example. This example compares our algorithm to* nonlinear enhancement algorithm*(NLE) [GAA00] and* multiple regressors algorithm*(MR) [TRF04].*

erative Methods, 2nd Edition. SIAM, Philadelphia, PA, 1994.

[BPD06] BAE S., PARIS S., DURAND F.: Two-scale tone management for photographic look. In *ACM SIGGRAPH* (2006), pp. 637–645.

[BS98a] BORMAN S., STEVENSON R.: Spatial resolution enhancement of low-resolution image sequences. a comprehensive review with directions for future research. *Lab. Image and Signal Analysis, University of Notre Dame, Tech. Rep.* (1998).

[BS98b] BORMAN S., STEVENSON R.: Super-resolution from image sequences-a review. *Proc. of Midwest Symposium on Circuits and Systems* (1998), 374–378.

[Can86] CANNY J.: A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence 8*, 6 (1986), 679–698.

[CYX04] CHANG H., YEUNG D.-Y., XIONG Y.: Super-resolution through neighbor embedding. In *IEEE CVPR* (2004), pp. 275–282.

[CZ01] CAPEL D., ZISSERMAN A.: Super-resolution from multiple views using learnt image models. In *Proc. IEEE CVPR 2001* (December 2001), pp. 627–634.

[DHX*07] DAI S., HAN M., XU W., WU Y., GONG Y.: Soft edge smoothness prior for alpha channel super resolution. In *IEEE CVPR* (2007).

[DKA04] DEDEOGLU G., KANADE T., AUGUST J.: High-zoom video hallucination by exploiting spatio-temporal regularities. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2004), pp. 151–158.

[DVKG*00] DAMERA-VENKATA N., KITE T., GEISLER W., EVANS B., BOVIK A.: Image quality assessment based on a degradation model. *IEEE Transactions on Image Processing 9*, 4 (2000), 636 – 650.

[Fat07] FATTAL R.: Image upsampling via imposed edge statistics. In *ACM SIGGRAPH* (2007).

[FPC00] FREEMAN W. T., PASZTOR E. C., CARMICHAEL. O. T.: Learning low-level vision. *IJCV*, 1 (2000), 25–47.

[FREM04] FARSIU S., ROBINSON D., ELAD M., MILANFAR P.: Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology 14*, 2 (2004), 47–57.

[GAA00] GREENSPAN H., ANDERSON C., AKBER S.: Image enhancement by nonlinear extrapolation in frequency space. *IEEE Transactions on Image Processing 9*, 6 (2000), 1035 – 1048.

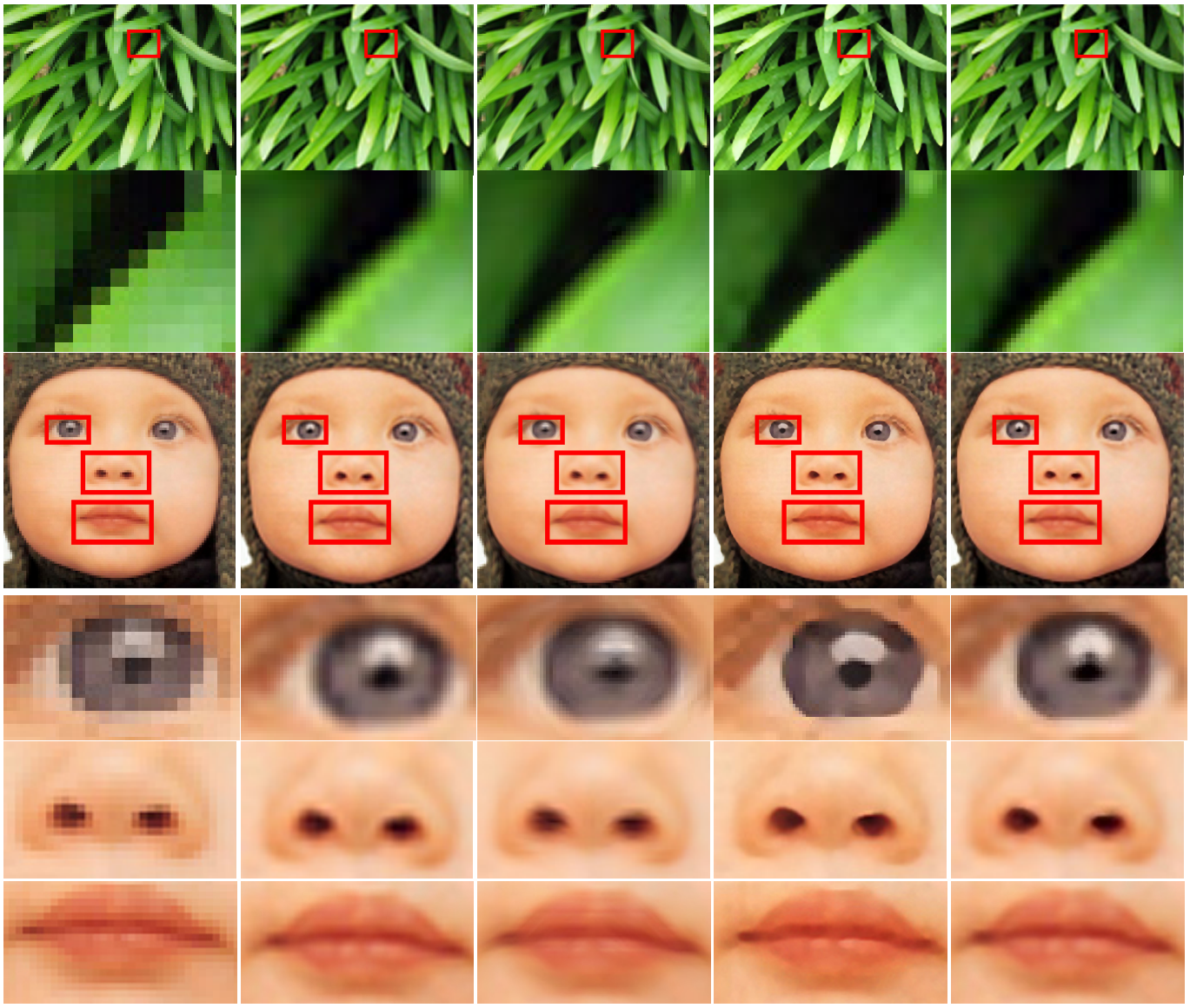
[GBA*03] GUNTURK B., BATUR A., ALTUNBASAK Y., HAYES M., M.MERSEREAU R.: Eigenface-domain super-resolution for face recognition. *IEEE Transactions on Image Processing*, 5 (2003), 597–606.

[IK01] ITTI L., KOCH C.: Computational modeling of visual attention. *Nature Reviews Neuroscience 2*, 3 (Mar 2001), 194–203.

[IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell. 20*, 11 (1998), 1254–1259.

[IP91] IRANI M., PELEG S.: Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, 3 (1991), 231–239.

[IP93] IRANI M., PELEG S.: Motion analysis for image enhancement: resolution, occlusion, and transparency.



(a) NN ×4 (b) Sharp Bicubic ×4 (c) NEDI ×4 (d) IES ×4, from [Fat07] (e) Ours ×4

Figure 7: *Super-resolution examples. These examples compare our algorithm to PhotoShop sharp bicubic [Ado], NEDI [LO01] and IES [Fat07].*

Journal of Visual Communication and Image Representation, 4 (1993), 324–335.

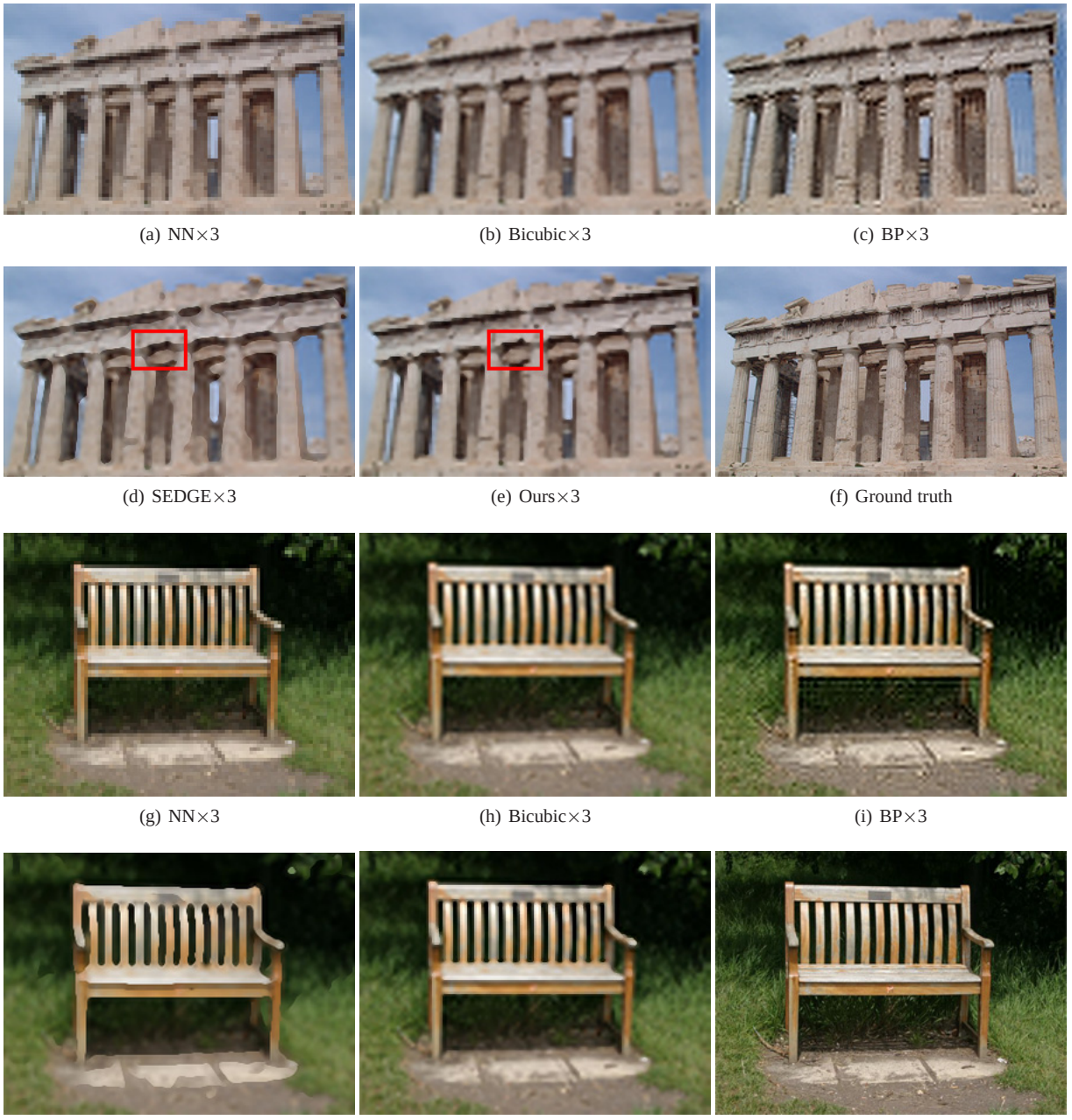
[JA95] JENSEN K., ANASTASSIOU D.: Subpixel edge localization and the interpolation of still images. *IEEE Trans. on Image Processing 4* (Mar. 1995), 285–295.

[JC04] JIJI C. V., CHAUDHURI S.: Pca-based generalized interpolation for image super-resolution. *Proc. of Indian Conference on Vision, Graphics & Image Processing'04* (2004), 139–144.

[JJC04] JIJI C. V., JOSHI M. V., CHAUDHURI S.: Single-frame image super-resolution using learned wavelet coefficients. *International Journal of Imaging Systems and Technology*, 3 (2004), 105–112.

[JS06] JIJI C., SUBHASIS C.: Single-frame image super-resolution through contourlet learning. *EURASIP Journal on Applied Signal Processing* (2006), 73767(11).

[KK95] KARUNASEKERA S., KINGSBURY N.: A distortion measure for blocking artifacts in images based on hu-



(a) NN×3 (b) Bicubic×3 (c) BP×3

(d) SEDGE×3 (e) Ours×3 (f) Ground truth

(g) NN×3 (h) Bicubic×3 (i) BP×3

(j) SEDGE×3 (k) Ours×3 (l) Ground truth

Figure 8: *Super-resolution results from 5 algorithms. Our algorithm creates reasonably sharp edges as shown in (e) and (k), and creates high contrasty local structure details as shown in the top of building in (e). BP creates visually faithful results, especially in the rich texture regions; however it introduces ringing artifacts along strong edges as shown in(c) and (i). SEDGE creates crisp and smooth edges; however the smooth edge diminishes the structure detail as shown in the top of building in (e). Also SEDGE creates flat image regions, such as the ground under the chair in (j).*

(a) NN ×4 (b) Ours ×4

Figure 9: *More examples.*

(a) $\lambda_{dt} = 0.2$ (b) $\lambda_{dt} = 1$ (c) $\lambda_{dt} = 2$

Figure 10: *Super-resolution result with different detail-enhancing weights* λ_{dt}.

(a) Ours×4 (b) BP×4 (c) SEDGE×4

Figure 11: *Typical artifacts of S.R. algorithms. (a) Our method introduces zigzag artifacts along long edges, (b) BP introduces ringing artifacts, and (c) SEDGE creates over smooth edge profiles and flat small image regions.*

man visual sensitivity. *IEEE Transactions on Image Processing 4*, 6 (June 1995), 713–724.

[KS93] KIM S., SU W.-Y.: Recursive high-resolution reconstruction of blurred multiframe images. *IEEE Transactions on Image Processing 2* (4 1993), 534 – 539.

[Lam91] LAMMING D.: Contrast sensitivity, chapter 5,. *Cronly-Dillon, J., Vision and Visual Dysfunction 5* (1991).

[LBAD*06] LAWRENCE J., BEN-ARTZI A., DECORO C., MATUSIK W., PFISTER H., RAMAMOORTHI R., RUSINKIEWICZ S.: Inverse shade trees for non-parametric material representation and editing. *ACM Trans. Graph. 25*, 3 (2006), 735–745.

[LLT05] LIU W., LIN D., TANG X.: Hallucinating faces: Tensorpatch super-resolution and coupled residue compensation. In *IEEE CVPR* (2005), pp. 478– 484.

[LO01] LI X., ORCHARD M.: New edge-directed interpolation. *IEEE Transactions on Image Processing 10*, 10 (Oct. 2001), 1521–1527.

[LSZ01] LIU C., SHUM H. Y., ZHANG C. S.: Two-step approach to hallucinating faces: global parametric model and local nonparametric model. In *IEEE CVPR* (2001), pp. 192–198.

[Not00] NOTHDURFT H.: Salience from feature contrast: additivity across dimensions. *Vision Research 40*, 11-12 (2000), 1183–1201.

[PPK03] PARK S. C., PARK M. K., KANG M. G.: Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine* (2003), 21–36.

[PRM00] POLESEL A., RAMPONI G., MATHEWS V.: Image enhancement via adaptive unsharp masking. *IEEE Transactions on Image Processing 9*, 3 (Mar. 2000), 505–510.

[Saa96] SAAD Y.: *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, 1996.

[STS03] SUN J., TAO H., SHUMF H.: Image hallucination with primal sketch priors. In *Proc. IEEE CVPR'03* (2003), pp. 729–736.

[TLZZ04] TONG H., LI M., ZHANG H.-J., ZHANG C.: No-reference quality assessment for jpeg2000 compressed images. In *IEEE ICIP* (2004), pp. 24–27.

[TRF04] TAPPEN M., RUSSELL B., FREEMAN W.: Efficient graphical models for processing images. In *IEEE CVPR* (2004), pp. 673–680.

[TTT06] TAI Y.-W., TONG W.-S., TANG C.-K.: Perceptually-inspired and edge-directed color image super-resolution. In *IEEE CVPR 2006* (2006), pp. 1948–1955.

[UAE95] UNSER M., ALDROUBI A., EDEN M.: Enlargement or reduction of digital with minimum loss of information. *IEEE Trans. Image Process*, 3 (Mar. 1995), 247–258.

[VV90] VALOIS R. L. D., VALOIS K. K. D.: *Spatial Vision*. Oxford University Press, 1990.

[WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing 13*, 4 (Apr. 2004), 600–612.

[Wil80] WILSON H. R.: A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics 38*, 3 (1980), 171–178.

[WS05] WANG Z., SIMONCELLI E. P.: An adaptive linear system framework for image distortion analysis. In *IEEE ICIP* (2005), pp. 1160–1163.

[WSB02] WANG Z., SHEIKH H., BOVIK A.: No-reference perceptual quality assessment of jpeg compressed images. In *IEEE ICIP* (2002), pp. Vol I: 477–480.

[WT04] WANG X., TANG X.: Hallucinating face by eigentransformation with distortion reduction. *Proc. of ICBA'04* (2004), 88–94.

[WTS05] WANG Q., TANG X., SHUM H.: Patch based blind image super resolution. In *Proc. of ICCV'05* (2005), no. 1, pp. 709–716.

[WWS*06] WANG Z., WU G., SHEIKH H., SIMONCELLI E., YANG E.-H., BOVIK A.: Quality-aware images. *IEEE Transactions on Image Processing 15*, 6 (2006), 1680 – 1689.