Joint Deconvolution and Imaging

Hyrum S. Anderson[a] and Maya R. Gupta[b]

University of Washington, Box 352500, Seattle, Washington 98195, USA
[a]hyrum@ee.washington.edu
[b]gupta@ee.washington.edu

Abstract. We investigate a Wiener fusion method to optimally combine multiple estimates for the problem of image deblurring given a known blur and a corpus of sharper training images. Nearest-neighbor estimation of high frequency information from training images is fused with a standard Wiener deconvolution estimate. Results show an improvement in sharpness and decreased artifacts compared to either the standard Wiener filter or the nearest-neighbor reconstruction.

Keywords: Wiener fusion, deconvolution, deblurring, superresolution

1 INTRODUCTION

Many applications in which images play a critical role are limited by the effective resolution of the imagery. For this reason, various approaches exist to estimate high-frequency image content to increase the level of detail. In this paper we assume that one knows the blurring filter and image statistics, and we focus on two relevant approaches: Wiener deconvolution for deblurring, and single-frame learning-by-example superresolution that estimates high-resolution detail for the blurred image from a corpus of sharper training images. We present a method that fuses the Wiener filtering approach and the learning-by-example approach optimally in terms of minimizing expected squared error.

Consider the imaging model in the Fourier domain

$$Y(k,l) = H(k,l)X(k,l) + V(k,l), \tag{1}$$

where $H(k,l)$, $V(k,l)$, $X(k,l)$ and $Y(k,l)$ represent the discrete Fourier transform of the point spread function (PSF) $h(m,n)$, additive noise $v(m,n)$, original image $x(m,n)$ and blurred image $y(m,n)$, respectively. We assume that $x(m,n)$ and $v(m,n)$ are stationary random processes with corresponding power spectra $S_x(k,l)$ and $S_v(k,l)$. For convenience, we will omit indices (k,l) and use f to denote a frequency of interest such that $Y(f) = Y(k,l)$.

Given a known PSF and the power spectra of the noise and image, Wiener deconvolution is the linear minimum mean squared error (LMMSE) estimator of the original image [1]. The Wiener filter inverts the blur more aggressively for frequencies that have a higher signal-to-noise ratio. As such, the Wiener deconvolution cannot recover information in the nullspace of the blur operator, and does little at frequencies for which the signal-to-noise ratio (SNR) is low.

Several recent methods in single-frame superresolution aim at learning high-frequency information from a corpus of sharp training images [2–6]. These methods implicitly assume that the lost high-frequency information $X(f_{\text{hi}})$ is correlated with the mid-frequency information $X(f_{\text{mid}})$. In fact, positive correlation between image frequency bands has been documented for wavelet filterbanks [7, 8]. Similar learning-by-example methods have also been successful for other image estimation tasks (see for example [9, 10]).

We contend that Wiener deconvolution and learning-by-example superresolution can be thought of as complementary tasks. Assuming $H(f)$ represents a low-pass filter, Wiener deconvolution provides the best LMMSE estimate of $X(f_{\text{mid}})$ from the blurred image, but recovers little information at f_{hi}. Conversely, single-frame superresolution provides novel information by estimating $X(f_{\text{hi}})$ from another source, but does not improve the estimate at f_{mid}.

In this work, we provide an optimal (LMMSE) method for fusing Wiener deconvolution with additional image information provided by a nearest-neighbor single-frame superresolution approach. Our proposed Wiener fusion could be used with other image estimates instead. We begin in Sec.2 by reviewing relevant methods in single-image superresolution, and describe our adaptation of a method presented in [2] to image deblurring. In Sec. 3, we detail the proposed multiple estimate (ME) Wiener filter. Sec. 4 details experiments, whose results are discussed in Section 5.

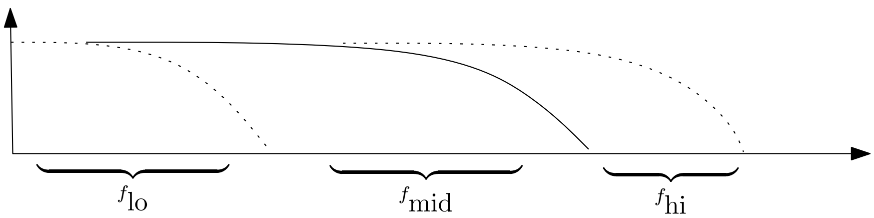


Fig. 1. Nearest neighbor reconstruction tries to recover $X(f_{\text{hi}})$ from $X(f_{\text{mid}})$. Low frequencies are discarded in the NN comparison.

2 NEAREST-NEIGHBOR SUPERRESOLUTION

We first review some relevant work in single-frame image superresolution, and describe how we adapt it for deblurring.

Freeman et al. [2] introduced a nearest-neighbor learning approach for upsampling a low-resolution image. The missing high-frequency information is inferred on a block-by-block basis by comparing a low-resolution image patch to a set of high-resolution training patches, down-sampled to match the test image resolution. High-frequency information is taken from the training patch that yields the best comparison. A boundary condition is included in the distance computation to regularize the nearest neighbor selection and form continuity in the reconstructed image by selecting high-resolution patches that match previously estimated high-resolution detail. A simple 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. For each patch z, the objective function for selecting the nearest-neighbor training patch x_k is given by

$$\|Hx_j - z\|_2^2 + \alpha \sum_{[i,j]\in \mathrm{B}} \left(\tilde{x}[i,j] - x_k[i-i_0, j-j_0]\right)^2 \tag{2}$$

where H is the decimation (or here, blurring) operator, $\tilde{x}$ is the current high-resolution estimate for the entire image, B is the set of indices of overlapping pixels, and (i_0, j_0) is the first overlap location. Equation (2) can be derived from a maximum a posteriori (MAP) objective $p(z|x_k)p(x_k)$ assuming additive white Gaussian noise, and a data-dependent Gaussian prior. Although Freeman et al. [2] applied the method to image upsampling, it is easily extended to debluring by choosing H to be the appropriate blur operator, as detailed below.

This nearest-neighbor (NN) reconstruction approach requires a large set of training patches that can be from images not necessarily similar to the test image. The approach is also sensitive to noise and image artifacts of the input image z [2]. Chang et al. presented a method which requires fewer training samples by posing the nearest-neighbor problem similar to manifold learning [3]. Jiji et al. proposed a method in which high-resolution detail was learned from a database using contourlet learning [4]. Jiji et al. evaluated various local and global single-frame superresolution reconstruction techniques [6], and proposed a global PCA-based technique to remove blur, noise, and aliasing artifacts. Datsenko and Elad used NN patches to form a prior for a global MAP objective, which they used to superresolve scanned documents containing text, equations and graphs [5]. In that NN approach, outliers in a group of candidate neighbors were pruned using a MAP cost function.

Here we adapt Freeman's method [2] for image deblurring. We assume that the input image is the result of linear time-invariant (LTI) blur of the true image with additive noise. The NN approach in [2] is sensitive to noise and image artifacts, therefore, we begin by performing Wiener denoising on the input image as a preprocessing step. (We have explored other preprocessing alternatives, and have found that the NN reconstruction is sensitive to image artifacts produced by other methods, as shown in Fig. 3.) Next, the denoised image is high-pass filtered, and the result is contrast normalized using a low-pass filter defined by Freeman et al. in [11]. Overlapping blocks are scanned in raster order, and for each block the nearest neighbor is found in the training images. The NN match specifies the high-frequency estimate for the test block. After the NN training path and its corresponding high-frequency information has been found for each block in the image, the original local contrast is reapplied. The result is an estimate of the high-frequency content of the image, which must subsequently be combined with the existing low-res image to form a sharp image. A block diagram is shown in Fig. 2.

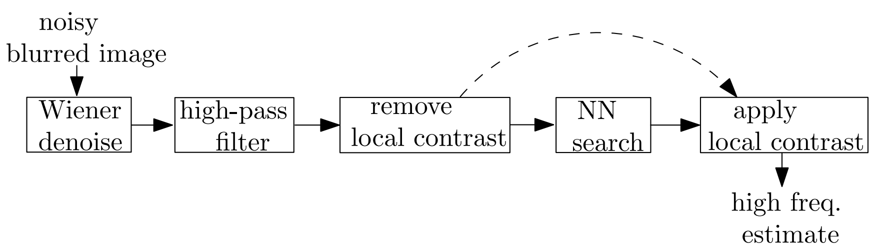
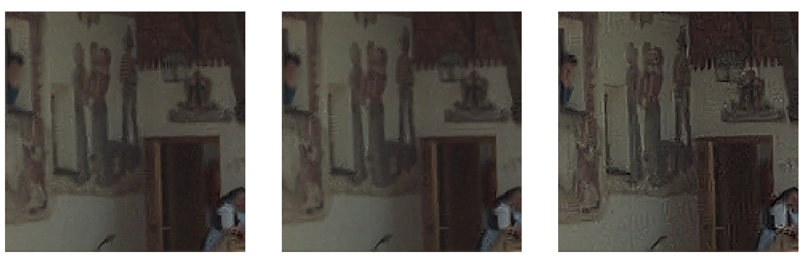


Fig. 2. Block diagram showing the steps used to estimate high-frequency information using a NN search.



NN given noisy image PSNR = 36.1 dB	NN given Wiener denoised image PSNR = 36.5 dB	NN given Wiener deconvolved image PSNR = 32.9 dB

Fig. 3. NN reconstruction given different choices of pre-processing before selecting the nearest-neighbor training patches. The original image degradation was rather mild in this case: the image was blurred with a Gaussian kernel with one pixel standard deviation, and corrupted with additive noise, with 44 dB PSNR. Wiener denoising before NN results in highest PSNR, and provides a good tradeoff between detail and image artifacts.

Training images are prepared in an analogous manner to Freeman et al. [11] by forming low-frequency input and high-frequency output pairs of image patches from a corpus of sharp training images. In our work, the sharp images are blurred with the known PSF as described in (2), and every $m \times m$ patch is extracted from each image. The output vector consists of the high frequency information that was discarded by the blur operator. For the NN match, a training vector is formed by augmenting the column-scanned low-res patch with $\sqrt{\alpha}$ times the top row and left column (raster-scan overlap regions) of high-resolution pixel values. Each test patch mirrors the same setup, so that the NN objective is precisely (2), where H encapsulates the blur.

3 MULTIPLE ESTIMATE WIENER FUSION

There are several theoretical motivations for using Wiener deconvolution for image deblurring. It is well known that the Wiener filter is the LMMSE inverse filter for estimating $X(f)$ in (1), so that when $E[X|Y]$ is linear (for example, when X and Y are jointly Gaussian), it is the MMSE inverse filter. In fact, for linear $E[X|Y]$, the Wiener filter minimizes any expected Bregman divergence $E[D(X,Y)]$, which can be stated as a corollary to the result that the conditional expectation is the unique minimizer of any Bregman divergence as shown in [12].

Wiener deconvolution given multiple noisy observations (such as from different cameras) dates back to at least 1980 [13]. In this paper, we propose treating different estimates of the reconstructed image as multiple noisy observations, and combining the multiple estimates using the Wiener filter. Given n estimates $Y_1, Y_2, \ldots, Y_n$, we assume that $Y_i = H_i X + V_i$, and let the reconstruction be $\hat{X} = G_1 Y_1 + G_2 Y_2 + \cdots + G_n Y_n$. Then the Weiner objective is to minimize $\epsilon = E\left[\left(\hat{X} - X\right)^2\right]$. To minimize ϵ, the partial derivative w.r.t. G_k^* is,

$$\begin{aligned}\frac{\partial \epsilon}{\partial G_k^*} &= E\left[(H_k^* X^* + V_k^*)\left(\sum_i G_i H_i X + \sum_i G_i V_i - X\right)\right] \\ &= G_k |H_k|^2 S_x + H_k^* S_x \sum_{i \neq k} G_i H_i + G_k S_{v_k} - H_k^* S_x.\end{aligned}$$

Equating to zero, and solving for G_k yields

$$G_k = \frac{H_k^* \left(1 - \sum_{i \neq k} G_i H_i\right)}{|H_k|^2 + \frac{S_{v_k}}{S_x}}. \tag{3}$$

The system of equations defined by (3) must be solved jointly for the n filters $G_1, G_2, \ldots, G_n$. The kth filter is given be

$$G_k = \frac{1}{H_k}\left(\frac{|H_k|^2 \prod_{j \neq k}^n S_{v_j}}{\sum_{i=1}^n |H_i|^2 \prod_{j \neq i}^n S_{v_j} + \frac{\prod_{i=1}^n S_{v_i}}{S_x}}\right). \tag{4}$$

Thus, the filter G_k weights the estimate Y_k by partially inverting the assumed relationship $H_k X_k$, where the amount of inversion of H_k depends on the assumed noise power of each estimate and on the relative H_i's.

In our work, we let Y_1 be the given blurred image, and Y_2 be the (high-pass part only) NN reconstruction described in Section 2. Note that when $H_2 = 0$ (no high-res detail available), G_1 reduces to the standard Wiener filter. For deblurring, we assume that $H_1(f)$ is lowpass in nature, and that the high-frequency detail estimated using NN reconstruction is such that $H_2(f) = 1 - H_1(f)$.

4 EXPERIMENTS

We compare the multiple-estimate Wiener deblurring (ME Wiener) given by (4) to the standard Wiener filter and to the NN reconstruction technique described in Sec. 2. We attempt to restore noisy, blurred versions of a benchmark Kodak image and a retinal fundus photograph; the relevant original snippets are shown in Fig. 4.

The test images were generated from the model in (1): the original test image was convolved with a Gaussian PSF $h_1[m, n] = \mathcal{N}(0, s^2 I)$ with standard deviation $s = 1$ or $s = 2$ pixels, and Gaussian white noise is added, with standard deviation $S_{v_1}(f) = \sigma_{v_1}^2 = (0.005)^2$ or $S_{v_1}(f) = \sigma_{v_1}^2 = (0.02)^2$ (the noise corresponds to peak SNR of 46 dB and 34 dB with respect to the sharp image). The luma (Y) and chroma planes (Cb and Cr) are made blurry and noisy, but only the luma plane is restored, and the noisy blurred chroma planes are added to the estimated luma plane.

Optimistically, the true blur, and noise and sharp image power spectra are used for Wiener deconvolution and for the ME Wiener filter, and the actual imaging error power spectrum S_{v_2} is used for the ME Wiener. We use the parameters specified by Freeman et al. for high-pass filtering and contrast normalization in the NN reconstruction [11].

5 RESULTS

The first set of results are shown in Figs. 5, 6, 7 and 8: the Kodak image NN reconstructions were trained on three other clean Kodak benchmark images that are similar to the blurred image in that they have manmade structures and trees, and the retina reconstruction in 8 was trained on three similar clean retina images.

The NN estimates are sharpest, but have disturbing artifacts; compare for example the old woman in the lower-right-hand corner in Fig. 5, the birds in Fig. 6, the edge of the roof in 7, and the yellow blind-spot in Fig. 8.

The ME Weiner combines the new NN high-frequency information with standard Weiner deblurring, and this combination reduces the sharpness a little compared to the NN estimate, but also reduces the artifacts significantly. In all cases the ME Weiner estimate is sharper than the standard Weiner estimate in terms of greater high-frequency energy, and the ME Weiner also reduces some of the upper-mid-frequency noise caused by the Weiner.

The peak SNR (PSNR) for each of the figures is shown in Table 1. In every case, the ME Wiener has highest PSNR. This may be unintuitive, as the Wiener filter is optimal in terms of MSE (and thus PSNR), but the gains over the Wiener filter are possible because the high-frequency information is correlated with the mid-frequency information, and the Wiener filter does not take this into account, whereas this correlation is the basis for the NN reconstruction. The ME Wiener is able to advantageously combine the two estimation approaches, as designed.

Next, we illustrate that the choice of training images matters. The images in Fig. 9 compare the results for the NN reconstruction and ME Wiener using either the Kodak images for training (as before) or using the retina images for training. One sees that different artifacts appear in both, and that training on the retina images arguably works better. Not shown are results comparing the reconstructions of the retina image with the two different training sets; those results show that the retina-trained estimation is also better. We hypothesize this is because of the relatively distinct edges prevalent in the retina images that have high local correlation between the corresponding mid and high-frequencies, and thus any edges picked up in the mid-frequency-based NN match correlate well with high-frequency edge information.

Unfortunately, using a larger corpus training images is not a good solution, as too large a corpus will have blocks that match well at low-resolution but have incorrect high-frequency information. (This is the motivation for outlier rejection presented in [5].) In fact, in Table 2, we show that using only one of the three Kodak training images results in higher PSNR than using all three training images; the corresponding images are shown in Fig. 10. We hypothesize that it may be advantageous to generate multiple NN estimates from multiple sets of training images, then fuse them using the ME Wiener. As a preliminary test of this hypothesis, we present in Fig. 10 results for using only one Kodak training image versus using two training images to form a NN estimate, and using two Kodak training images individually to form two NN estimates which are combined using (4). Evidentally, the ME Wiener formulation is a superior way to incorporate additional training information. Table 2 provides preliminary support that this approach can increase PSNR.

Fig. 4. True image snippets used in the experiments cropped from a Kodak benchmark image of a house (left three) and a retinal fundus photograph (right).

Table 1. PSNR (in dB) of reconstruction estimate for the Kodak house image. Highest PSNR for each column is in bold.

	Low blur Low noise	Low blur High noise	High blur Low noise	High blur High noise
Figure 4 Estimates:				
Blurred/Noisy Original	31.4	28.3	30.2	27.7
Wiener	32.9	31.2	32.5	31.1
NN	32.3	31.2	30.4	29.9
ME Wiener	**33.2**	**31.9**	**32.8**	**31.5**
Figure 5 Estimates:				
Blurred/Noisy Original	28.8	26.9	27.5	26.0
Wiener	29.9	28.6	29.5	28.3
NN	29.8	28.8	27.7	27.2
ME Wiener	**30.3**	**29.4**	**30.0**	**28.8**
Figure 6 Estimates:				
Blurred/Noisy Original	23.3	22.7	22.1	21.6
Wiener	24.1	23.3	23.5	22.7
NN	24.0	23.5	22.2	22.0
ME Wiener	**24.6**	**23.9**	**24.0**	**23.1**
Figure 7 Estimates:				
Blurred/Noisy Original	33.1	29.0	31.9	28.6
Wiener	33.7	32.6	34.1	32.6
NN	33.9	32.8	32.7	31.8
ME Wiener	**34.0**	**33.1**	**34.6**	**33.1**

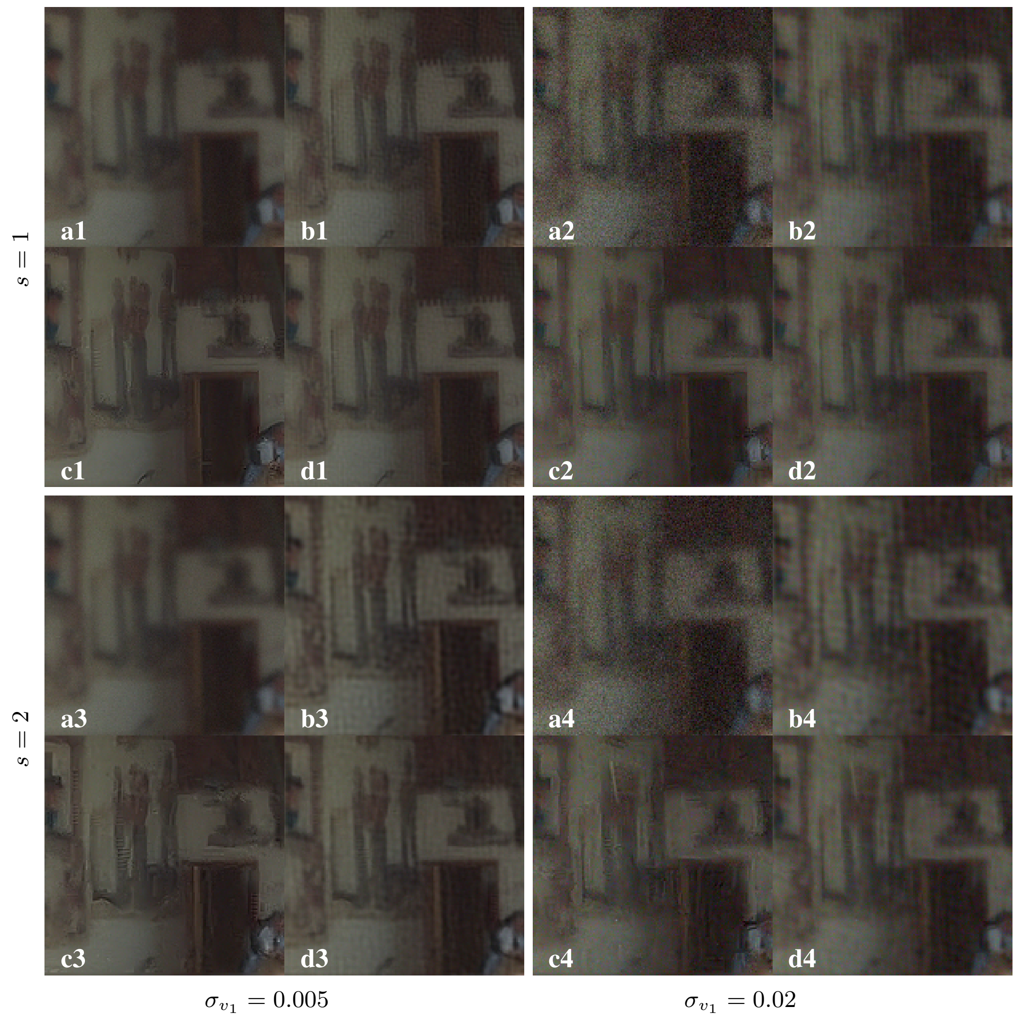


Fig. 5. Results for snippet 1 of Kodak image. The subblocks in each quadrant represent (a) the original blurred image, (b) the Wiener deconvolution estimate, (c) the NN reconstruction and (d) the ME Wiener estimate.

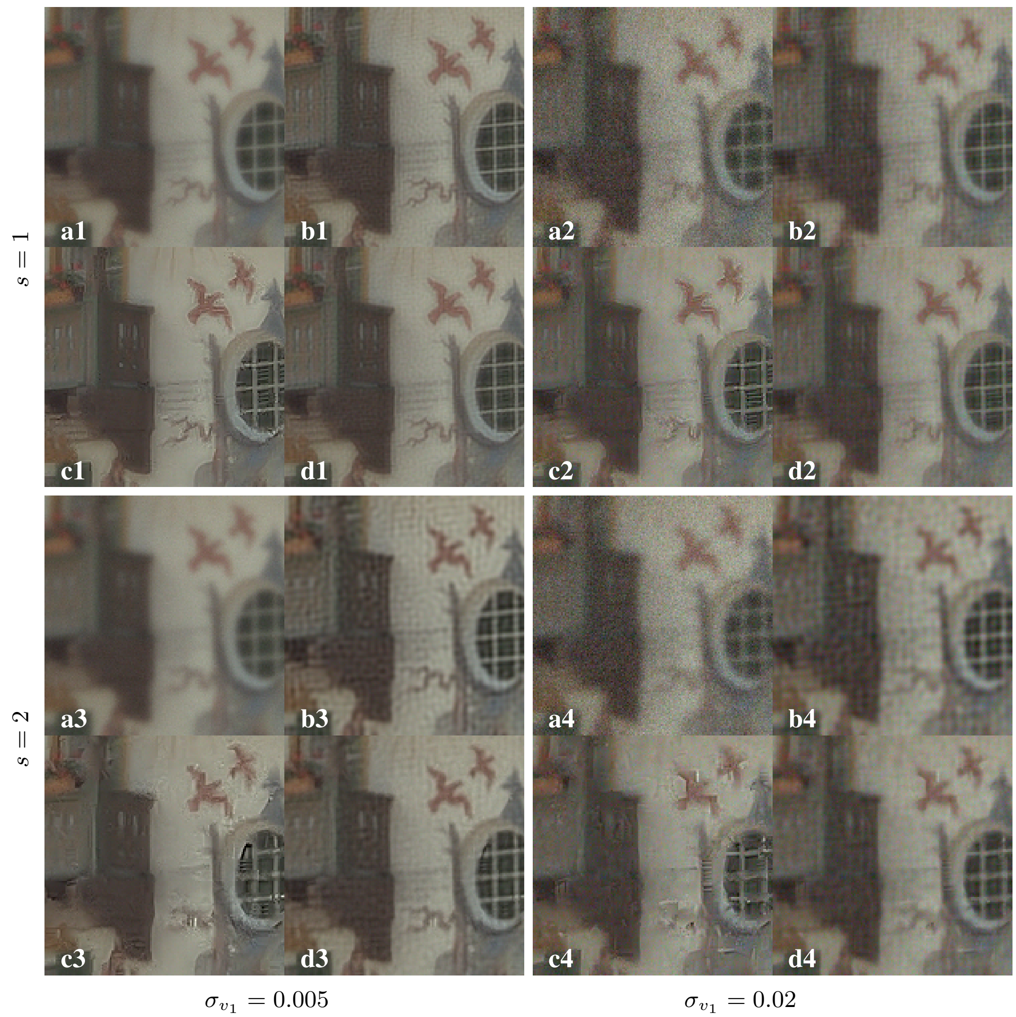


Fig. 6. Results for snippet 2 of Kodak image. The subblocks in each quadrant represent (a) the original blurred image, (b) the Wiener deconvolution estimate, (c) the NN reconstruction and (d) the ME Wiener estimate.

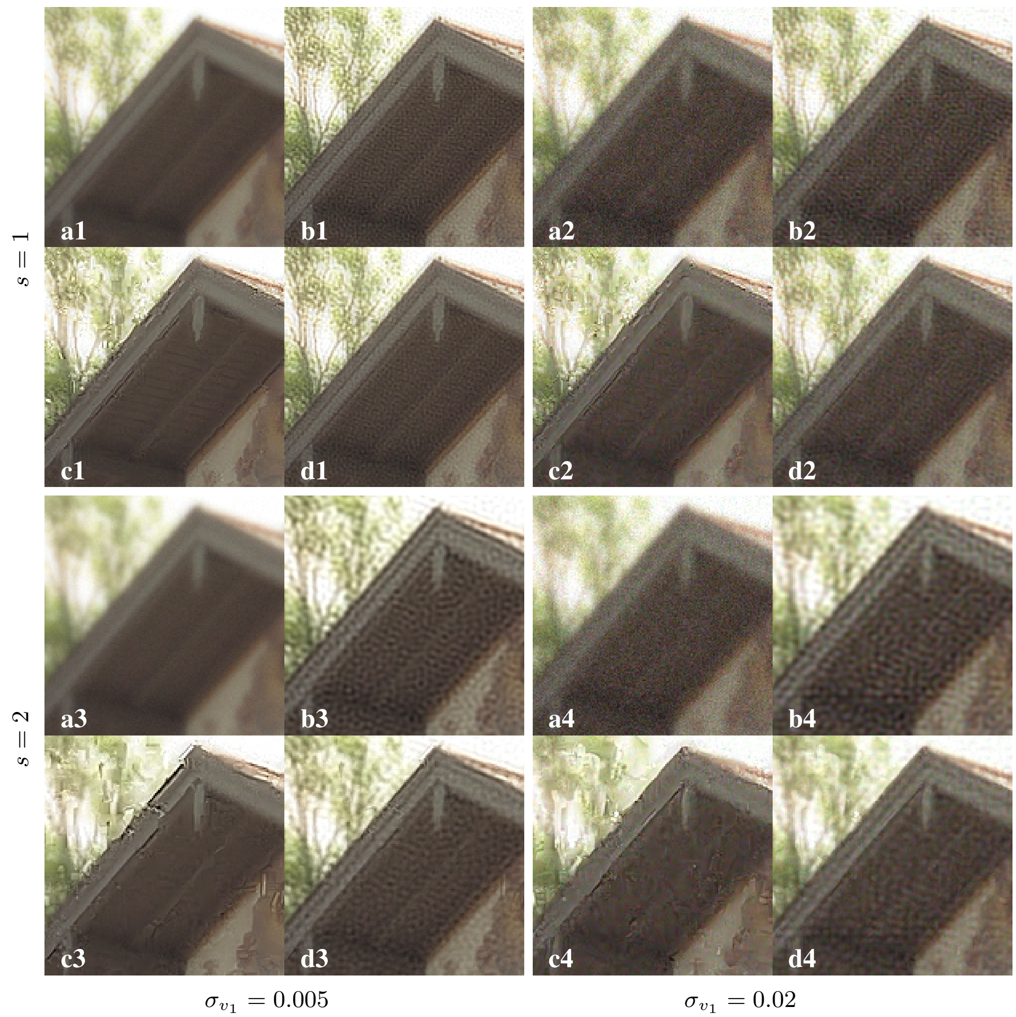
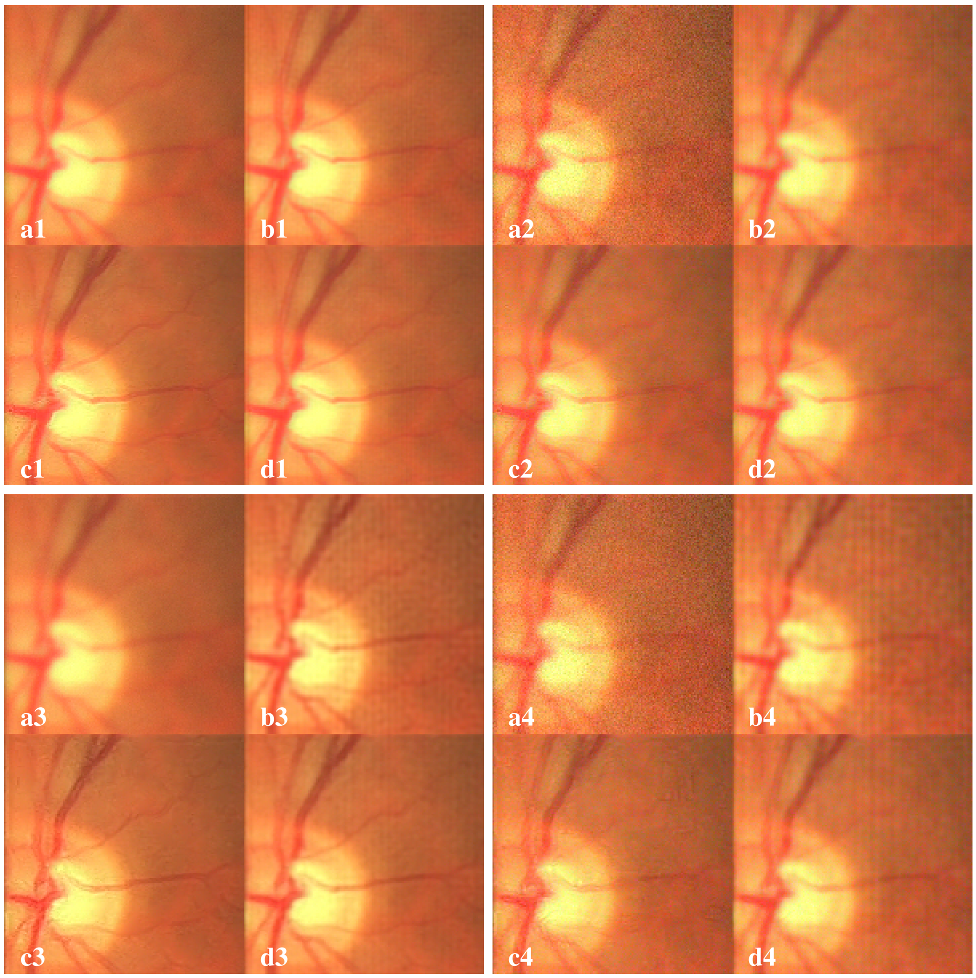


Fig. 7. Results for snippet3 of Kodak image. The subblocks in each quadrant represent (a) the original blurred image, (b) the Wiener deconvolution estimate, (c) the NN reconstruction and (d) the ME Wiener estimate.



$s = 1$

$s = 2$

$\sigma_{v_1} = 0.005$ $\sigma_{v_1} = 0.02$

Fig. 8. Results for snippet of retina image. The subblocks in each quadrant represent (a) the original blurred image, (b) the Wiener deconvolution estimate, (c) the NN reconstruction and (d) the ME Wiener estimate.

Table 2. PSNR (in dB) for estimates comparing different training images, and comparing with using two NN estimates for the images in Figs. 9 and 10. Highest PSNR in bold.

	Low blur, Low noise
Blurred/Noisy Original	28.8
Wiener	29.9
NN trained with three Kodak images	29.8
ME Wiener trained with three Kodak images	30.3
ME Wiener trained with Kodak image 1 and 2	32.4
ME Wiener trained with Kodak image 1 only	32.3
ME Wiener trained with Kodak image 2 only	32.4
ME Weiner w/ two NN estimates trained on Kodak image 1, and on Kodak image 2	**32.7**

NN Kodak training	NN Retina training	ME Wiener Kodak training	ME Wiener Retina training

Fig. 9. Comparison of NN and ME Wiener given either Kodak training images or retina training images.

ME Wiener Trained on 2 Kodak Images	ME Wiener Trained on Kodak Image 1	ME Wiener Trained on Kodak Image 2	ME Wiener Two NN Estimates Trained on Kodak Images 1 and 2

Fig. 10. Comparison of ME Wiener given two NN estimates.

6 CONCLUSIONS AND OPEN QUESTIONS

Theoretically, the proposed ME Weiner fusion is the optimal tradeoff (in MSE) between the deblurring capability of the Wiener filter and the high-frequency detail offered from the NN reconstruction method. Visually, it appears to be a reasonable tradeoff between accepting sharpness and rejecting image artifacts of the NN reconstruction estimate. However, minimizing mean-squared error should not be expected to be the optimal approach for producing aesthetically-pleasing images. One simple fix is to use a human visual model to determine the nearest-neighbor rather than squared error, but the results will then be dependent on viewing distance. Minimizing mean-squared error is a reasonable metric for producing images that are used for image analysis applications.

Here, we optimistically assumed knowledge of the required power spectrums. Although it is helpful to analyze algorithms when their assumptions are correct, future work would consider the impact of noisy estimates of the power spectrums.

We have demonstrated that the choice of training image(s) can result in significant changes in the quality of NN and ME Weiner reconstructions. In part to address this issue, we have hypothesized that using multiple NN estimates with different training sets can be useful, and provided some preliminary evidence supporting this hypothesis. However, further evidence and methodology would be needed, and this approach highlights the open question of how to estimate the error power spectrum for each of the NN estimates.

References

[1] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall (1993).

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

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

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

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

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

[7] E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," *Proc. SPIE* **3813**, 188–195 (1999).

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

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

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

[11] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int'l J. Comp. Vis.* **40**(1), 25–47 (2000).

[12] A. Banerjee, X. Guo, and H. Wang, "On the optimality of conditional expectation as a Bregman predictor," *IEEE Trans. Information Theory* **51**, 2664–2669 (2005).

[13] S. M. Bozic, *Digital and Kalman Filtering*, Halstead Press (1980).