Subjective evaluations of example-based, total variation, and joint regularization for image processing

Hyrum S. Anderson[a], Maya R. Gupta[b], and Jon Hardeberg[c]

[a]Sandia National Laboratories, USA;
[b]Department of Electrical Engineering, Univ. of Washington, USA;
[c]Dept. of Computer Science and Media Technology, Gjøvik University, Norway.

ABSTRACT

We report on subjective experiments comparing example-based regularization, total variation regularization, and the joint use of both regularizers. We focus on the noisy deblurring problem, which generalizes image superresolution and denoising. Controlled subjective experiments suggest that joint example-based regularization and total variation regularization can provide subjective gains over total regularization alone, particularly when the example images contain similar structural elements as the test image. We also investigate whether the regularization parameters can be trained by cross-validation, and we compare the reconstructions using cross-validation judgments made by humans or by fully automatic image quality metrics. Experiments showed that of five image quality metrics tested, the structural similarity index (SSIM) correlates best with human judgement of image quality, and can be profitably used to cross-validate regularization parameters. However, there is a significant quality gap between images restored using human or automatic parameter cross-validation.

Keywords: subjective evaluation, deblurring, image quality metrics

1. INTRODUCTION

Given the wide availability of high-quality images, many researchers have studied how one can use example images as training examples to solve image estimation problems. However, even with advanced algorithms, these example-based (EB) approaches can produce objectionable artifacts, as illustrated in Fig. 1, and image quality can be heavily dependent on the choice of parameters. The objectionability of such artifacts is often not accurately captured by simple image quality metrics such as MSE or PSNR. We questioned whether EB methods can result in more visually-appealing reconstructions than the standard approach of using total variation (TV) regularization, but were unable to find comparative experiments in the literature.

In this paper, we address this question with subjective experiments that compare EB, TV, and a joint EB-TV regularizer for the task of deblurring. All three approaches require choosing how much regularization to use. We set these regularization parameters by finding the parameter values that maximize quality on a set of training images. This enables a second experiment: we compare human judgement to five full-reference image quality metrics (IQMs) to cross-validate the regularization parameters.

Our results show that the combined regularization can perform better than either regularizer alone, and that example-based regularization is most effective on images with strong structure and when the example images are similar. In subjective human experiments, we found that parameters can be effectively cross-validated by human judges, but are significantly less optimal when chosen by an IQM; we also found that SSIM correlated best with human judgements of deblurred image quality out of the five IQMs tested for the proposed algorithm.

Further information: contact hander@sandia.gov.

Computational Imaging X, edited by Charles A. Bouman, Ilya Pollak,
Patrick J. Wolfe, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8296, 82960S
 · doi: 10.1117/12.917710

noisy blurred image EB deblurring TV deblurring EB-TV deblurring

Figure 1. Far left: A noisy blurred image. Second from left: an EB reconstruction that exhibits crisp edges but objectionable artifacts. Second from right: a total-variation (TV) reconstruction. Far right: a joint example-based TV (EB-TV) reconstruction (far left). Regularization parameters for the TV and EB-TV images were optimized over a finite set of parameter choices by maximizing the average score of three independent observers, as described in Section 5.1.

2. EXAMPLE-BASED IMAGE PROCESSING

One of the first image processing solutions to explicitly use example images was for the problem of inverse halftoning:[1] halftoned blocks of an image were mapped to continuous-tone values based on training example pairs of continuous blocks and their halftones. Another early direct use of example images was in restoring lossy-compressed images.[2]

Recently, a number of EB methods have focused on the task of single-frame superresolution. Freeman et al. proposed a method that estimates missing high-frequency components of a low-resolution (LR) image by comparing small blocks in the image to a training corpus of high-resolution images. The high-res images are anti-aliased and downsampled, and from them a nearest-neighbor image block is located; the high-frequency edges corresponding to the low-resolution nearest-neighbor match are added to the low-resolution image to be super-resolved.[3] Note that estimating high-frequencies from image blocks with similar mid-frequencies can be effective because edges induce correlation of adjacent frequency bins.[4,5] Since the procedure is performed on a per-block basis, a boundary condition between adjacent blocks is used to regularize the nearest neighbor selection and promote continuity across block boundaries in the reconstructed image. A simple and effective implementation of such a boundary condition is to find the nearest-neighbor image patches in raster-scan order, and overlap each new patch with previously reconstructed image pixels.[3]

Freeman et al. noted that this approach of learning from nearest-neighbor patches can overfit to noise in the test image and create objectionable artifacts in the reconstructed image.[3] When we applied the strict analogue of this EB approach to the deblurring problem, we saw similar objectionable reconstruction artifacts,[6] and the risk of such overfitting artifacts generally increased if we used a larger number of example images. To reduce artifacts, we have previously proposed an optimal combination of the nearest-neighbor reconstruction with Wiener deblurring by treating them as multiple observations in a linear minimum mean squared error (LMMSE) framework.[6] However, that approach requires impractical knowledge about power spectra, assumes inaccurately that the image statistics are stationary, and focuses on maximizing PSNR rather than visual image quality.

A number of improvements to the basic idea of learning from nearest-neighbor patches have been proposed in the superresolution community. Chang et al. presented a method which requires fewer training samples by posing the nearest-neighbor problem similar to manifold learning.[7] Jiji and Chaudhury proposed a method in which high-resolution detail was learned from a database using contourlets.[8] Jiji et al. evaluated various local and global single-frame superresolution reconstruction techniques,[9] and proposed a global PCA-based technique to remove blur, noise, and aliasing artifacts. Ma et al. have shown that example-based superresolution can be effective for error concealment in scalable video coding where previously decoded frames are used to create example images that can be highly similar to the image to be reconstructed.[10]

In this paper we explicitly use examples to regularize the deblurring estimate (as opposed to model-based regularizations that use examples to learn the model parameters, such as the work of Zhu and Mumford[11]). Explicit example-based regularization was perhaps first proposed by Baker and Kanade in 2002, who regularized the gradients of their superresolution solution to match the gradient of the corresponding nearest-neighbor from a set of example images;[12] they termed this *image hallucination*. Datsenko and Elad proposed a regularizer

based on an (image-dependent) example-based prior for a global maximum a posterior objective, and showed good results superresolving scanned documents containing text, equations and graphs,[13] and face images.[14]

3. DEBLURRING WITH EXAMPLE-BASED AND TOTAL VARIATION REGULARIZERS

Given a column-scanned $M \times N$ noisy blurred image $z \in \mathbb{R}^{MN}$ and a known blur matrix $H \in \mathbb{R}^{MN \times MN}$, a TV-regularized least-squares estimate of the original image is the $x^* \in \mathbb{R}^{MN}$ that solves:

$$\min_x \|Hx - z\|_2^2 + \lambda_{\text{TV}} \|x\|_{\text{TV}}, \tag{1}$$

where $\lambda_{\text{TV}} \in \mathbb{R}$ is a regularization parameter that must be specified, and $\|\cdot\|_{TV}$ denotes total variation:

$$\|x\|_{TV} \triangleq \sum_{j,k} \sqrt{|(\nabla_v x)_{j,k}|^2 + |(\nabla_h x)_{j,k}|^2},$$

where the subscripts v and h denote the vertical and horizontal gradients respectively, and the subscript (j,k) indexes the $((k-1)N + j)$th entry of the vector.

Similarly, an EB-regularized least-squares estimate of the original image is the $x^* \in \mathbb{R}^{MN}$ that solves:

$$\min_x \|Hx - z\|_2^2 + \lambda_{\text{EB}} \|x - \hat{x}_{\text{EB}}\|_2^2 \tag{2}$$

where $\hat{x}_{\text{EB}}$ is an example-based reconstruction (our implementation details given below).

EB and TV regularization are different approaches to enhancing image quality, and using them together might be useful. To test this hypothesis as well as compare EB and TV regularization, we defined the following joint EB-TV objective function for deblurring, which has the following features:

- it uses an auxilliary image $v \in \mathbb{R}^{MN}$ for computational efficiency, following the work of Chambolle;[15]
- it has three regularization parameters: $\lambda_{\text{AUX}}, \lambda_{\text{EB}}, \lambda_{\text{TV}}$ corresponding to weights on the auxiliary term, the EB regularizer, and the TV regularizer respectively;
- and the example-based regularization term is defined as $\|x - \sum_i T_i \alpha_i\|^2$, where the columns of each $T_i \in \mathbb{R}^{MN \times k}$ contain the k "most similar" $m \times m$ training blocks to the ith block of the input image z (with entries outside the block set to zero), the ith training blocks are weighted by the corresponding vector α_i, and the α_i are jointly solved for to minimize the error between x and the nearest neighbor reconstruction $\sum_i T_i \alpha_i$. This example-based regularizer is based on the work of Datsenko and Elad[13,14] but differs in that we regularize towards a linear combination of neighborhood blocks rather than summing regularizations to each block, which showed a slight performance advantage in preliminary experiments.

Specifically, the joint EB-TV objective function we used solves for an image estimate $x \in \mathbb{R}^{MN}$, an auxiliary image $v \in \mathbb{R}^{MN}$, and weights on example-blocks $\{\alpha_i \in \mathbb{R}^k\}_{i=1}^B$ such that:

$$\min_{x,v,\{\alpha_i\}_{i=1}^B} \|Hv - z\|_2^2 + \lambda_{\text{AUX}} \|x - v\|_2^2 + \lambda_{\text{EB}} \left\| x - \sum_i T_i \alpha_i \right\|_2^2 + \lambda_{\text{TV}} \|x\|_{\text{TV}}. \tag{3}$$

This paper's subjective experiments use (3) with different values of the regularization parameters:

- $\lambda_{\text{EB}} = 0$ corresponds to TV-only regularization as implemented by Bresson and Chan;[16]
- $\lambda_{\text{TV}} = 0$ corresponds to EB-only regularization, analogous to the work of Freeman et al.[3] and Elad and Datsenko;[14]

- $\lambda_{\mathrm{EB}} > 0$ and $\lambda_{\mathrm{TV}} > 0$ then (3) is a joint EB-TV regularization. A key aspect in the joint case is the interaction of the $\{\alpha_i\}$ in the example-based regularization with the other terms caused by jointly minimizing over x, v, and $\{\alpha_i\}$. Usually in example-based reconstruction, one or more neighbors that are close to z are solved for and combined in some way. Here, the neighbor blocks $\{T_i\}$ are fixed based on closeness to z, but the combination weights $\{\alpha_i\}$ are solved for jointly with x and v, and the example-based reconstruction $\sum_i T_i \alpha_i$ is solved to match the reconstruction x. This interaction tempers the propensity of the example-based reconstruction to add artifacts, because the total variation regularization will cause the reconstruction x to be smoother and have cleaner edges, and thus T_i that are smoother and have cleaner edges will be rewarded with stronger α_i's.

We provide our complete algorithm implementation details in the appendix.

4. EXPERIMENTAL SET-UP

In this section, we describe the details of the experimental set-up. In the next section, we describe the actual experiments and results.

4.1 Image Distortion

For all experiments, each sharp image was blurred by convolving it with a Gaussian blur kernel with a 2.25 pixel bandwidth and adding Gaussian white noise such that the PSNR with respect to the original sharp image was 34 dB PSNR. For the reconstructions where $\lambda_{\mathrm{EB}} > 0$, example image blocks were extracted from all of the the images in the training set if reconstructing a test image, or if a training image was being reconstructed (to cross-validate parameters), then the example image blocks were extracted from only those images in the training set that were not being reconstructed (that is, at no point did we deblur an image using examples from the same image).

For all experiments, each color plane of the original 24 bit RGB color image was independently subjected to blur and noise. Wen et al. have investigated color-plane-dependent TV deblurring,[17] but to make for a more valid comparison with the IQMs, we converted the corrupted RGB image to the YCbCr color space, and only deblur the Y plane. Because the Y plane roughly corresponds to human perception of luminance and luminance edges are most noticeable,[18] this is a good approximation to full deblurring. We chose the YCbCr color space over a more perceptually accurate color space such as CIELAB because the Y channel is a linear function of the original RGB channels, and thus the blur affecting Y is still H. After deblurring the Y channel, we converted the YCbCr image back to an RGB image, and then all visual comparisons were performed with RGB images. However, the IQM's only used the Y color plane as specified below.

4.2 Training and Test Images

The training images and test images used are shown in Fig. 2 and are from the publicly available Kodak benchmark image dataset. We selected the test images randomly, and selected the training images to fall in one of two coarse image categories: `structure` or `human`.

In preliminary experiments we studied the effect of the number of training images on EB. While it may seem that the larger the set of training images the better for EB regularization, we found in preliminary experiments that this was not true, as having too many choices makes it possible to find very close nearest neighbor examples, but does not necessarily increase the chance that those nearest neighbor examples have the same high-frequency content as the block to be reconstructed. This *overfitting* in the nearest neighbor selection led to objectionable artifacts. The more degraded the test image, the more risk there is of overfitting. In addition, we found that using a fewer number of training images results in a greater diversity in the selected neighbor examples for a given block, and this had benefits in the EB reconstruction.

4.3 Subjects and Environment

All nineteen subjects had normal color vision, normal or corrected-to normal vision. All experiments took place in a room with controlled D50 lighting, using a calibrated NEC MultiSync LCD 2490WUXi monitor in sRGB mode, and with the observer sitting at roughly 18 inches from the monitor.

Figure 2. Test images `house`, `boat` and `face` (top), `structure` training set (middle), and `human` training set (bottom); all images are 768×512 pixels.

4.4 Parameter Choices

The set of possible parameter values was $\lambda_{\text{AUX}} \in \{0.1, 0.5, 1\}$, $\lambda_{\text{EB}} \in \{0, 0.001, 0.1, 1\}$ and $\lambda_{\text{TV}} \in \{0, 0.001, 0.1\}$, for a total of 36 different combinations of parameter values. These values were chosen in a preliminary experiment such that the parameters were far enough apart to produce perceptually different reconstructions.

5. SUBJECTIVE EXPERIMENTS

We conducted controlled subjective experiments to compare EB, TV, and EB-TV regularizers on the deblurring task. In addition, we investigated how well the objective function parameters for each method can be optimized by cross-validation on two different categories of training images. Lastly, we compared how well different IQMs could choose optimal parameters from training images compared to humans.

5.1 Experiment Set One: Parameter Training Experiments

The first round of experiments was parameter training. In this experiment, we tested our hypothesis that a good set of parameter choices for a particular image will be the same as for similar images, so that one can learn good

parameter choices using similar training images. Here, by similar, we mean having similar spatial features and edge-content.

For each of the five training images shown in Fig. 2, the 36 different-parameter-setting reconstructions were ordered randomly, and each of three subjects blindly and independently scored each of the 5×36 reconstructions according to the following categories: very good (score=5), good (score=4), fair (score=3), unsatisfactory (score=2) and objectionable (score=1). Scoring was averaged across the three subjects to yield the "human score." Although the three subjects' scores showed high correlation, the subjects differed in their sensitivity to blur, to overall noise, and to objectionable artifacts (of the type shown in Fig. 1, second from right). In addition, for each of the three test images shown in Fig. 2, the 36 different-parameter-setting reconstructions were made for each of the two training image sets (structure and human), and the $3 \times 36 \times 2$ resulting reconstructions were scored by the same three subjects and the scores averaged.

Fig. 3 shows the results. Each marker corresponds to one of the 36 parameter settings (zoom in to see that some markers overlap). The x-axis location of a marker denotes the mean score of reconstructed training images with that parameter setting. The y-axis location of a marker denotes the score on a particular test image.

If all of the markers lay on the diagonal, it would show that one could perfectly choose regularization parameters for a test image by using the optimal regularization parameters for the training images. For example, the top-left plot shows that the parameter scores for the `house` test image correlate well with the parameter scores for the `structure` training images, but the top-right plot shows that the parameter scores for the `face` test image correlate poorly with the parameter scores for the `structure` training images. Reviewing all of the plots, we conclude that the more similar the image structure, the easier it is to choose regularization parameters on training images that will also work well on test images.

As a separate issue, Fig. 3 shows what parameter choices are optimal for each test image. The red squares correspond to parameter choices where only $\lambda_{\mathrm{EB}} = 0$ (TV regularization only), the green diamonds correspond to parameter choices where $\lambda_{\mathrm{TV}} = 0$ (EB regularization only), and the blue triangles mark parameter choices that use both EB regularization and TV regularization. For example, the parameter set that produced the highest-quality `boat` reconstruction given the `structure` images did not use EB regularization, but the highest-quality `face` reconstruction used both EB and TV regularization. While many of the top scores use both EB and TV regularization, we do not think the results show strong evidence that both types of regularization are needed to achieve high-quality reconstruction. However, in practice it may be easier and more robust to train the parameters for joint regularization - we test this hypothesis in the following subsection.

5.2 Cross-validating Deblurring Parameters Using IQMs

In practice it would be desirable to automatically cross-validate the regularization parameters on training images, removing the need for human subjects. To that end we consider whether a full-reference IQM can be used to assess the best parameter choices on the training images, since a high-quality training image is available to compare to the 36 reconstructions with different parameter choices. We evaluated five candidate IQMs: PSNR, the 95% highest error as measured by ΔE CIELAB, the 95% highest error as measured by ΔE of s-CIELAB,[19] structural similarity index (SSIM),[20] and visual information (VIF).[21] For each IQM, the Y plane of the true image is compared to the Y plane of the reconstructed image.

The scatterplot in Fig. 4 shows how IQM scores correlate with human scores when averaged over each training set. We hoped to see a strong positive correlation for PSNR, SSIM and VIF 95% ΔE and a strong negative correlation for 95% s-ΔE (because low ΔE scores are desirable).

To determine which IQM performs closest to the human observers, we consider two metrics, shown in the tables in Fig. 4. First, the correlation between each IQM and the human score gives a global indication of how well the IQM mimics human performance. However, since the goal is to select the regularization parameters that achieve the best IQM score, we are especially interested in whether the highest-ranking IQM parameters correspond to high-ranking human scores. Thus for a second metric, we compute the average human score of the top five IQM-selected parameters. For both metrics and for both the `structure` and `human` training sets, SSIM performs better than the other IQMs.

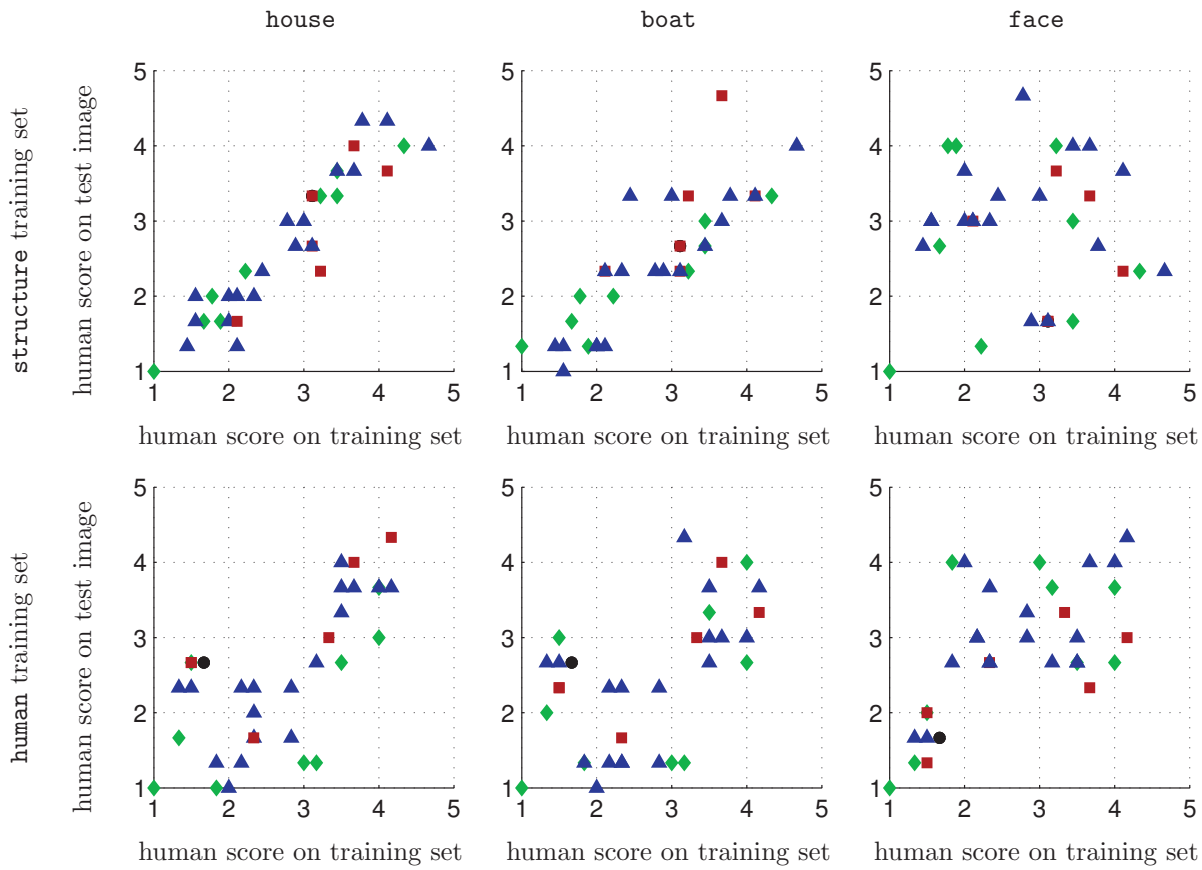
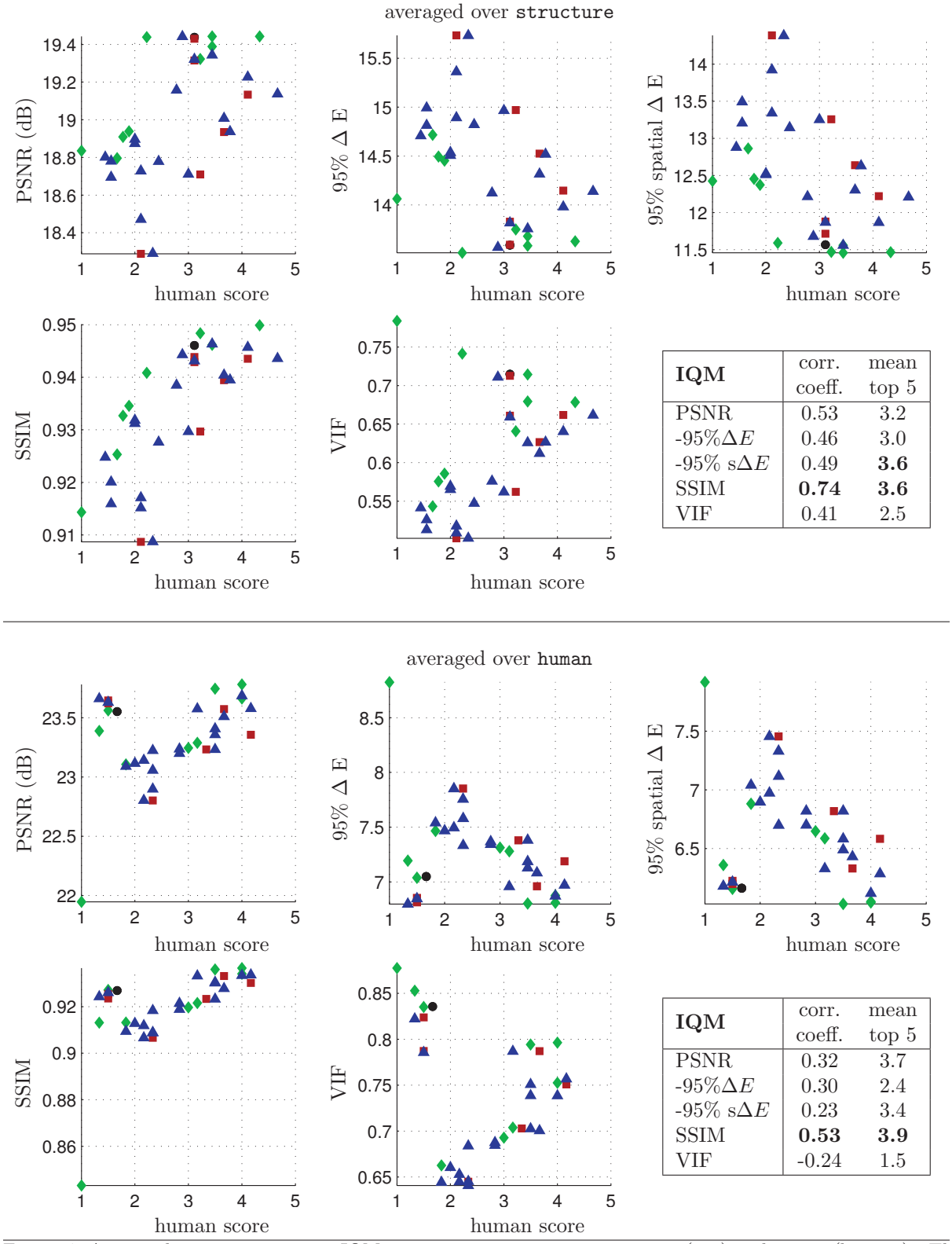


Figure 3. Correlation of Parameter-settings Between Test and Training: Each marker in each plot corresponds to a parameter-setting, and its x-axis position corresponds to its average human score on the training set, while its y-axis position corresponds to its average human score on a particular test image. The test images `house` (left column), `boat` (middle column) and `face` (right column) were reconstructed using the `structure` (top row) and `human` (bottom row) training sets. Scores for the `house` and `boat` exhibit better correlation to the `structure` training set, while `face` exhibits better correlation to the `human` training set. The marks denote images for which $\lambda_{\text{EB}} = \lambda_{\text{TV}} = 0$ (●), $\lambda_{\text{EB}} = 0$, $\lambda_{\text{TV}} > 0$ (■), $\lambda_{\text{EB}} > 0$, $\lambda_{\text{TV}} = 0$ (◆), and $\lambda_{\text{EB}}, \lambda_{\text{TV}} > 0$ (▲).

5.3 Experiment Set Two: Comparisons With Trained Parameters

In our second round of experiments, we performed larger-scale subjective comparisons of reconstructions using the parameter values chosen by cross-validation. As given in Table 1, eight parameter settings were compared, covering all combinations of three binary factors: *(i)* which of the two training sets was used, *(ii)* whether or not EB regularization was used, and *(iii)* whether the parameters were cross-validated on the training set using human judgement or SSIM as detailed in Sec. 5.1.

Subjective comparisons for the second round of experiments were made by a set of sixteen subjects that did not include the subjects from the first experiment set. Subjects compared the eight different parameter settings on each of the three test images for a total of 24 images evaluated by each subject. The forced-choice pair comparison protocol was used.[22–24] In the beginning of the experiment, the observers were shown the original (non-blurred) images as a reference, while during the comparisons only each pair of reconstructed images to be compared were shown at any one time. The observers were informed that the images they would be comparing were high-quality images that had been subjected to blur and noise and then processed with different deblurring methods, and they were instructed to select the reconstruction in each pair that they *preferred*. All pairs were presented twice with the two reconstructions on different sides of the monitor to avoid possible systematic errors in which people prefer one side of the monitor to the other when reconstructions seem indistinguishable. The observers' responses were analyzed according to Case V of Thurstone's Law of Comparative Judgement in order



IQM	corr. coeff.	mean top 5
PSNR	0.53	3.2
-95%ΔE	0.46	3.0
-95% sΔE	0.49	**3.6**
SSIM	**0.74**	**3.6**
VIF	0.41	2.5

IQM	corr. coeff.	mean top 5
PSNR	0.32	3.7
-95%ΔE	0.30	2.4
-95% sΔE	0.23	3.4
SSIM	**0.53**	**3.9**
VIF	-0.24	1.5

Figure 4. Average human score versus IQM score over training set `structure` (top) and `human` (bottom). The marks denote images for which $\lambda_{\text{EB}} = \lambda_{\text{TV}} = 0$ (•), $\lambda_{\text{EB}} = 0$, $\lambda_{\text{TV}} > 0$ (■), $\lambda_{\text{EB}} > 0$, $\lambda_{\text{TV}} = 0$ (♦), and $\lambda_{\text{EB}}, \lambda_{\text{TV}} > 0$ (▲). The table lists the correlation coefficient (negative correlation for the ΔE metrics) and the average human score of the top 5 IQM parameters.

Table 1. Regularization parameters for the proposed algorithm and TV chosen by cross-validation on each training set by a human-in-the-loop and by SSIM.

	λ_{AUX}	λ_{EB}	λ_{TV}
trained on `structure`			
No EB Reg. (human)	1.0	—	0.001
No EB Reg. (SSIM)	1.0	—	0
With EB Reg. (human)	0.5	0.001	0.001
With EB Reg. (SSIM)	0.5	0.1	0
trained on `human`			
No EB Reg. (human)	1.0	—	0.001
No EB Reg. (SSIM)	0.5	—	0.001
With EB Reg. (human)	0.5	0.1	0.001
With EB Reg. (SSIM)	1.0	0.1	0

to obtain Z-scores.[22–24]

Results for this experiment are given in Fig. 5. The left column shows results using the `structure` training images, and the right column shows results using the `human` training images. We focus first on the parameters chosen by cross-validation using human judgement. For human-judgement parameters and training on `structure` images, the joint EB-TV deblurring is a statistically significant preference on the `house` and `boat` test images, which are psychovisually more similar to the `structure` training images than the `face` image. On the `face` image, there was not a statistically significant difference with the addition of EB regularization when trained with the `structure` images. Further, when trained on `human` training images, the joint EB-TV regularization is preferred to TV regularization only, but the differences are not statistically significant for any of the test images.

Next, consider the results using SSIM cross-validated parameters. Fig. 5 shows that SSIM cross-validated parameters almost always led to worse results than using human cross-validated parameters, and sometimes dramatically worse. When using SSIM cross-validated parameters, the joint EB-TV regularization is statistically significantly better than TV-only for the cases of `face` and `house` trained with `structure`, and those are the only statistically significant differences.

When subjects were asked afterward their evaluation to comment on what was important to them in making their preference judgements, three main attributes were mentioned: sharpness, presence of artifacts (described as jagged edges/ghosting/ringing), and noise (described as graininess/dottedness/pockiness). Typically the observers reported making tradeoffs between these attributes. Notably, according to these post-experiment interviews, there was a large variation in how the observers ranked these three different attributes.

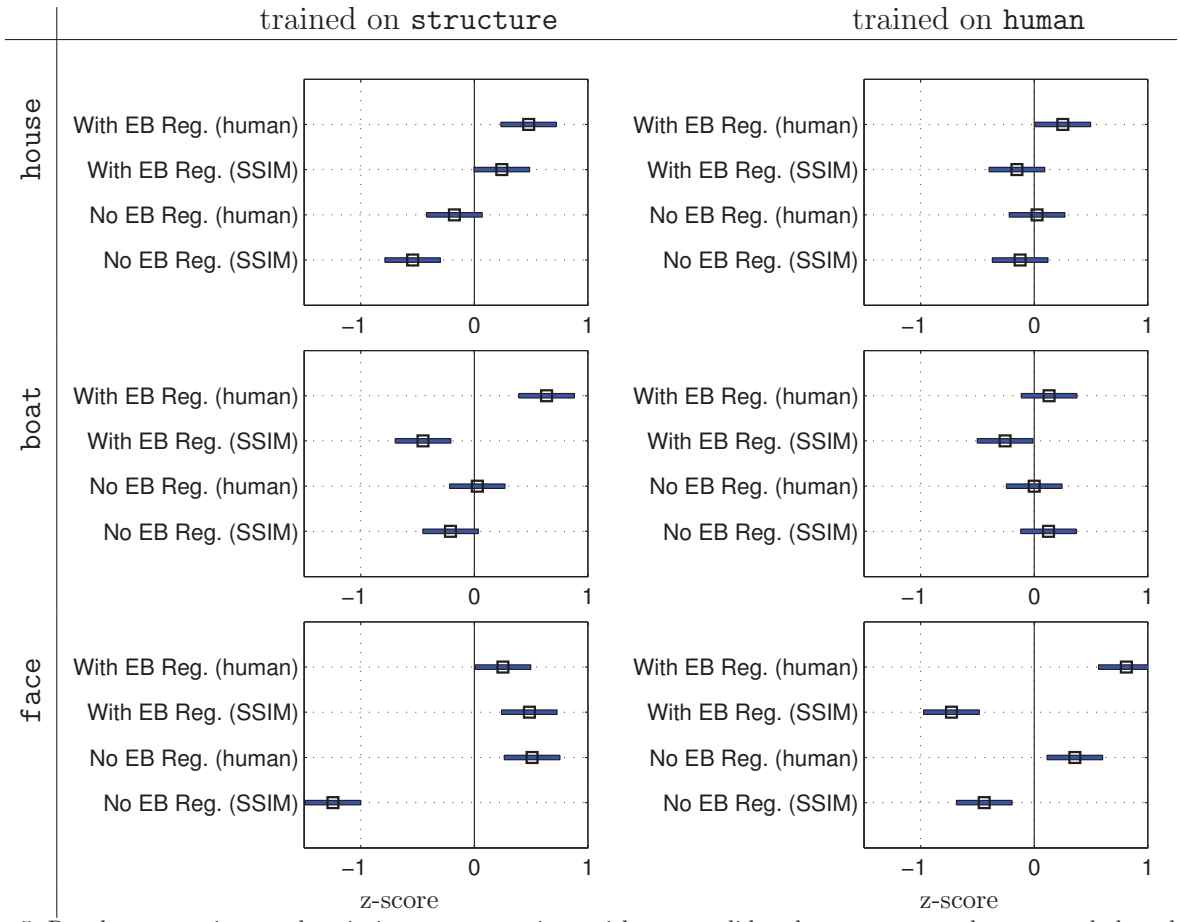


Figure 5. Results comparing total variation reconstructions with cross-validated parameters and no example-based regularization, or with additional example-based regularization.

6. CONCLUSIONS

In this paper we presented the results of subjective experiments comparing the value of example-based regularization and total-variation regularization as well as their combined joint regularization effect. In addition, we compared how well the regularization parameters for both methods can be learned by cross-validation, either with human judgement or an IQM.

Our experiments suggest that good deblurring parameters are learnable by cross-validation with human scores, but more so if the training and test data are psychovisually well-matched, and even more so for example-based regularization if the images have structure. Of five IQM's, we found that SSIM correlated best with human judgements of quality of the reconstructions for all images. Overall, the parameters chosen by SSIM yielded much lower quality reconstructions than the parameters chosen by humans. Comparing the use of EB regularization to no EB regularization, we found that using EB regularization was significantly better than using TV regularization only when the test and training images were psychovisually matched, and particularly for more structured images. The effect was stronger using parameters trained with human judgement rather than by SSIM.

We found the results depended heavily on the training set. We believe there are two main reasons the training set matters. First, different training sets may have different quantities and qualities of structure. Second, as shown in Fig. 3, the appropriate set of parameters for different images may differ; we hypothesize that this is primarily because humans judge image quality differently depending on the image content, for example noise appears to be judged more harshly in face images, and that affects the choice of regularization parameters. We did not try to quantify how the training set should match the test set, which is an important open question for future research.

Although we focused on image deblurring with a known point spread function in this paper, we believe that the knowledge gained from these experiments will be useful for many image processing problems where total variation has already been shown to be effective.

APPENDIX A. APPENDIX: FULL ALGORITHM IMPLEMENTATION DETAILS

In the following subsections, we provide full details on how we solved (3).

A.1 Pre-Processing the Training Images

To generate $\{T_i\}_{i=1}^{B}$, we pre-process training images as prescribed in Freeman et al.:[25] edges are extracted using a high-pass filter, and the images are (locally) contrast normalized. Training image blocks are generated by extracting every $m \times m$ block from the corpus of training images; we use $m = 9$. For each $m \times m$ block of the test image in raster-order, we select the $k = 5$ (rather than $k = 1$ as done in[3, 25]) blocks that, when blurred by H, are the k nearest neighbors to the test block in terms of Euclidean distance. To promote continuity across block boundaries, we overlap subsequent blocks by one row (or column) of pixels, and use the same distance calculation between blocks as Freeman et al.:[25] the distance between blocks is the Euclidean distance between the blurred blocks plus a scalar ($\alpha = 0.1$) times the Euclidean distance between the current reconstruction estimate for the overlapped pixels and the unblurred training blocks. Each of the k nearest neighbor training blocks inherits the local contrast and low-frequency information of the test block; this is done so that the low-frequency content of x is not regularized towards zero for the λ_{EB} term in (3).

Searching for the nearest-neighbors can be expedited by creating a kd-tree, however this requires extracting, processing and storing every overlapping block from each training image, which leads to redundancy in a straightforward implementation. Instead, we found it quite efficient to exploit the fact that the cross-term of $\|a - b_i\|^2 = \|a\|^2 - 2\langle a, b_i\rangle + \|b_i\|^2$ can be computed by cross-correlation of image block a with $\{b_i\}_{i=1}^{B}$ quickly using FFTs, while the other two terms can be computed once prior to the search.

A.2 Minimizing the Deblurring Objectives

We describe how we minimize the joint EB-TV objective (3), the EB-only and TV-only objectives are special cases. Since (3) is jointly convex in its variables, its minima can be found by alternating minimizations over each variable.[26] Specifically, we first optimize over $\{\alpha_i\}_{i=1}^B$ for fixed x and v, then optimize over x and v for fixed $\{\alpha_i\}_{i=1}^B$. The latter optimization over x and v is accomplished with an inner alternating minimizations loop. That is, (3) is implemented by solving

$$\left\{\begin{array}{ll} \text{Step 1:} & \min\limits_{\{\alpha_i\}_{i=1}^B} \|x - \sum\limits_i T_i\alpha_i\|_2^2 \\ \left\{\begin{array}{ll} \text{Step 2a:} & \min\limits_x \lambda_{\text{AUX}}\|x - v\|_2^2 + \lambda_{\text{EB}}\|x - \sum\limits_i T_i\alpha_i\|_2^2 \\ & +\lambda_{\text{TV}}\|x\|_{TV} \\ \text{Step 2b:} & \min\limits_v \|Hv - z\|_2^2 + \lambda_{\text{AUX}}\|x - v\|_2^2, \end{array}\right. \end{array}\right. \tag{4}$$

where each bracket denotes a loop of alternating minimizations. We next detail how we solve the subproblems that arise in each of the steps.

Since the blocks $\{T_i\}$ overlap, optimizing Step 1 requires that all $\{\alpha_i\}_{i=1}^B$ be found jointly using, however, for the related example-based superresolution problem, Freeman et al. found that a raster-order method performed well compared to a joint solution.[3] With that motivation, we similarly solve Step 1 in (4) by solving for each block in raster scan order. Then solving for each α_i is equivalent to solving a least-squares linear regression problem, and thus each α_i has a closed-form solution $\alpha_i = (T_i^T T_i)^{-1} T_i^T x$ (solved efficiently by exploiting the sparsity of T_i; the only nonzero entries of T_i correspond to the support of the ith block).

We derived a solution for Steps 2a and 2b in (4) by extending recent work on efficient total variation deblurring. Chambolle showed that the solution to $\min\limits_x \frac{1}{2}\|x - g\|^2 + \gamma\|x\|_{TV}$ can be found using a dual formulation for total variation.[15] In particular, note that $\|x\|_{TV} = |\nabla x| = \max\limits_{\|p\|\leq 1} \langle p, \nabla x\rangle$. By substituting this dual formulation for $\|x\|_{TV}$, applying the minimax theorem to swap max and min, then minimizing using the Euler-Lagrange method with respect to x, Chambolle shows that

$$x = g - \gamma(\nabla \cdot p), \tag{5}$$

which gives a relationship between the image x, the dual variable p and the target image g. From this, Chambolle derived a semi-implicit gradient descent method.[15] Bresson and Chan[16] and Wen et al.[17] extended this approach for total-variation image deblurring of full-color images; each work employed an alternating-minimizations algorithm. Wang et al. took a slightly different approach to obtain computational efficiency for total variation deblurring,[27] which we believe could also be generalized to implement the proposed objective function.

Analogously, let $f = \sum_i T_i\alpha_i$, substitute $\|x\|_{TV} = \max_{\|p\|\leq 1}\langle p, \nabla x\rangle$, and minimize Step 2a in (4) with respect to x using the Euler-Lagrange technique to yield

$$x = \frac{\lambda_{\text{AUX}} v + \lambda_{\text{EB}} f}{\lambda_{\text{AUX}} + \lambda_{\text{EB}}} - \frac{\lambda_{\text{TV}}}{2(\lambda_{\text{AUX}} + \lambda_{\text{EB}})}(\nabla \cdot p). \tag{6}$$

This can be written as Chambolle's solution given in (5) with $g = \frac{\lambda_{\text{AUX}} v + \lambda_{\text{EB}} f}{\lambda_{\text{AUX}} + \lambda_{\text{EB}}}$ (a convex combination of the auxillary image v and the example-based image f) and $\gamma = \frac{\lambda_{\text{TV}}}{2(\lambda_{\text{AUX}} + \lambda_{\text{EB}})}$, which enables us to implement Steps 2a and 2b of (4) using the alternating minimizations approach of Bresson and Chan:[16]

$$\left\{\begin{array}{l} x = \frac{\lambda_{\text{AUX}} v + \lambda_{\text{EB}} f}{\lambda_{\text{AUX}} + \lambda_{\text{EB}}} - \frac{\lambda_{\text{TV}}}{2(\lambda_{\text{AUX}} + \lambda_{\text{EB}})}(\nabla \cdot p) \\ v = \left((H^T H + \lambda_{\text{AUX}} I)\right)^{-1}\left(A^T z + \lambda_{\text{AUX}} x\right). \end{array}\right. \tag{7}$$

Note that $v = (H^T H + \lambda_{\text{AUX}} I)^{-1}(A^T z + \lambda_{\text{AUX}} x)$ can be solved efficiently by approximating H as block circulant (corresponding to 2D circular convolution); it is then diagonalizable using the 2D FFT, and we need only compute the inverse of a diagonal matrix.

A.3 Pixel-wise EB Regularization

In preliminary experiments, we observed that the example-based term in (3) can induce artifacts in the reconstruction especially when λ_{EB} is large. The severity of the artifacts depends on how well the training blocks match image blocks in the image z. To control this effect, we form a pixel-specific regularization parameter $\tilde{\lambda}_{EB,i} \in [0, \lambda_{\mathrm{EB}}]$ for each pixel i, based on the EB image residual $z - Hf$, where $f = \sum_i T_i \alpha_i$. For pixels where $\|z - Hf\|$ is large, we wish to decrease the effect of the example-based regularization. To this end, we modify (7) so that the update on x is on a per-pixel basis:

$$x_i = \frac{\lambda_{\mathrm{AUX}} v_i + \tilde{\lambda}_{EB,i} f_i}{\lambda_{\mathrm{AUX}} + \tilde{\lambda}_{EB,i}} - \frac{\lambda_{\mathrm{TV}}}{2(\lambda_{\mathrm{AUX}} + \tilde{\lambda}_{EB,i})} (\nabla \cdot p)_i,$$

where the subscript i denotes the ith element of the vector. In our experiments, we form $\tilde{\lambda}_{EB}$ by starting with the residual image $r = z - Hf$, using each pixel's residual to define a pixel weight $q_i = e^{-50 r_i^2}$, then spatially smoothing the weight-image q by convolving it with a 2×2 spatial filter s whose taps are each $1/4$, resulting in the pixel-wise regularization parameter $\tilde{\lambda}_{EB,i} = \lambda_{\mathrm{EB}} (q ** s)_i$, where $**$ denotes spatial (two-dimensional) convolution. We found that the performance of this pixel-wise regularizer was robust to the exact specification of the q and s.

REFERENCES

[1] Mese, M. and Vaidyanathan, P. P., "Look up table method for inverse halftoning," *IEEE Trans. on Image Processing* **10**(10), 1566–1578 (2001).

[2] Wang, Z., Yu, Y., and Zhang, D., "An information loss restoration technique for block-based image coding systems," *IEEE Trans. Image Processing* **7**(7), 1056–1061 (1998).

[3] Freeman, W. T., Jones, T. R., and Pasztor, E. C., "Example-based super-resolution," *IEEE Comp. Graph. and Appl.* **22**, 56–65 (Mar/Apr 2002).

[4] Simoncelli, E., "Modeling the joint statistics of images in the wavelet domain," *Proc. SPIE* , 188–195 (1999).

[5] Crouse, M. S., Nowak, R. D., and Baraniuk, R. G., "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing* **46**, 886–902 (1998).

[6] Anderson, H. S. and Gupta, M. R., "Joint deconvolution and imaging," *Proc. SPIE Conf. on Computational Imaging* (2009).

[7] Chang, H., Yeung, D. Y., and Xiong, Y., "Super-resolution through neighbor embedding," *Proc. IEEE Comp. Vis. and Pat. Rec.* **1**, 275–282 (June-July 2004).

[8] Jiji, C. V. and Chaudhury, S., "Single-frame image super-resolution through contourlet learning," *EURASIP Jour. Appl. Sig. Proc.* **2006**, 1–11 (2006).

[9] Jiji, C. V., Chaudhuri, S., and Chatterjee, P., "Single frame image super-resolution: should we process locally or globally?," *Multidim. Syst. Sig. Process.* **18**, 123–152 (2007).

[10] Ma, Q., Wu, F., Lou, J., and Sun, M., "Frame loss error concealment for spatial scalability using hallucination," *Packet Video Workshop* (May 2009).

[11] Zhu, S. C. and Mumford, D., "Prior learning and Gibbs reaction-diffusion," *IEEE Trans. Pattern Analysis Machine Intelligence* **19**(11), 1236–1250 (1997).

[12] Baker, S. and Kanade, T., "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Analysis Machine Intelligence* **24**(9), 1167–1183 (2002).

[13] Datsenko, D. and Elad, M., "Example-based single document image super-resolution: a global MAP approach with outlier rejection," *Multidim. Syst. Sig. Process.* **18**, 103–121 (2007).

[14] Elad, M. and Datsenko, D., "Example-based regularization deployed to super-resolution reconstruction of a single image," *The Computer Journal* **50**(4), 1–16 (2007).

[15] Chambolle, A., "An algorithm for total variation minimization and applications," *J. Math. Imaging and Vision* **20**, 89–97 (2004).

[16] Bresson, X. and Chan, T. F., "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Problems and Imaging* **2**(4), 455–484 (2008).

[17] Wen, Y., Ng, M., and Huang, Y., "Efficient total variation minimization methods for color image restoration," *IEEE Trans. Image Processing* **17** (November 2008).
[18] Fairchild, M. D., [*Color Appearance Models 2nd edition*], Addison Wesley, Reading, MA (2005).
[19] Zhang, X. and Wandell, B. A., "A spatial extension of CIELAB for digital color image reproduction," *SID Symposium Technical Digest* **27**, 731–734 (1996).
[20] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing* **13**, 600–612 (Apr 2004).
[21] Sheikh, H. R. and Bovik, A. C., "Image information and visual quality," *IEEE Trans. Image Processing* **15**, 430–444 (Feb 2006).
[22] Thurstone, L. L., "A law of comparative judgment," *Psychological Review* **34**, 273–286 (1937).
[23] Tsukida, K. and Gupta, M. R., "How to analyze paired comparison data," *University of Washington Dept. of Electrical Engineering Technical Report UWEETR-2011-0004* (2011).
[24] Engeldrum, P. G., [*Psychometric Scaling: A toolkit for imaging systems development*], Imcotek Press, Winchester, Massachusetts (2000).
[25] Freeman, W. T., Pasztor, E. C., and Carmichael, O. T., "Learning low-level vision," *Intl. J. Comp. Vis.* **40**(1), 25–47 (2000).
[26] Yeung, R., [*A First Course in Information Theory*], Springer, New York, NY (2002).
[27] Wang, Y., Yang, J., Yin, W., and Zhang, Y., "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences* **1**(3), 248–272 (2008).