Bias Modeling for Image Denoising

Priyam Chatterjee and Peyman Milanfar
Department of Electrical Engineering
University of California, Santa Cruz
1156 High Street, Mailstop SOE
Santa Cruz, CA 95064, USA
Email : {priyam, milanfar}@soe.ucsc.edu

***Abstract*—In this paper, we study the bias characteristics of image denoising algorithms. Recently introduced state-of-the-art denoising methods produce biased estimates of pixel intensities. The bias in each case is dependent on the underlying image geometry. Hence, we cluster the image into groups of patches that share a common underlying structure and study the bias independently in each cluster. We show that the bias in each cluster can be modeled effectively by an affine function, where the parameters of the model differ between clusters and algorithms. We validate our model through experimental results, both visually and quantitatively.**

I. Introduction

Image denoising is a well studied problem in image processing. Recently published papers [1–5] have amply illustrated the considerable progress that has been made in this field. Interestingly, all these methods produce biased estimates. In this work, we analyze these methods and their results as a first step to understanding the problem of denoising through the evaluation of the bias of the best current methods. The problem of denoising can be defined as an estimation problem where the actual pixel intensity (z_i) is to be estimated at each pixel location from its noisy observation

$$y_i = z_i + \eta_i \tag{1}$$

where η_i is the corrupting noise. Most modern denoising methods work on a group of contiguous image pixels (called patches) to obtain better results. The patch-wise data model can then be written as

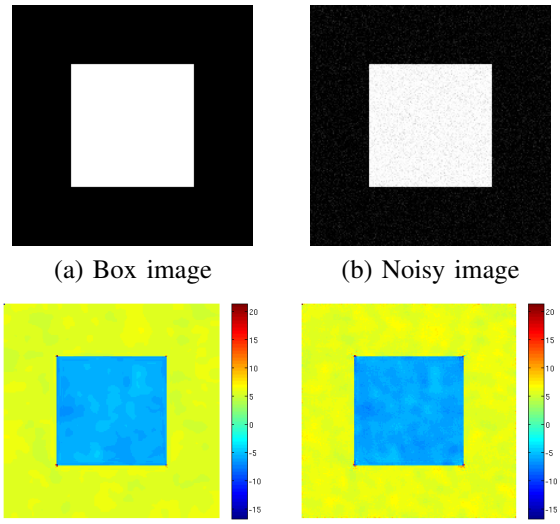
$$\mathbf{y}_i = \mathbf{z}_i + \boldsymbol{\eta}_i \tag{2}$$

where $\mathbf{y}_i$ is the (column stacked) vectorized version of the patch with y_i as its central pixel. The denoising problem can then be formulated as that of estimating the $\mathbf{z}_i$ vectors. In our work, we assume that $\mathbf{z}_i$ vectors are realizations of a vector random variable $\mathbf{z}$. Thus, the denoising bias for any given patch can be defined as

$$\mathbf{b}(\mathbf{z}_i) = E\left[\widehat{\mathbf{z}}_i\right] - \mathbf{z}_i \tag{3}$$

where $\widehat{\mathbf{z}}_i$ is an estimate of the vector $\mathbf{z}_i$.

In statistics, it is well known that the introduction of bias can lead to a lower mean squared error (MSE) [6]. This is achieved through lowering of the variance of the estimate. This is also true for most denoising methods, where certain parameters can be tuned to control the bias-variance tradeoff. For example, the bandwidth of kernel based methods [1, 4, 5, 7] can be adjusted to reduce the variance in the estimated denoised image, leading to a more visually pleasant smooth output. That modern denoising methods produce biased estimates can be readily seen from Fig. 1 where we show the bias obtained in denoising a very simple simulated box image from two of the popular denoising methods (namely, BM3D [2] and K-SVD [3]). In this paper, we analyze the bias characteristics of some recently introduced denoising methods [1–4] and propose a model for the bias. In the next section we show that the bias can be modeled to be an affine function, where the parameters of the model depend on the geometric structure of the patches. In Sec. III the bias model is verified qualitatively as well as quantitatively. We finally conclude in Sec. IV with a few words on the usefulness of modeling the bias.



(a) Box image (b) Noisy image

(c) BM3D bias (d) K-SVD bias

Fig. 1. Example of bias in denoising produced by some modern denoising methods: (a) Box image, (b) noisy image of standard deviation 15, and bias produced by (a) BM3D [2], and (d) K-SVD [3].

II. Modeling Denoising Bias

In this section we derive a model for the bias in image denoising. In Fig. 1, we show the bias produced by BM3D [2] and K-SVD [3] when denoising a simple simulated box

This work was supported in part by the U.S. Air Force under Grant F49620-03-1-0387.

image. Such bias is to be expected as the final denoised estimate of a patch is derived as a weighted average of other similar patches in the image. Mathematically, we can write the denoised estimate of any given patch (say $\mathbf{z}_i$) to be

$$\hat{\mathbf{z}}_i = \sum_{j=1}^{N} \mathbf{W}_{ij}\mathbf{y}_j, \tag{4}$$

where, the (data-dependent) weight matrix[1] $\mathbf{W}_{ij}$ provides a measure of similarity between patches $\mathbf{y}_i$ and $\mathbf{y}_j$ and N such similar patches are considered in the denoising process. In general, two noisy patches are deemed to be similar if they can be expressed in the form

$$\mathbf{y}_j = \mathbf{y}_i + \boldsymbol{\epsilon}_{ij} \quad \text{such that} \quad \|\boldsymbol{\epsilon}_{ij}\|^2 \leqslant \gamma \tag{5}$$

where γ is some small threshold value and $\boldsymbol{\epsilon}_{ij}$ is a vector. Using the data model of Eq. 2, and Eq. 5 above, we can express Eq. 4 as

$$\begin{aligned}\hat{\mathbf{z}}_i &= \sum_j \mathbf{W}_{ij}\mathbf{y}_j = \sum_j \mathbf{W}_{ij}(\mathbf{y}_i + \boldsymbol{\epsilon}_{ij}) \\ &= \sum_j \mathbf{W}_{ij}(\mathbf{z}_i + \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_{ij}).\end{aligned} \tag{6}$$

The expected value of this estimate can then be written as

$$\begin{aligned}E[\hat{\mathbf{z}}_i] &= E\left[\sum_j \mathbf{W}_{ij}(\mathbf{z}_i + \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_{ij})\right] \\ &= E\left[\sum_j \mathbf{W}_{ij}\right]\mathbf{z}_i + \sum_j E\left[\mathbf{W}_{ij}\left(\boldsymbol{\epsilon}_{ij} + \boldsymbol{\eta}_i\right)\right].\end{aligned} \tag{7}$$

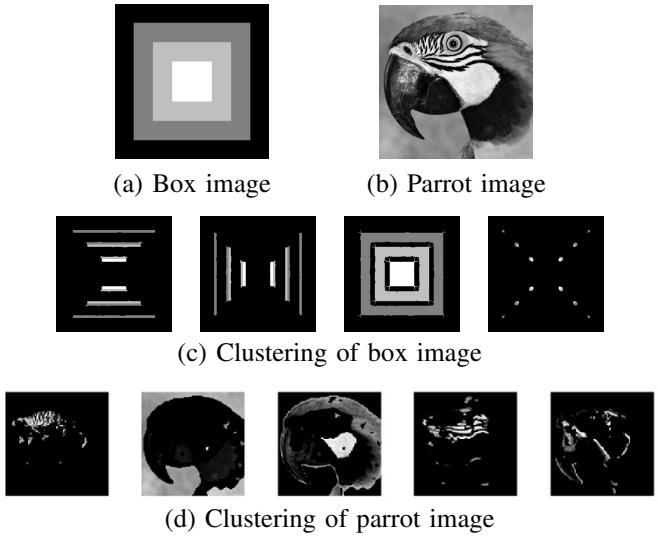
This allows us to calculate the bias of such non-linear weighted averaging methods as

$$\begin{aligned}\mathbf{b}(\mathbf{z}_i) &= E[\hat{\mathbf{z}}_i] - \mathbf{z}_i \\ &= \left(E\left[\sum_j \mathbf{W}_{ij}\right] - \mathbf{I}\right)\mathbf{z}_i + \sum_j E\left[\mathbf{W}_{ij}\left(\boldsymbol{\epsilon}_{ij} + \boldsymbol{\eta}_i\right)\right] \\ &= \mathbf{M}_i\mathbf{z}_i + \mathbf{u}_i\end{aligned} \tag{8}$$

where $\mathbf{M}_i = \left(E\left[\sum_j \mathbf{W}_{ij}\right] - \mathbf{I}\right)$ and $\mathbf{u}_i = \sum_j E\left[\mathbf{W}_{ij}\left(\boldsymbol{\epsilon}_{ij} + \boldsymbol{\eta}_j\right)\right]$. This provides us with an expression for the bias for each patch as an approximately affine (since $\mathbf{M}$ and $\mathbf{u}$, strictly speaking, depend on the data) function of $\mathbf{z}_i$. This is in keeping with the bias shown in Fig. 1 where it can be seen that the bias produced is a function of the underlying image structure. For example, in the smooth regions, the bias produced by either method is structurally approximately smooth. Also note that transitions in the bias occur along transitions in image structure, that is, at edges and corners for the box image of Fig. 1. Moreover, the parameters of the affine function, namely $\mathbf{M}_i$ and $\mathbf{u}_i$ are also dependent on the data. However, if we restrict ourselves to consider only patches that have similar geometric structure, these parameters can be effectively approximated by a single set of parameters (a matrix $\mathbf{M}$ and a vector $\mathbf{u}$). This leads us to model the bias along structurally similar regions of the image as

$$\mathbf{b}(\mathbf{z}_i) = \mathbf{M}\mathbf{z}_i + \mathbf{u}. \tag{9}$$

It is important to note that this model holds only when the patches considered are structurally similar, even though their intensities may differ. Thus, given an image, we first attempt to segment it into clusters such that each cluster contains patches that are geometrically similar. To obtain such clustering, we employ a geometric clustering technique that we proposed in [4]. As an example, Fig. 2 shows the clustering of some images into geometrically similar regions. Specifically, it can be seen (Fig. 2(c)) that the box image of Fig. 2(a) is divided into segments of horizontal and vertical edges, *flat* regions and corners. The bias in each cluster is then modeled as affine where the model parameters differ across clusters. In the next section we demonstrate that such a cluster-wise affine model is reflective of the bias produced by the current state-of-the-art denoising methods.



(a) Box image (b) Parrot image

(c) Clustering of box image

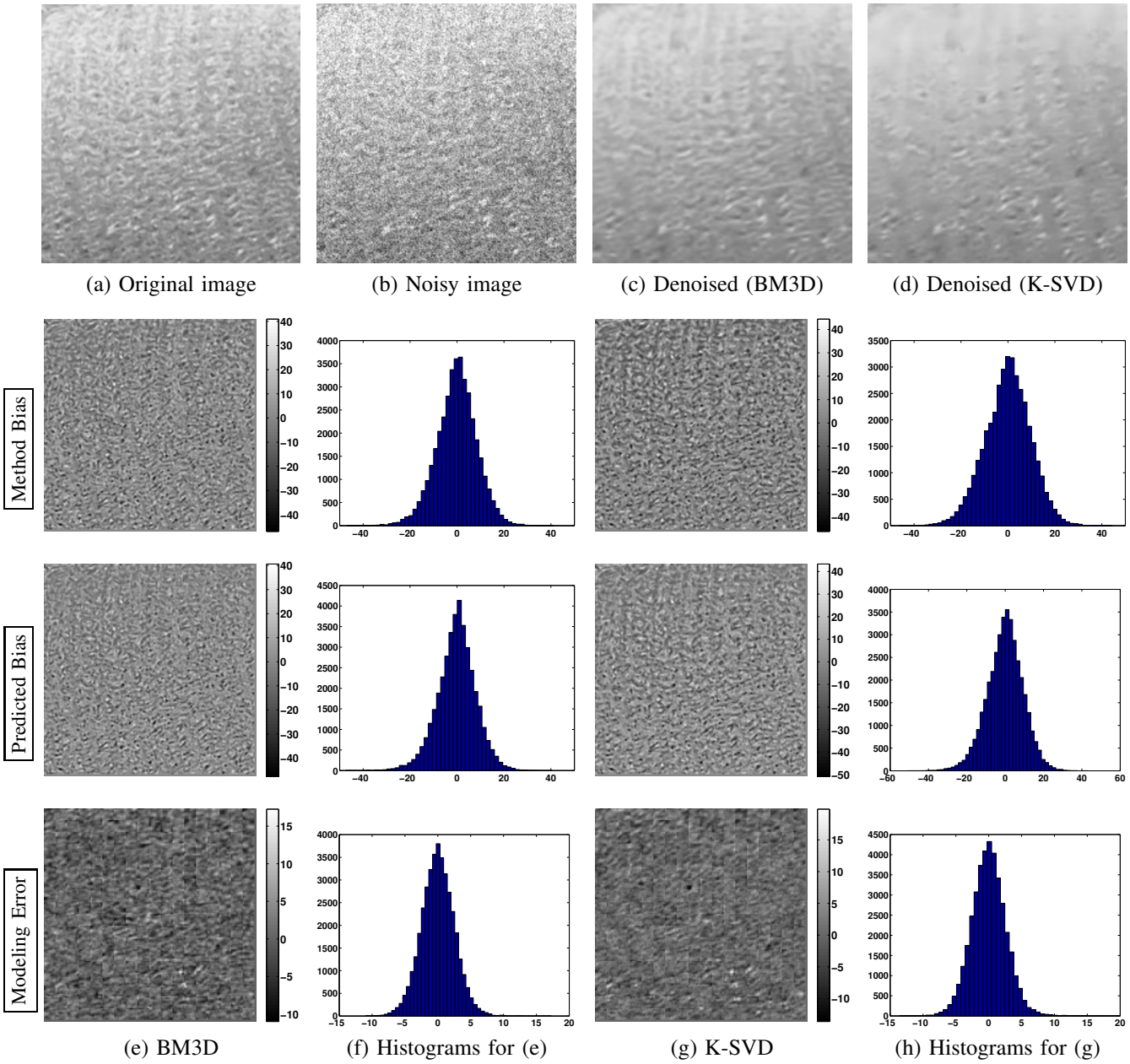
(d) Clustering of parrot image

Fig. 2. Clustering of a simple image based on geometric similarity. Note how pixels in any particular cluster can have quite different intensities but similar geometric structure.

III. Model Validation

Until now, we have argued that current non-linear denoising methods that perform denoising by a weighted averaging scheme produce biased estimates of the patches to be denoised and such bias is inherently an affine function of the underlying patch. Moreover, under the constraint of only geometrically similar patches being considered, a single set of parameters of the bias model (namely, $\mathbf{M}$ and $\mathbf{u}$) can be used to describe the bias for the entire group of patches. In this section we provide experimental validation for such an approximation. As a first step, we consider an image that contains patches that are roughly geometrically similar, although radiometrically the patches may be quite varied. Such images do not require any clustering to group together patches of similar geometry. An example of such an image is the towel image of size 200×200 shown in Fig. 3(a). There it can be seen that

[1]For most methods [1, 3–5, 7] that perform denoising in the spatial domain, a scalar measure of similarity between patches is used, which is a special case of Eq. 4.



(a) Original image (b) Noisy image (c) Denoised (BM3D) (d) Denoised (K-SVD)

(e) BM3D (f) Histograms for (e) (g) K-SVD (h) Histograms for (g)

Fig. 3. Visual comparison of the actual bias obtained from BM3D [2] & K-SVD [3] and reconstructed bias using affine model fit for the 200×200 towel image. The patch size chosen was 11×11. We can see that the histograms for the modeling errors in both the cases are centered around zero and have short tails.

although fine texture exists, the image patches are roughly similar in pattern across the entire image. In order to calculate the bias produced by each denoising method, we corrupt the towel image with 10 different realizations of additive white Gaussian noise (AWGN) of standard deviation 25 to produce 10 noisy images, one of which is shown in Fig. 3(b). Each of the noisy images are then denoised by each denoising method and the mean denoised image is estimated from which we calculate the bias for a particular method. The second row of Fig. 3 shows the bias in denoising and their histograms for two popular denoising methods, namely BM3D [2] and K-SVD [3]. Bias patches of size 11×11 are then used (in a vectorized form) to learn the affine model parameters (namely $\mathbf{M}$ and $\mathbf{u}$) of Eq. 9 in a least squares setting. Using the estimated $\widehat{\mathbf{M}}$ and $\hat{\mathbf{u}}$, the bias vectors are predicted. These predicted bias vectors are tiled and shown in an image form along with their corresponding pixel-wise histograms in the third row of Fig. 3. The modeling error is then calculated and shown as the difference between the calculated method bias and the predicted bias in each case. The tiled image form of the patch-wise model errors for each method and their corresponding pixel-wise histograms are shown in the last row of Fig. 3. There it can be seen that the affine model produces a reasonable approximation of the bias that is visually validated by the error histograms that are centered around zero with short tails.

Although the histograms act as fair qualitative indicators of the goodness of fit, we provide further quantitative eval-

(a) Grass image (b) House image (c) Parrot image

Fig. 4. Some examples of images that we will use to model the denoising bias of various methods.

uation for the effectiveness of the affine bias model. To this end, we make use of the coefficient of determination as the quantitative measure of the goodness of fit. The coefficient of determination [8] is mathematically defined as

$$R^2 = 1 - \frac{\sum_i \left\| \mathbf{b}(\mathbf{z}_i) - \widehat{\mathbf{b}}(\mathbf{z}_i) \right\|_2^2}{\sum_i \left\| \mathbf{b}(\mathbf{z}_i) - \overline{\mathbf{b}}(\mathbf{z}) \right\|_2^2} \tag{10}$$

where i indexes all the patches in the image, $\mathbf{b}(\mathbf{z}_i)$ is the actual bias of the estimated intensity of the i-th patch, $\overline{\mathbf{b}}(\mathbf{z})$ is the mean bias obtained by the denoising method across all patches in the image and $\widehat{\mathbf{b}}(\mathbf{z}_i) = \widehat{\mathbf{M}}\mathbf{z}_i + \hat{\mathbf{u}}$ is the predicted bias obtained from the estimated parameters $\widehat{\mathbf{M}}$ and $\hat{\mathbf{u}}$ of the affine model. Conceptually, the R^2 value provides a measure of the amount of variability in the bias vectors that is effectively explained by the affine model. The values of R^2 range from 0 to 1, where a higher value indicates a better model. For the bias modeling of the towel image of Fig. 3(a) and the much textured grass image (Fig. 4(a)), we obtained high R^2 values for some of the recently proposed denoising methods [1–4]. These values are tabulated in Table I.

Until now, we have shown experiments only on images where the patches are roughly geometrically homogeneous. However, this is not necessarily true for most general images (see Figures 4(b) & (c)). As mentioned in Sec. II, we deal with such images by performing a geometric clustering proposed in [4] to group together patches of similar geometry, irrespective of their radiometric intensities. For the house and parrot images of Fig. 4, the patches were grouped into 5 clusters. In each cluster of structurally similar patches, the bias is calculated and modeled and the coefficient of determination is calculated. In Table I, the average R^2 values across different clusters for these images are tabulated. There it can be seen that a high R^2 value is obtained for all the methods considered.

To further justify the necessity of restricting the bias model to patches displaying structural similarity, we devised another experiment with general images where patches were randomly sampled from an image and a single $\mathbf{M}$ and $\mathbf{u}$ were estimated from the sampled patches, irrespective of their structural dissimilarities. That is to say that we enforced an affine model for the bias over geometrically inhomogeneous patches. For all the images, we consistently obtained R^2 values that were substantially lower ($R^2 < 0.6$) than those reported in Table I. This supports our claim that such an affine model is a good approximation of the denoising bias only when patches of similar geometric structure are considered.

TABLE I

R^2 VALUES FOR THE AFFINE MODEL FIT OF THE BIAS PRODUCED BY DIFFERENT METHODS FOR DIFFERENT IMAGES CONSIDERING ADDITIVE WHITE GAUSSIAN NOISE OF STANDARD DEVIATION 25.

Image	BM3D [2]	K-SVD [3]	SKR [1]	K-LLD [4]
Towel	0.913	0.928	0.864	0.880
Grass	0.863	0.801	0.833	0.810
House	0.916	0.955	0.959	0.963
Parrot	0.957	0.963	0.946	0.954

IV. CONCLUSIONS

In this paper, we studied the bias produced by some of the recently published works on image denoising. This was done by formulating the image denoising problem in a patch-wise intensity estimation framework. Our experiments show that irrespective of the domain in which such estimation is performed, the denoising estimate can be modeled effectively to be affine. However, such a model holds only when the patches involved are geometrically similar. To validate the model, we performed various experiments that show, qualitatively and quantitatively, that such an affine model is a good approximation of the cluster-wise bias in denoising that is produced by recent state-of-the-art denoising methods.

Our motivation for modeling the bias stems from the fact that introduction of bias can offset the variance of an estimator resulting in decreased MSE of the denoised output. An interesting application of formulating an effective model for the bias is the ability to study the theoretical performance limits of denoising any given image. Such a study has enabled us to understand how the current state-of-the-art compares to the limits of denoising performance [9]. Moreover, given a model for the bias, one can also attempt to design a denoising method that optimizes the bias and hence improves on the MSE obtained by the current most popular denoising methods. The latter is a useful applications that we consider as a direction for our future research.

REFERENCES

[1] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel Regression for Image Processing and Reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 349–66, February 2007.

[2] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, August 2007.

[3] M. Elad and M. Aharon, "Image Denoising via Sparse and Redundant Representations over Learned Dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, December 2006.

[4] P. Chatterjee and P. Milanfar, "Clustering-Based Denoising with Locally Learned Dictionaries," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1438–1451, July 2009.

[5] C. Kervrann and J. Boulanger, "Optimal Spatial Adaptation for Patch-Based Image Denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 2866–2878, October 2006.

[6] S. Kay and Y. C. Eldar, "Rethinking Biased Estimation," *Signal Processing Magazine*, vol. 25, no. 3, pp. 133–136, May 2008.

[7] A. Buades, B. Coll, and J. M. Morel, "A Review of Image Denoising Methods, with a New One," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[8] N. Draper and H. Smith, *Applied Regression Analysis*, 3rd ed., ser. Probability and Statistics. Hoboken, N.J.: Wiley-Interscience, April 1998.

[9] P. Chatterjee and P. Milanfar, "Is Denoising Dead?" *IEEE Transactions on Image Processing*, in press.