Unsupervised, Information-Theoretic, Adaptive Image Filtering for Image Restoration

Suyash P. Awate, *Student Member*, *IEEE*, and Ross T. Whitaker, *Member*, *IEEE Computer Society*

Abstract—Image restoration is an important and widely studied problem in computer vision and image processing. Various image filtering strategies have been effective, but invariably make strong assumptions about the properties of the signal and/or degradation. Hence, these methods lack the generality to be easily applied to new applications or diverse image collections. This paper describes a novel unsupervised, information-theoretic, adaptive filter (UINTA) that improves the predictability of pixel intensities from their neighborhoods by decreasing their joint entropy. In this way, UINTA automatically discovers the statistical properties of the signal and can thereby restore a wide spectrum of images. The paper describes the formulation to minimize the joint entropy measure and presents several important practical considerations in estimating neighborhood statistics. It presents a series of results on both real and synthetic data along with comparisons with current state-of-the-art techniques, including novel applications to medical image processing.

Index Terms—Filtering, restoration, nonparametric statistics, information theory.

1 Introduction

The restoration of corrupted images is an important and widely studied problem in computer vision and image processing.[1] By *image restoration*, we mean the recovery of an image from a *degraded* version whose quality has been undermined by some stochastic process. Most research addresses the removal of additive, independent, random noise, which is also the focus of many of the results in this paper. However, such degradations can include a wide variety of processes such as correlated noise, spatially varying blurring ("smudging"), and nonstationary reductions in contrast.

Research in image restoration has led to a plethora of algorithms based on diverse strategies such as linear systems, statistics, information theory, and variational calculus. However, most of the image filtering strategies make strong assumptions about the properties of the signal and/or degradation. Therefore, they lack the generality to be easily applied to diverse image collections, and they break down when images exhibit properties that do not adhere to the underlying assumptions. Hence, there is still a need for general image filtering strategies that are effective for a wide spectrum of restoration tasks and are easily adaptable to new applications.

This paper describes a novel *unsupervised information-theoretic adaptive filter* (UINTA) for image restoration. It is a more complete version of some of our previous work [3]. UINTA restores pixels by comparing pixel values with other pixels in the image that have similar neighborhoods. The underlying formulation relies on an information-theoretic measure of goodness combined with a nonparametric model of image statistics. UINTA minimizes a penalty function that captures the entropy of the patterns of intensities in image regions. Entropy is a nonquadratic functional of the image intensities and, therefore, the filtering, obtained as the derivation of the entropy, is nonlinear. UINTA operates without a priori knowledge of the geometric or statistical structure of the signal, but relies instead on some general observations about the entropy of natural images. It does not rely on labeled examples to shape its output and is therefore *unsupervised*. UINTA automatically *learns* the true image statistics from the degraded input data and constructs a filtering strategy based on that model, making it *adaptive*. Moreover, UINTA adjusts virtually all its important internal parameters automatically using a data-driven approach and information-theoretic metrics. Because UINTA is nonlinear, nonparametric, adaptive, and unsupervised, it can restore a wide spectrum of images with very little parameter tuning.

The remainder of the paper is organized as follows: Section 2 discusses recent works in image filtering and their relationship to UINTA. Section 3 describes the mathematical formulation of UINTA and motivates the choice of the particular information-theoretic measure based on joint entropy. Entropy optimization entails the estimation of probability density functions (PDFs). Hence, Section 4 describes a nonparametric multivariate density estimation technique. It also describes the general problems associated with density estimation in high-dimensional spaces and provides some intuition behind the success of UINTA in spite of these difficulties. Section 5 formulates a gradient-descent scheme to optimize the joint-entropy measure and addresses several important practical issues pertaining to statistical estimation and its application to image neighborhoods. Section 6 gives experimental results and analyzes UINTA's behavior on numerous real and synthetic images. Section 7 summarizes the contributions of the paper and presents ideas for further exploration.

1. We mean *image* in the most general sense—a scalar or vector valued function defined on an n-dimensional domain, sampled on a dense, Cartesian grid.

• *The authors are with the School of Computing, University of Utah, 3190 Merril Engineering Building, 50 South Central Campus Drive, Salt Lake City, UT 84112. E-mail: {suyash, whitaker}@cs.utah.edu.*

Manuscript received 20 Sept. 2004; revised 27 May 2005; accepted 7 July 2005; published online 13 Jan. 2006.
Recommended for acceptance by J. Goutsias.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0497-0904.

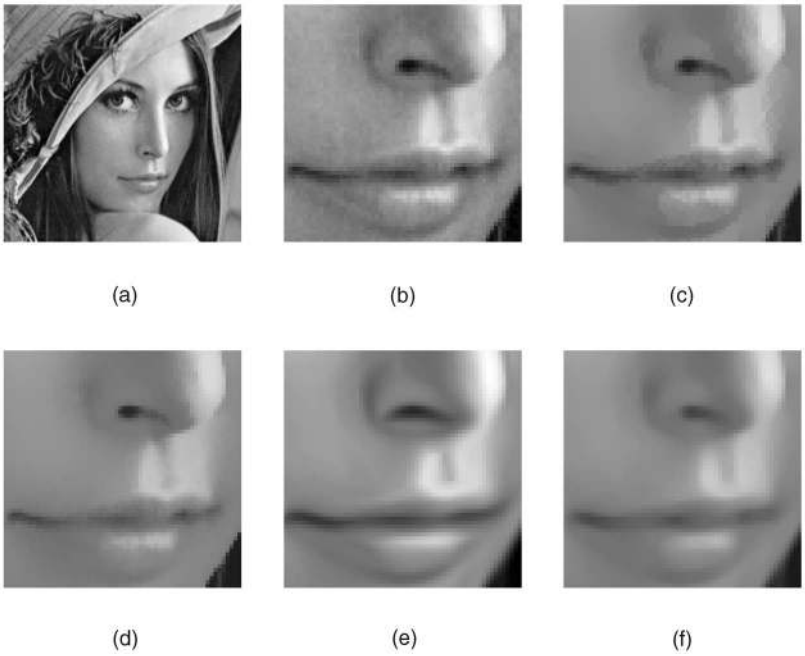

Fig. 1. (a) Degraded *Lena* image: gray-scale values range: 0-100 gray-scale unit (G.U.). Zoomed insets of: (b) the degraded image and the results with PDE-based methods, e.g., (c) anisotropic diffusion: $K = 0.5\mathrm{G.U.s}$, 20 iterations, (d) bilateral filtering: $\sigma_{domain} = 3$ pixels, $\sigma_{range} = 12$ G.U., (e) coherence enhancing diffusion: $\sigma = 0.1$ pixels, $\rho = 2$ pixels, $\alpha = 0.0001$, $C = 0.0001$, 15 iterations, and (f) curvature flow: time step = 0.2, eight iterations.

2 Related Work

The literature on signal and image restoration is vast and a comprehensive review is beyond the scope of this paper. This section establishes the relationship of this work to several important, relevant areas of nonlinear image filtering. Nonlinear filtering approaches are typically based on either variational methods, leading to algorithms based on partial differential equations (PDEs), or statistical methods, leading to nonlinear estimation problems.

PDE-based image processing methods became widespread after the work of Perona and Malik [27], where they propose a modified version of the heat equation (calling it *anisotropic diffusion*) that adapted the diffusivity to image features. The anisotropic diffusion equation is also the first variation of an image energy [25], [39] that seeks piecewise constant solutions (in 1D—the situation is somewhat more complex in multiple dimensions). Because such variational approaches prefer certain image geometries, we refer to these local geometric configurations as *models*. A multitude of nonlinear PDE models have been developed for a wide variety of images and applications [32], [46], including the total variation model by Rudin et al. [33], PDE versions [6] of the Mumford and Shah [24] variational model, the cartoon-texture model by Vese and Osher [43], the coherence enhancing flow by Weickert [47], and various algorithms based on level sets [2], [36], [41], [26], [6]. These nonlinear PDE models have proven to be very effective, but only for particular applications where the input data is well suited to the model's underlying geometric assumptions. Moreover, the parameter tuning is a challenge because it entails fuzzy thresholds that determine which image features are enhanced and which are smoothed away.

Statistical formulations have given rise to a wide variety of image filters. For instance, the median and other order-statistics on image neighborhoods can be quite effective [23]. Tomasi and Manduchi [42] describe a bilateral filter, which does a robust averaging in Gaussian-weighted image neighborhoods. A great deal of image processing work develops from a stochastic model of image structure given by Markov random fields (MRFs). Geman and Geman [18] exploit the equivalence between MRFs and Gibbs distributions to model images with Gibbs distributions, in which case, the optimal image estimate is given as a fixed point of an iterative procedure that relies on neighborhood-dependent updates. The conditional probabilities for image neighborhood configurations, namely, *cliques*, play a similar role to the image energy in the variational approaches. The most widely used models penalize intensity differences, but simultaneously estimate hidden parameters that explicitly model intensity edges, which pushes the iterative process toward piecewise-constant solutions. The cliques in the MRF approach encode a set of probabilistic assumptions (priors) about the geometric properties of the signal and, thus, they are effective only when the signal conforms sufficiently well to the prior. UINTA also exploits the Markov property of the images, but in a different context. Rather than imposing a particular model on the image, UINTA *learns* the relevant conditional PDFs from the input data and updates pixel intensities to decrease the randomness of these conditional PDFs.

Fig. 1 demonstrates the effects of such strong models on image filtering[2] by showing the effects of some of the prevalent nonlinear techniques on the *Lena* image. Anisotropic diffusion (Fig. 1c) restores the cheeks but introduces spurious edges near the nose and the lips. Bilateral filtering [42] tends to smooth away fine textures, resulting in their elimination, e.g., on the lips in Fig. 1d. Both of these algorithms entail two free parameters, i.e., scale and contrast, and require significant tuning. The coherence enhancing diffusion forces specific elongated shapes in images, as seen in the enlarged nostril and the lips' curves in Fig. 1e. On the

2. Please refer to the electronic copy to notice/distinguish subtle features/differences in images.

other hand, Fig. 1f shows the curvature flow [36], [26], which is very similar to the total variation strategy of [33], that tends to shrink features by rounding them off. The *Lena* image, which appears to be a very typical gray-scale photograph, does not adhere very well to the basic geometric models underlying these algorithms.

An alternative to filtering with variational models is to construct nonlinear transforms in the frequency domain. In this context, the wavelet literature addresses image denoising extensively. The current state-of-the-art wavelet denoising methods [31], [35], [28], [40] treat the wavelet coefficients as random variables and model their a priori marginal/joint PDFs parametrically. They then estimate the coefficients of the noiseless image given the observed coefficients of the noisy image via various schemes such as Bayesian estimation. The limitations of these methods stem both from the choice of the particular wavelet decomposition basis and the parametric models imposed on the coefficients. A very recent work [30] aims at the blind removal of correlated Gaussian/mesokurtotic noise using Gaussian-scale-mixture signal models in the wavelet domain. It adapts to the noise statistics by estimating the noise covariance from the input image. The sparse-code shrinkage strategy [21] chooses the transformation based on the statistical properties of the data, using noiseless training data, in order to concentrate the energy in only a few components and then shrinking the sparse component values akin to wavelet-based methods.

Researchers analyzing the statistics of natural images in terms of local neighborhoods also draw conclusions that are consistent with MRF models of images. For instance, Huang and Mumford [20] find the mutual information between the adjacent pixel intensities to be rather large, attributing this to the presence of spatial correlation in images. Lee et al. [22] and de Silva and Carlsson [11] analyze the statistics of 3×3 patches in images, in the corresponding high-dimensional spaces, and find the data to be concentrated in clusters and low-dimensional manifolds exhibiting a nontrivial topology. UINTA also relies on the hypothesis that natural images exhibit some regularity in neighborhood structure and it discovers that for each image individually in a nonparametric manner.

The literature shows several statistically-based image processing algorithms that do rely on information theory. The *mean-shift* algorithm [17], [38], [7], [9], [5] modifies image intensities so that they move uphill on the PDF associated with the gray-scale histogram of the image. At steady state (assuming appropriate windowing strategies), all samples converge to the nearest mode. The mean-shift procedure, thus, can be said to be a *mode seeking* process. However, the mean-shift algorithm operates only on image intensities (be they scalar or vector valued) and does not account for neighborhood structure in images. Thus, mean shift resembles a kind of data-driven thresholding process, particularly in the algorithm proposed by [9], in which the density estimate is static as the algorithm iterates. This paper shows the mathematical relationship between the mean-shift procedure and entropy reduction and thereby shows UINTA to be a generalization of the mean-shift algorithm, which incorporates image neighborhoods to reduce the entropy of the associated conditional PDFs.

Weissman et al. [48] propose the DUDE algorithm that addresses the problem of denoising data sequences generated by a discrete source and received over a discrete, memoryless channel. It assumes no knowledge of the source statistics and yet performs (asymptotically) as well as any denoiser (e.g., one that knows the source statistics), thereby making DUDE *universal* and *optimal*. DUDE assigns image values based on the similarity of neighborhoods gathered from image statistics, which resembles the construction of conditional probabilities in UINTA. However, the DUDE approach does not account for noise in the neighborhoods that are used to condition the probabilities for the reconstruction and it is limited to discrete-valued signals. While the DUDE algorithm is demonstrably effective for removing *replacement noise*, it is less effective in case of additive noise, which is more important in scientific applications.

There are some examples in the literature of algorithms that learn the statistics of image neighborhoods. Popat and Picard [29] were among the first to use nonparametric Markov sampling in images. Their approach models the higher-order nonlinear image statistics via cluster-based, nonparametric density estimation and they apply it to image restoration, image compression, and texture classification. However, their approach relies on training samples, which limits its practical use—the proposed method learns the statistics of the signal directly from the degraded data. The literature dealing with *texture synthesis* also sheds some light on the principles underlying UINTA. Texture synthesis algorithms rely on image statistics from an input image to construct novel images that bear a qualitative resemblance to the input [12], [15], [45]. This is a different application and these algorithms do not rely on information-theoretic formulations, but they demonstrate the power of neighborhood statistics and mode-seeking processing in capturing essential aspects of image structure.

3 Joint Entropy-Based Image Filtering

This section presents the mathematical formulation of UINTA, which relies on a random-field image model. It concludes with a high-level algorithmic description of UINTA.

3.1 Random-Field Image Model

A random field [13] is a family of random variables $X(\Omega; T)$, for some index set T, where, for each fixed $T = t$, the random variable $X(\Omega; t)$ is defined on the sample space Ω. If we let T be a set of points defined on a discrete Cartesian grid and fix $\Omega = \omega$, we have a realization of the random field called the *digital image*, $X(\omega, T)$. In this case, $\{t\}_{t \in T}$ is the set of pixels in the image. For two-dimensional images, t is a two-vector. If we fix $T = t$ and let ω vary, then $X(t)$ is a random variable on the sample space. We denote a specific realization $X(\omega; t)$ (the intensity at pixel t) a deterministic function $x(t)$.

If we associate with T a family of pixel neighborhoods $N = \{N_t\}_{t \in T}$ such that $N_t \subset T$, $t \notin N_t$, and $u \in N_t$ if and only if $t \in N_u$, then N is called a neighborhood system for the set T. Points in N_t are called neighbors of t. We define a random vector $Y(t) = \{X(t)\}_{t \in N_t}$ corresponding to the set of intensities at the neighbors of pixel t. We also define a random vector $Z(t) = (X(t), Y(t))$ to denote image regions, i.e., pixels combined with their neighborhoods. For the formulation in this paper, we assume a stationary ergodic process (in practice, this assumption can be relaxed somewhat, as explained in Section 5.2). We denote the original (not

degraded) image by $X(\omega, T)$ and its associated set of neighborhood intensities by $Y(\omega, T)$ and regions by $Z(\omega, T)$. Correspondingly, for the observed degraded image, we use $\tilde{X}(\omega, T)$, $\tilde{Y}(\omega, T)$, and $\tilde{Z}(\omega, T)$. For notational simplicity, we use the shorthand for random variables $X(t)$ as X and their realizations $x(t)$ as x, dropping the index t.

3.2 Neighborhood Entropy for Image Filtering

The UINTA strategy is to reduce the entropy, $h(\tilde{X}|\tilde{Y} = \tilde{y})$, of the conditional PDF for each pixel-neighborhood pair, $(\tilde{X} = \tilde{x}, \tilde{Y} = \tilde{y})$, by manipulating the value of each center pixel $\tilde{x}$. For this, UINTA employs a gradient descent strategy. Note that a gradient descent on $h(\tilde{X}|\tilde{Y})$ has components corresponding to both the center pixel $\tilde{x}$ and the neighborhood $\tilde{y}$ and, thus, the entire region, $(\tilde{x}, \tilde{y})$, could be updated in a gradient descent scheme. In practice, however, we update only the center pixel $\tilde{x}$, that is, we project the gradient onto the direction associated with the center pixel. Given this projection, UINTA is a reweighted gradient descent on either the joint entropy $h(\tilde{X}, \tilde{Y})$ or the conditional entropy $h(\tilde{X}|\tilde{Y})$—they are equivalent for this particular descent strategy.

The choice of entropy as the optimization measure follows from several observations. We assume independent and identically distributed (IID) additive zero-mean noise (in some of our other work, we explore adapting UINTA to other specific noise models [4]). Then, the addition of two independent random variables, i.e., the signal and additive noise, increases the entropy [37], [10]. Entropy reduction reduces the randomness in corrupted PDFs and tries to counteract noise. Of course, continued entropy reduction might also eliminate some of the normal variability in the signal (original image). However, we have found that noise-free images tend to have very low entropy relative to their noisy counterparts. Therefore, entropy reduction first affects random degradations substantially more than the signal. Furthermore, the entropy reduction is limited by an information-based stopping criterion, as described in Section 5.2.3.

3.3 High-Level Structure of the UINTA Algorithm

This section gives a high-level view of the UINTA algorithm.

1. The input degraded image I has a set of intensities $\{\tilde{x}\}_{t \in T}$, neighborhoods $\{\tilde{y}\}_{t \in T}$, and regions $\{\tilde{z}\}_{t \in T} = \{(\tilde{x}, \tilde{y})\}_{t \in T}$. These values form the initial values ($I^0 = I$) of a sequence of images $I^0, I^1, I^2, \ldots$, with corresponding intensities $\hat{x}^0, \hat{x}^1, \hat{x}^2, \ldots$.
2. For each image region $\hat{z}^m$, compute
$$\partial h(\hat{X}|\hat{Y} = \hat{y}^m)/\partial \hat{x}^m.$$
3. Construct image I^{m+1}, using finite forward differences on the gradient descent, with intensities $\hat{x}^{m+1} = \hat{x}^m - \lambda \partial h/\partial \hat{x}^m$, where λ is the time step explained in Section 5.2.1.
4. Check stopping criteria, as explained in Section 5.2.3. If not done, go to Step 2, otherwise, I^{m+1} is the output.

Although each step of the UINTA algorithm operating on a single pixel in Step 3 is merely a gradient descent on the center pixel, the interactions from one iteration to the next are quite complex. The updates on the center-pixel intensities in Step 3 affect, in the next iteration, not only the center pixels but also the neighborhoods. This is because the image regions overlap and the set of pixels that form the centers of regions is the same as that which form the neighborhoods. Thus, UINTA filtering consists of two kinds of processes. One is the first-order *optimization process*, which computes updates for pixels based on their neighborhoods. The other second-order process causes updates of the neighborhoods based on the role of those pixels as centers in the previous iteration.

4 Nonparametric Multivariate Density Estimation

Entropy optimization entails the estimation of higher-order conditional PDFs. This introduces the challenge of high-dimensional, scattered-data interpolation, even for modest sized image neighborhoods. High-dimensional spaces are notoriously challenging for data analysis (regarded as *the curse of dimensionality* [38], [34]) because they are so sparsely populated. Despite theoretical arguments suggesting that density estimation beyond a few dimensions is impractical, the empirical evidence from the literature is more optimistic [34], [29]. The results in this paper confirm that observation. Furthermore, stationarity implies that the random vector $\tilde{Z} = (\tilde{X}, \tilde{Y})$ has identical marginal PDFs, thus lending itself to more accurate density estimates [34], [38]. Also, UINTA relies on the neighborhoods in natural images having a lower-dimensional topology in the multidimensional feature space [22], [11]. Therefore, in the feature space, *locally* the PDFs of images are lower-dimensional entities that lend themselves to better density estimation.

We use the Parzen-window nonparametric density estimation technique [14] with an n-dimensional Gaussian kernel $G_n(\tilde{z}, \Psi_n)$. Having no a priori information on the structure of the PDFs, we choose an isotropic Gaussian, i.e., $\Psi_n = \sigma I_n$, where I_n is the $n \times n$ identity matrix. Using optimal values of the Parzen-window parameters is critical for success and that can be difficult in such high-dimensional spaces; we describe a method in Section 5.2 for automatically choosing this parameter.

For a stationary ergodic process, the multivariate Parzen-window estimate is

$$P(\tilde{Z} = \tilde{z}_i) \approx \frac{1}{|A_i|} \sum_{t_j \in A_i} G_n(\tilde{z}_i - \tilde{z}_j, \Psi_n), \tag{1}$$

where the set A_i is a small subset of T, chosen at random for each t_i, and $\tilde{z}_j$ is shorthand for $\tilde{z}(t_j)$. Thus, each iteration of UINTA requires a computation time $O(|A_i||T|)$ and results in a stochastic approximation for the PDFs.

4.1 A Stochastic Approximation for Entropy

Entropy is the expectation of negative log-probability and, therefore, we can approximate it with the sample mean [44]. For a stationary ergodic process,

$$h(\tilde{Z}) = -E_P[\log P(\tilde{Z})] \approx -\frac{1}{|T|} \sum_{t_i \in T} \log \left[\frac{1}{|A_i|} \sum_{t_j \in A_i} G_n(\tilde{z}_i - \tilde{z}_j, \Psi_n) \right]. \tag{2}$$

The set A_i, which generates the density estimate $P(\tilde{z}_i)$, should not contain the point t_i itself—because this biases the entropy estimates. Typically, the samples in set A_i are a small fraction of those in T, chosen at random. This has two

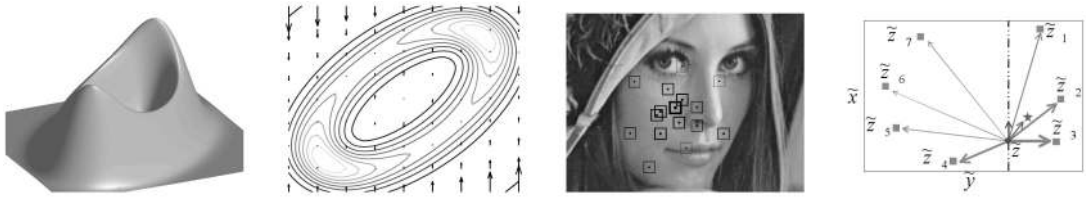

Fig. 2. (a) An example 2D PDF, $P(\tilde{x}, \tilde{y})$, on feature space, $< \tilde{x}, \tilde{y} >$. (b) A contour plot of the PDF depicts the forces (vertical arrows) that reduce the entropy of the conditional PDFs $P(\tilde{X}|\tilde{Y} = \tilde{y})$, as in (3). (c) Some pixels in A_i (black dots) along with the neighborhoods (squares around the dots) yielding feature space samples $\tilde{z}_i$. The thickness of the squares indicate the weights, as in (4), for the intensities of pixels in A_i. The thickest square denotes the neighborhood around the pixel being processed. (d) Attractive forces (arrow width $\equiv$ force magnitude) act on a sample ($\tilde{z}$: circle) toward other samples ($\tilde{z}_j$: squares) in the set A_i, as per (4). The resultant force acts toward the weighted mean (star) and the sample $\tilde{z}$ moves based on its projection (vertical arrow).

important implications. First, it significantly reduces the computational cost for the entropy estimation, from $O(|T|^2)$ to $O(|A_i||T|)$. Second, because A_i is different for each element of T for each iteration, the entropy estimate $h(\tilde{Z})$ is stochastic. Hence, a gradient descent entropy optimization technique results in a stochastic-gradient algorithm [19]. The stochastic-gradient alleviates the effects of spurious local maxima introduced in the Parzen-window density estimate [44].

5 Entropy Minimization on Conditional PDFs

Entropy minimization in UINTA relies on the derivative of the entropy with respect to $\tilde{x}_i \equiv \tilde{x}(t_i)$ for each $t_i \in T$. Each $\tilde{x}_i$ undergoes a gradient descent based on the entropy of the conditional PDF estimated from A_i. The gradient descent is

$$\frac{\partial \tilde{x}_i}{\partial \tau} = -\frac{\partial h(\tilde{X}|\tilde{Y} = \tilde{y}_i)}{\partial \tilde{x}_i} \approx \frac{1}{|T|}\frac{\partial \log P(\tilde{x}_i|\tilde{y}_i)}{\partial \tilde{x}_i} = \frac{1}{|T|}\frac{\partial \log P(\tilde{z}_i)}{\partial \tilde{x}_i} \quad (3)$$

$$= -\frac{1}{|T|}\frac{\partial \tilde{z}_i}{\partial \tilde{x}_i}\sum_{t_j \in A_i}\frac{G_n(\tilde{z}_i - \tilde{z}_j, \Psi_n)}{\sum_{t_k \in A_i} G_n(\tilde{z}_i - \tilde{z}_k, \Psi_n)}\Psi_n^{-1}(\tilde{z}_i - \tilde{z}_j), \quad (4)$$

where $\partial \tilde{z}_i / \partial \tilde{x}_i$ is a projection operator that projects an n-dimensional vector $\tilde{z}_i$ onto the dimension associated with the center pixel intensity $\tilde{x}_i$ and τ is a dummy evolution parameter. Fig. 2 elucidates this process.

If we choose a $\sigma > 0$ (and finite), the entropy for a finite set of samples is always bounded. We perform a (projected) gradient descent on a bounded energy function and, thus, for sufficiently small time steps, the process converges. Steady-state convergence is rarely the goal, however, and stopping criteria are presented as a practical consideration in Section 5.2.3.

5.1 Relationship to the Mean-Shift Procedure

The mean-shift procedure [17], [38], [7], [9] moves each sample in a feature space to a weighted average of other samples using a weighting scheme that is similar to Parzen windowing. We can also view this as moving samples uphill on a PDF. Comaniciu and Meer [9] propose an iterative mean-shift algorithm for image intensities, where the PDF does not change with iterations, for image segmentation. Each gray-scale or vector pixel intensity is drawn toward a local maximum in the corresponding PDF.

This section shows how UINTA relates to the mean-shift procedure. Consider, as an example, a gradient descent on the entropy of the gray-scale pixel intensities. This gives

$$\frac{\partial \tilde{x}_i}{\partial \tau} = -\lambda\frac{\partial h(\tilde{X})}{\partial \tilde{x}_i} \approx -\frac{\lambda}{|T|}\sum_{t_j \in A_i}\frac{G_1(\tilde{x}_i - \tilde{x}_j, \Psi_1)}{\sum_{t_k \in A_i} G_1(\tilde{x}_i - \tilde{x}_k, \Psi_1)}\Psi_1^{-1}(\tilde{x}_i - \tilde{x}_j). \quad (5)$$

Finite forward differences, $\tilde{x}^{m+1} = \tilde{x}^m - \lambda \partial h / \partial \tilde{x}^m$, with a time step $\lambda = |T|\sigma^2$ give

$$\tilde{x}_i^{m+1} = \tilde{x}_i^m + \left[\frac{\sum_{t_j \in A_i}\tilde{x}_j^m G_1(\tilde{x}_i^m - \tilde{x}_j^m, \Psi_1)}{\sum_{t_k \in A_i} G_1(\tilde{x}_i^m - \tilde{x}_k^m, \Psi_1)} - \tilde{x}_i^m\right] = \sum_{t_j \in A_i}\tilde{x}_j^m W_j(\tilde{x}_i^m, \tilde{x}_j^m, A_i, \Psi_1). \quad (6)$$

Thus, each new pixel value is a weighted average of a selection of pixel values from the previous iteration with weights $W_j(\cdot) > 0$ and $\sum_j W_j(\cdot) = 1$. Taking $A_i = T$ gives exactly the mean-shift update proposed by Fukunaga and Hostetler [17]—note that, in UINTA, the PDFs on which the samples climb get updated after every iteration. Thus, the mean-shift algorithm is a gradient descent on the entropy associated with the gray-scale intensities of an image. We observe that samples $\tilde{x}$ are being attracted toward every other sample, with a weighting term that diminishes with the distance between the two samples. The UINTA updates have the same form, except that the weights are influenced not only by the distances between intensities $\tilde{x}$, but also by the distances between the neighborhoods $\tilde{y}$. That is, pixels in the image with similar neighborhoods have a relatively larger impact on the weighted mean that drives the updates of the center pixels.

5.2 Implementation Issues

The UINTA algorithm as presented in previous sections presents a number of significant engineering issues that are important to its effectiveness. These concern the setting of various free parameters, accounting for the nonstationary statistics of real images, and avoiding artifacts associated with the grid and grid boundaries.

5.2.1 Update Rates

The update via gradient descent introduces a constant, the time step λ, which controls the proportionality between the derivatives of the energy and the rate at which the solution is updated. As discussed in Section 5.1, a time step of $\lambda = |T|\sigma^2$ gives rise to a mean-shift on the conditional PDFs. However,

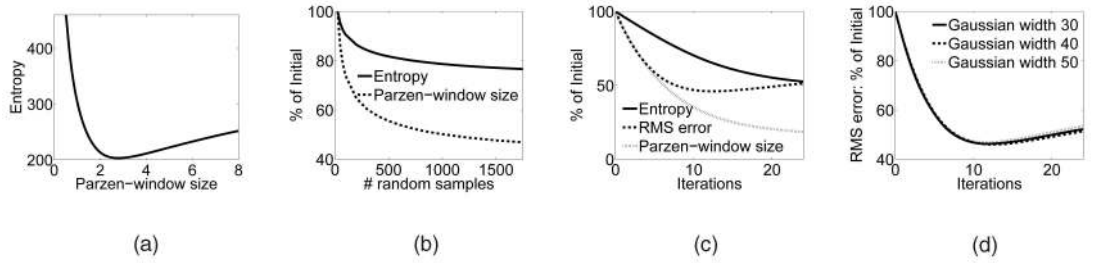

Fig. 3. (a) $h(\tilde{X},\tilde{Y})$ as a function of σ ($|A_i| = 1,000$) (for the *Lena* image in Fig. 1). (b) $h(\tilde{X},\tilde{Y})$ and σ are almost unaffected for $|A_i| > 1,000$. To give smoother curves, each measurement, for a particular $|A_i|$, is averaged over three different random sets A_i. The following graphs use the *Lena* image corrupted with IID additive Gaussian noise $N(0, 100)$. (c) $h(\tilde{X},\tilde{Y})$, σ and RMS error. (d) Effect of local sampling scheme on UINTA filtering.

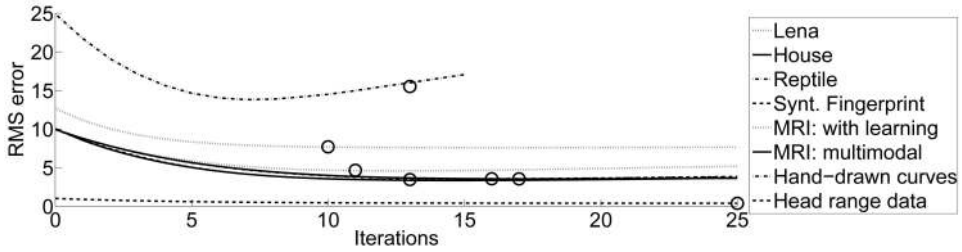

Fig. 4. RMS errors versus iterations for several images (see Section 6) with varying additive-noise levels. The circles represent the points where the residual equals the noise level.

because of the interactions between neighborhoods from one iteration to the next, that time step can cause the algorithm to overshoot and oscillate in an unstable manner. Because the pixels in the neighborhoods each perform a mean shift, the resulting motion in the feature space can be as much as $\sqrt{|N_t|+1}$ times the mean shift update, where $|N_t|$ is the number of pixels in the neighborhood. Thus, a stable, reliable time step is $\lambda = |T|\sigma^2/\sqrt{|N_t|+1}$. We have found that this is more conservative than it needs to be and we have used $\lambda = 0.2|T|\sigma^2$ for all of the results in this paper.

5.2.2 Parameters for Density Estimation

Parzen windowing, using a finite number of samples, is very sensitive to the value of σ [14]. Many algorithms/applications with low dimensional features spaces (e.g., 2 or 3) operate by manually tuning the scale parameter. However, because UINTA relies on a sparsely populated high dimensional space, it is very difficult to manually find values for σ that allow the samples to interact properly without excessively smoothing the PDF. Also, UINTA being iterative and dynamic, the best scale parameter changes every iteration. UINTA finds σ via a data-driven approach. Because the goal is to minimize joint entropy, a logical choice is to choose a value for σ that minimizes the same. Fig. 3a confirms the existence of a unique minimum. Fig. 3b shows that, for sufficiently large $|A_i|$, the choice of σ is *not* sensitive to the value of $|A_i|$, thereby enabling UINTA to automatically choose $|A_i|$ to an appropriate value before the filtering begins. We have implemented both differential (Newton's method) and discrete (Fibonacci search) methods for finding the optimal scale and both offer acceptable results. Fig. 3c depicts the decreasing trend for σ as the filtering progresses, which is common to every example and is consistent with UINTA's entropy-reducing action bringing samples closer in the feature space.

5.2.3 Stopping Criteria

An analysis of simple examples shows the existence of nontrivial steady states, e.g., an image which is a discrete sampling of a linear function such as a ramp. Empirical evidence shows that the filtering algorithm does sometimes converge to interesting results. However, for most applications, convergence to a fixed point is not a useful goal. As with many other iterative filtering strategies, several iterations of the gradient descent are sufficient for acceptable restoration, but this requires either parameter tuning or the definition of suitable stopping criteria.

The choice of stopping criteria for this algorithm depends on a number of factors. For instance, in the absence of *any* knowledge of the signal, noise, or other types of degradation, the algorithm will inevitably require some parameter tuning. We assume that noiseless images have conditional PDFs with low entropy and degradations substantially increase this randomness. We have found empirically (and it seems reasonable) that entropy reduction via gradient descent starts by counteracting the randomness introduced by the noise much more than reducing the inherent randomness in the signal. Thus, an effective strategy is to stop when the relative rate of change of entropy, from one iteration to the next, falls below some threshold.

When the level of additive noise is known, UINTA can iterate until the root-mean-square (RMS) difference (residual) between input and the processed image equals the noise level. We have found empirically that this method is quite effective (see Fig. 4) and we have used this approach in all of the examples for which noise levels are known.

5.2.4 Piecewise Stationarity and Local Sampling

In practice, image statistics are not homogeneous and are more accurately modeled as piecewise stationary-ergodic. Thus, the set of samples A_i used for processing pixels $\tilde{x}_i$ should consist of pixel neighborhoods that are spatially nearby. To achieve this, we choose a unique set of samples for each pixel at random using a Gaussian distribution on the

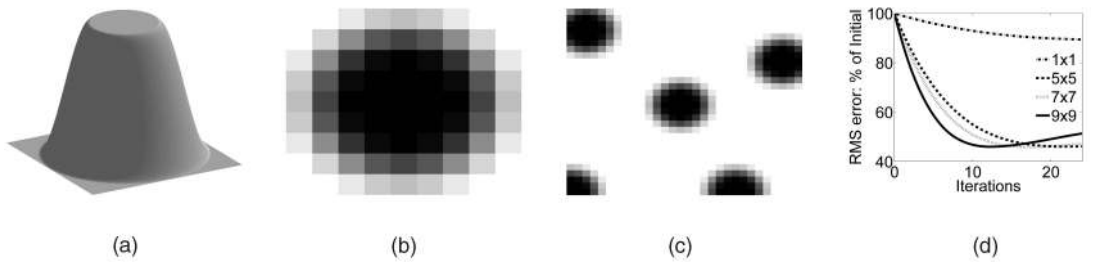

Fig. 5. (a) Preserving rotational invariance via a neighborhood mask consisting of a flat central circular plateau with cubic splines on the sides. (b) The discrete sampling of the mask (black $\equiv$ 1, white $\equiv$ 0) for a 9×9 neighborhood. (c) Anisotropic neighborhoods at boundaries. (d) Effect of neighborhood size on UINTA filtering (for the *Lena* image corrupted with IID Gaussian $N(0, 100)$ noise).

image coordinates, centered at the pixel in question with $\sigma = 40$. Thus, the set A_i is biased and contains more pixels near the pixel being processed. This strategy gives consistently better results than uniform sampling. Furthermore, we have found that it performs well for any choice of σ that encompasses more than several hundred pixels. Fig. 3d also shows that UINTA's performance degrades gracefully for suboptimal values of this parameter.

5.2.5 Rotational Invariance and Boundary Conditions

Square neighborhoods generate results with artifacts exhibiting preferences for grid-aligned features. A solution is to weight the intensities, making neighborhoods more isotropic. UINTA incorporates such fuzzy weights by using an anisotropic feature-space distance metric, $\|z\|_M = \sqrt{z^T M z}$, where M is a diagonal matrix with the elements being the appropriate weights on the influence of the neighbors on the center pixel. For the random vector Z to have identical marginal PDFs that aid in density estimation, as explained in Section 4, we require the weights to be somewhat homogeneous. Figs. 5a and 5b shows a disk-shaped mask that achieves this balance. The intensities near the center are unchanged ($m_i = 1$), while the intensities near the corners are weighted down gradually (via cubic-spline interpolation) to zero. The proposed isotropic mask is a gray-scale version of the DUDE [48] strategy of using a binary disc-shaped mask for discrete (half-toned) images. Note that scaling the center-pixel intensity more than its neighbors leads to an elongated space $(\tilde{X}', \tilde{Y}')$ along $\tilde{X}'$—in the limit, when all neighbors are weighted zero, leading to a thresholding as in the mean-shift algorithm [9]. Fig. 5d also shows that UINTA's performance degrades gracefully for suboptimal values of the neighborhood size.

Typical image boundary conditions, e.g., replicating pixels or toroidal topologies, can produce neighborhoods that distort the feature-space statistics. UINTA handles boundary neighborhoods by collapsing the feature space along the dimensions corresponding to the neighbors falling outside the image. UINTA crops the square regions crossing image boundaries and processes them in the lower-dimensional subspace, as in Fig. 5c. This strategy results in important modifications in two stages of UINTA. First, the cropped intensity vectors take part in a mean-shift process reducing entropies of the conditional PDFs in the particular subspace where they reside. Second, UINTA chooses the Parzen window size, σ, based only on the regions lying completely in the image interior.

6 Experiments and Results

This section gives experimental results on numerous real and synthetic images along with the analysis of UINTA's behavior and qualitative and quantitative comparisons with the state-of-the-art wavelet methods. The only parameters that UINTA exposes to the user are: 1) the size of the neighborhoods and 2) the width of the Gaussian that defines the extent from which local samples are taken for density estimation (for truly stationary images, samples would be global). Empirical results show that UINTA's performance degrades gracefully—no drastic effects as in typical PDE-based filtering schemes—for suboptimal values of these parameters. We use masked, rotationally symmetric, 9×9 neighborhoods, as described in Section 5.2.5. Parzen windowing in all of the examples uses a local Gaussian random sampling (standard deviation 40 pixels) in the image domain with 1,000 samples (i.e., $|A_i| = 1,000$), as explained in Section 5.2.4. We use additive zero-mean IID Gaussian noise. We recompute the size of the Parzen window σ after each iteration, as explained in Section 5.2.2.

All original images have intensities ranging from 0 to 100. As a visualization aid for comparing different images/results, the intensities of all images within a set have been consistently rescaled to span the available range of intensities. The UINTA implementation in this paper relies on the Insight Toolkit [1].

Fig. 6 shows the result of UINTA filtering on the *Lena* image. UINTA preserves and enhances fine structures, such as strands of hair or feathers in the hat, while removing random noise without imposing a piecewise constant intensity profile. The results are noticeably better than any of those obtained using other methods shown in Fig. 1. Fig. 6 also shows the results of processing an MRI image of a human head. These examples show UINTA's ability to adapt to a variety of gray-scale features in real images approximated by piecewise stationary models.

The fingerprint image in Fig. 7 is an example where the degradation involves smudges (blurring) and is clearly not additive noise. UINTA enhances the light and dark lines without significant shrinkage. UINTA performs a kind of multidimensional classification of neighborhoods—therefore, some features in the top-left are lost because they resemble the background more than the ridges. For the stopping criteria, we use the relative change in entropy as described in Section 5.2.3. Fig. 7 also presents the results with other restoration strategies for visual comparison with UINTA. The piecewise smooth image models associated with anisotropic smoothing, bilateral filtering, and curvature

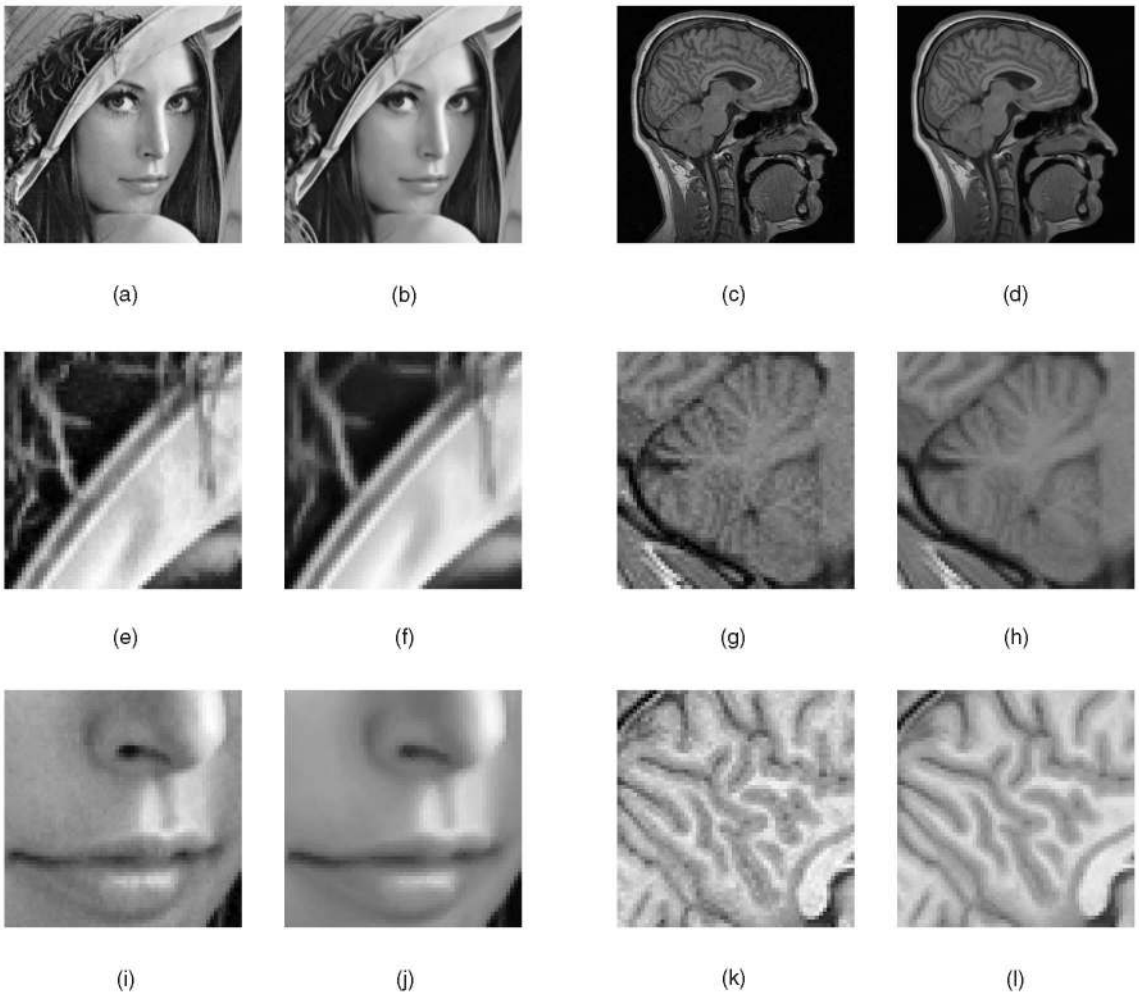

Fig. 6. (a) and (c) Noisy images: *Lena* and MRI head. (b) and (d) UINTA filtered images. (e), (g), (i), (k) and (f), (h), (j), (l) show magnified portions of the degraded and filtered images, respectively.

flow (Figs. 7g, 7h, and 7i) are clearly inappropriate for this image. The coherence enhancing filter (Fig. 7j) does not succeed in retaining or enhancing the light-dark contrast boundaries. It also forces some elongated structures to grow or connect. An unrestricted mean-shift filtering (Fig. 7k) on image intensities (with the PDF not changing with iterations) yields a thresholded image, while retaining most of the noise. Wavelet denoising (Fig. 7l) is unable to get rid of the smudges and excessively smoothes other regions of the image.

Fig. 8 gives an example of restoring the standard *House* image corrupted with IID additive noise $N(0, 100)$. Table 1 shows the RMS errors with the standard test images of the *House*, *Lena*, *Barbara*, and *Peppers*. The wavelet denoising technique yields a lower RMS error for this image, but introduces ringing-like artifacts in smooth regions.

Fig. 9 shows the application of UINTA to an image of hand-drawn curves (noise $N(0, 625)$). The noise level is high enough so that thresholding cannot yield the original image. UINTA learns the pattern of black-on-white curves and forces the image to adhere to this pattern. However, UINTA does make mistakes when curves become too close, exhibit very sharp bends, or when the noise introduces ambiguous gaps. The wavelet denoised image depicts significant artifacts around the edges, giving a higher RMS error (Table 1).

The entropy reduction associated with UINTA *does* impose a kind of statistical simplification on the image and that statistical simplicity corresponds, in many cases, to geometric simplicity. Fig. 10 shows the results of many UINTA iterations on the hand-drawn image of Fig. 9a. UINTA has no explicit geometrical model and yet it gradually smooths out the kinks in these curves producing progressively simpler geometric structures. The entropy of straighter curves is lower because of reduced variability in the associated neighborhoods. The result is qualitatively similar to that of curvature-reducing geometric flows [36], [26], suggesting a strong link between variational and statistical characterizations of images.

In order to better analyze the behavior of UINTA and compare its performance with state-of-the-art wavelet denoisers, we present results with a diverse collection of synthetic images. We provide examples on the simulated fingerprint image (Fig. 11a), the simulated range data of the human head (Fig. 11e), and the synthetic *Reptile* image [16] (Fig. 11i). Table 1 shows the RMS errors. UINTA performs better on the fingerprint, almost equally well on the range data and poorer on the *Reptile* image. Thus, it seems that UINTA performs better as a denoiser when it can find sufficiently many patterns in the degraded image to be able to distinguish the degradation from the underlying signal.

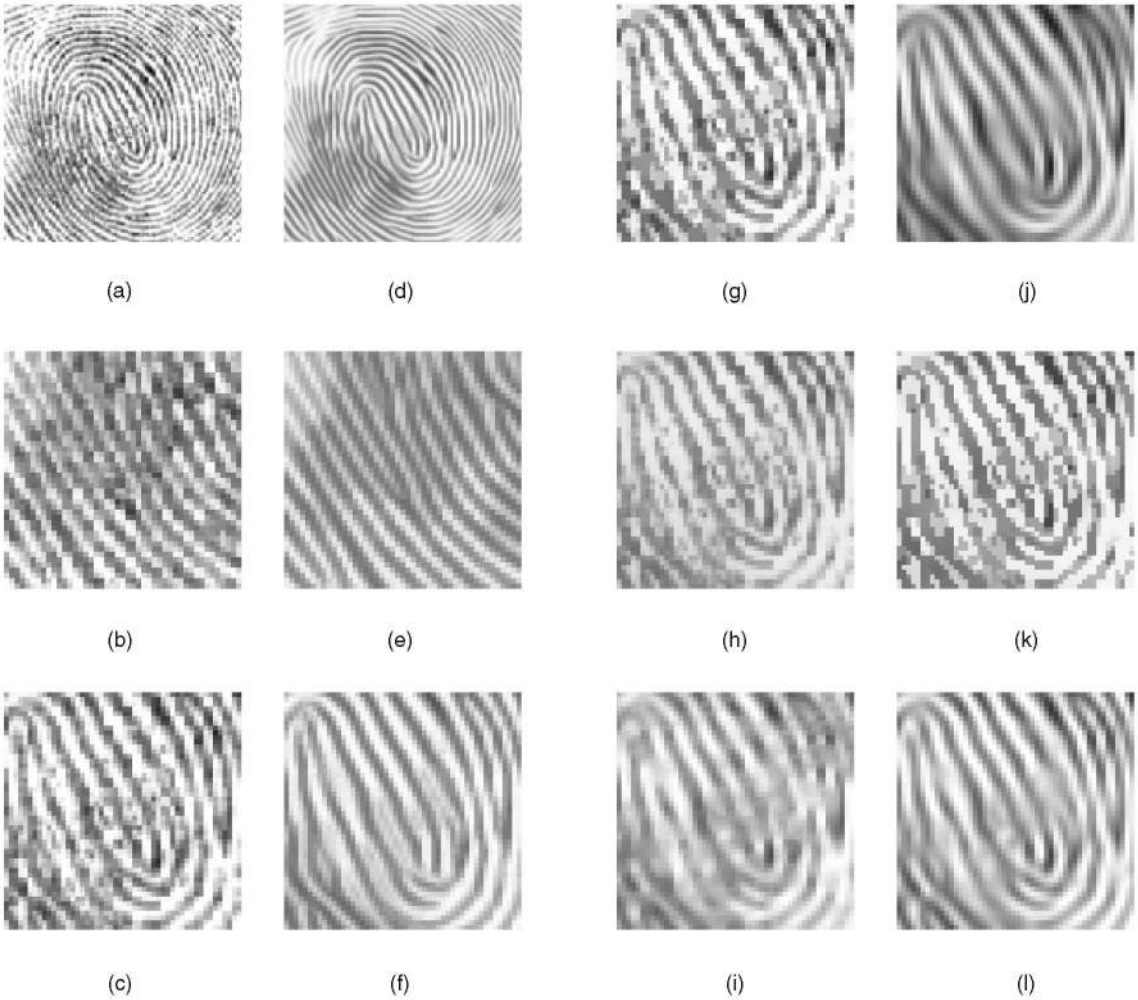

Fig. 7. (a) Degraded fingerprint image with (b), (c) zoomed insets. (d) UINTA restored image with (e), (f) zoomed insets. Zoomed insets of the fingerprint image processed with (g) anisotropic diffusion: $K = 0.45$ gray-scale values, 99 iterations, (h) bilateral filtering: $\sigma_{\text{domain}} = 3$ pixels, $\sigma_{\text{range}} = 15$ gray-scale values, (i) curvature flow: time step = 0.2, five iterations, (j) coherence enhancing diffusion: $\sigma = 0.1$ pixels, $\rho = 2$ pixels, $\alpha = 0.0001$, $C = 0.0001$, 15 iterations, (k) unrestricted mean shift [5]: $\sigma_{\text{domain}} = 2$ pixels, $\sigma_{\text{range}} = 5$ gray-scale values, five iterations, and (l) wavelet denoising [31]: $\sigma_{\text{noise}} = 14$ gray-scale values.

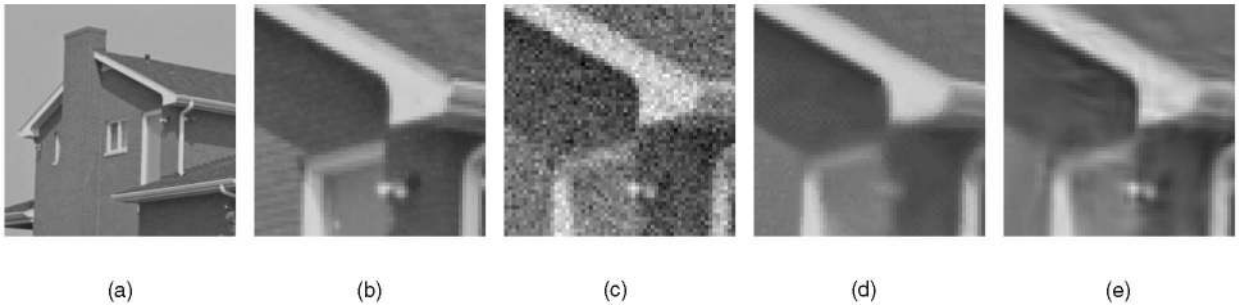

Fig. 8. (a) *House* image and its (b) zoomed inset. Zoomed insets of the (c) noisy image, (d) UINTA filtered image, and (e) Wavelet denoised [31] image.

Moreover, the statistical models underlying the wavelet denoisers are empirically derived from photographs, like the *Reptile* image. Fig. 12 shows a real image of a building facade that exhibits a certain degree of redundancy. UINTA is able to exploit that to perform almost as well as the best wavelet denoiser in terms of RMS error (see Table 1) and with fewer visual artifacts.

When operating within a specific application domain, UINTA can perform much better by learning from ideal or noiseless image examples. Fig. 13 shows a demonstration of this concept on simulated MRI data from the BrainWeb [8] project. We corrupt a head MRI T1 image with IID additive Gaussian noise and use two other similar, but not identical, images for learning the neighborhood statistics of typical brain MR images. Fig. 13a shows one of the two images representing the nonparametric prior model. This example shows the power of such learning—the UINTA restored image exhibits structures that are barely visible in the degraded version and fares considerably better than the wavelet denoiser, both qualitatively and quantitatively.

TABLE 1
RMS Errors Comparing UINTA with the Current State-of-the-Art Wavelet Denoisers

Example	Initial RMS error	UINTA	[31]	[35]	[28]
Standard image: *House*	10.0	3.5	2.9	3.1	3.5
Standard image: *Lena*	10.0	4.6	3.6	3.8	4.1
Standard image: *Barbara*	10.0	4.8	3.8	4.2	4.5
Standard image: *Peppers*	10.0	4.5	3.5	3.7	3.9
Hand-drawn curves	25.0	15.4	16.0	18.5	18.0
Simulated fingerprint	10.0	3.4	4.1	4.7	4.7
Simulated range data (head)	1.0	0.35	0.34	0.36	0.5
Reptile	10.0	3.5	2.9	3.0	3.4
Building Facade	10.0	4.5	4.4	5.1	5.4
MRI (with learning)	10.0	3.1	3.4	3.7	3.9
MRI (multimodal)	10.0	3.3	3.4	3.7	3.9

Note: The standard test images of Barbara and Peppers [31] do not appear in this paper.

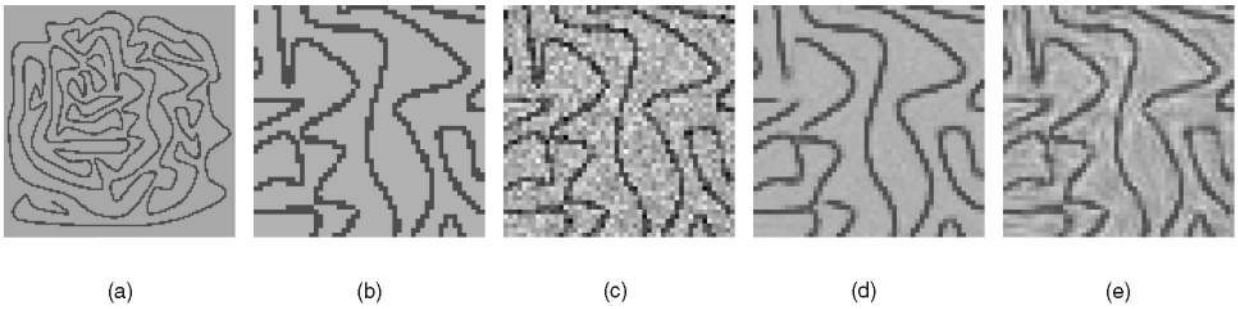

Fig. 9. (a) Hand-drawn curves with a (b) zoomed inset. Zoomed insets of the (c) noisy image, (d) UINTA-filtered image, and (e) wavelet denoised [31] image.

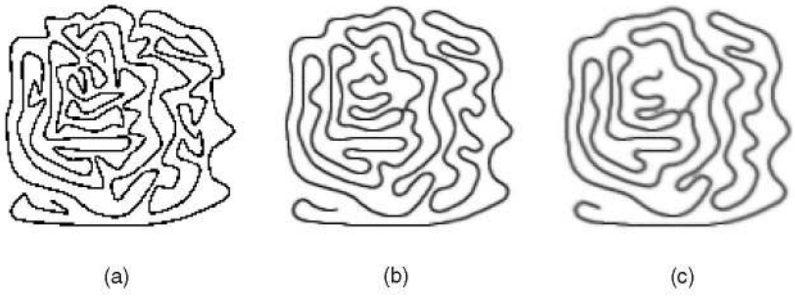

Fig. 10. (a) Hand-drawn curves. (b) and (c) show UINTA filtered images after 100 and 200 iterations, respectively.

The UINTA formulation also generalizes easily to simultaneous restoration of a sequence of images, e.g., multimodal MRI, exploiting the relationships between images to further enhance performance. Fig. 14 shows an example with multimodal restoration. This entails a simultaneous restoration of T1, T2, and PD images in a coupled manner, treating the combination of three images as an image of vectors and looking at PDFs in the *combined* probability space. Although, in this paper, we show results with multimodal images that are well aligned, our experiments suggest that the restoration is fairly robust to minor registration errors. Here again, UINTA fares better than the wavelet denoiser.

7 Conclusions and Discussion

UINTA is a novel, unsupervised, information-theoretic, adaptive filter that improves the predictability of pixel intensities from the intensities in their neighborhoods by decreasing the joint entropy. UINTA can preserve and enhance structures in a way that resembles many nonlinear, variational filters, but does so without any explicit geometric image model. Because it is nonparametric, it can adapt to the statistics of the input image and, therefore, it applies quite readily to new applications with little parameter tuning. The stochastic gradient-descent algorithm for minimizing joint entropy entails density estimation in high-dimensional spaces and relies on Parzen windowing with automatic parameter selection. In order to be effective for image processing the UINTA algorithm operates with a feature-space metric that preserves rotational symmetry and allows for boundary conditions. The UINTA algorithm is a generalization of the mean-shift classification algorithm [17], [9], [7] that conditions the distribution based on the pixel neighborhood. Results show that the statistics of image neighborhoods are sufficiently regular for reliable image restoration.

Despite these promising results, this paper presents only a preliminary implementation that could benefit from some engineering advances. For instance, the method of density estimation with single-scale isotropic Parzen-window kernels is clearly insufficient for all situations and it is reasonable that kernels be chosen adaptively to accommodate the signal

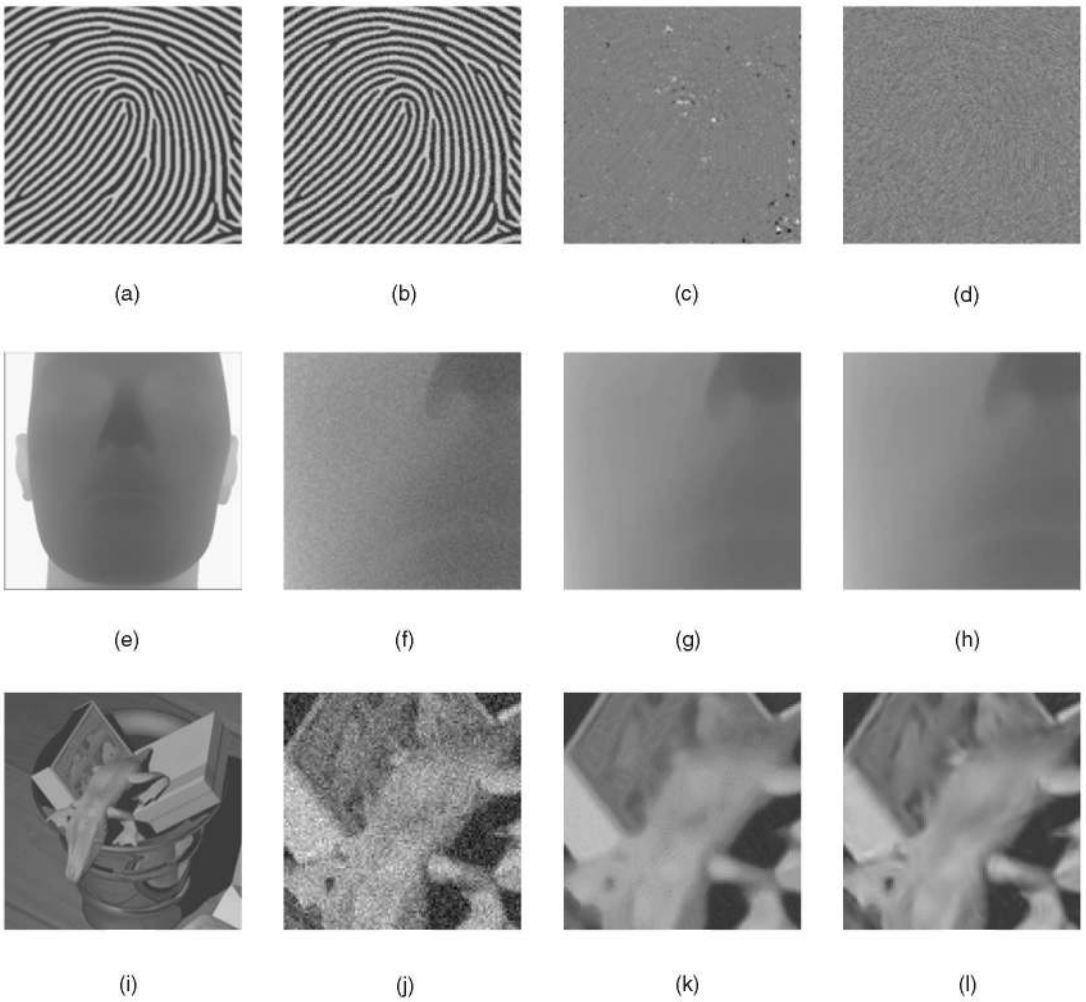

Fig. 11. (a) Simulated fingerprint image. (b) Noisy image. Difference between the filtered and the noiseless images for (c) UINTA and (d) the wavelet denoiser [31]. (e) Head range data. (i) Reptile image [16]. Zoomed insets of the (f)-(j) noisy images, (g)-(k) UINTA filtered images, and (h)-(l) wavelet denoised images [31].

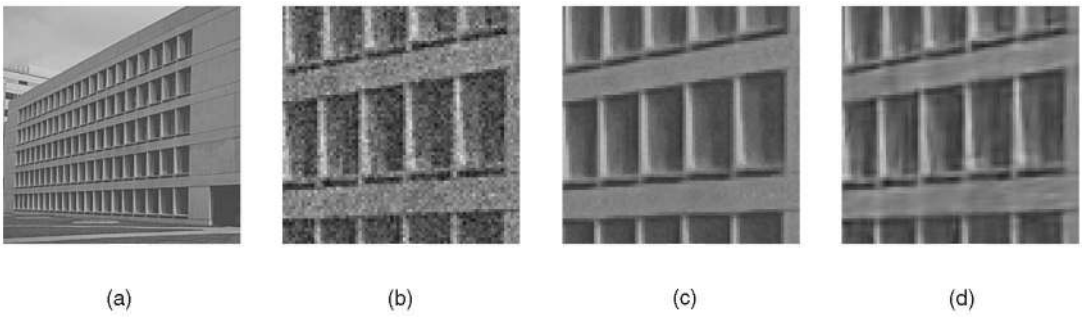

Fig. 12. (a) Building facade image. Zoomed insets of the (b) noisy image, (c) UINTA filtered image, and (d) wavelet denoised image [31].

or noise. The computation times for the implementation are impractical for most applications. This is because the algorithmic complexity $O(|T||A_i|E^D)$, where D is the image dimension and E is the extent of the neighborhood along a dimension, is exponential in E. Improving the computational scheme is an important area of future work.

An intrinsic limitation of UINTA is that its performance degrades for images not having truly stationary statistics. Indeed, comparisons against wavelet-based denoisers shed light on the kinds of images where UINTA does particularly well. One of the interesting empirical outcomes of this paper is that the method degrades elegantly as these conditions are relaxed.

The implications of the empirical results in this paper are significant. They show that it is possible to construct nonparametric density estimations in the very high-dimensional spaces of image neighborhoods. These results also suggest that the statistical structure in these spaces capture important geometric properties of images. The UINTA formulation also generalizes in several different ways. All of the mathematics, statistics, and engineering in this paper is appropriate for any kind of densely sampled data,

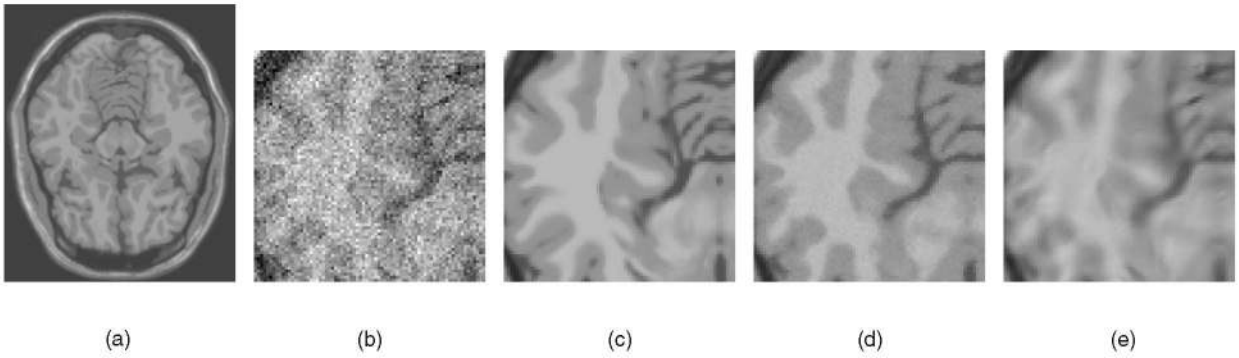

Fig. 13. (a) Image used for learning neighborhood statistics. Zoomed insets of the (b) noisy image, (c) original image, (d) UINTA filtered image, and (e) wavelet denoised image [31].

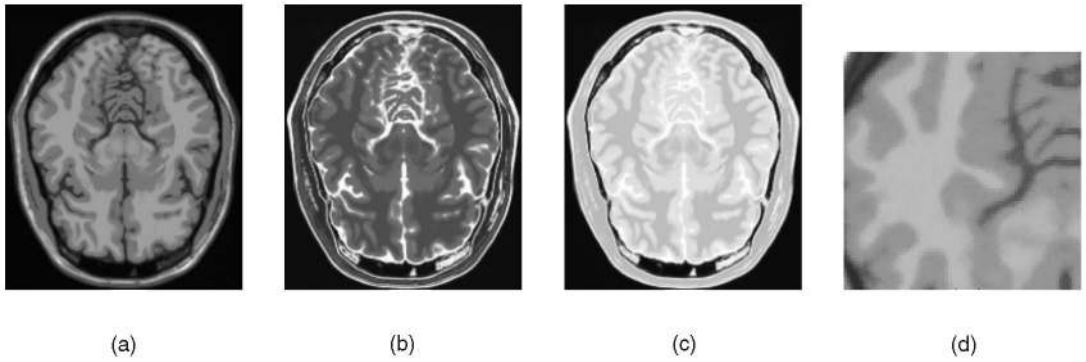

Fig. 14. (a), (b), and (c) Multimodal MR images comprised T1, T2, and PD scans. (d) Zoomed inset of the UINTA restored image.

including higher-dimensional image domains and vector-valued data. The challenge is the increase in computation time, which is already quite significant. The same scheme could easily apply to other image representations, such as image pyramids, wavelets, or local geometric features.

ACKNOWLEDGMENTS

This work was supported by US National Science Foundation (NSF) EIA0313268 and NSF CAREER CCR0092065 grants.

REFERENCES

[1] NLM Insight Segmentation and Registration Toolkit, http://www.itk.org, 2005.
[2] D. Adalsteinsson and J. Sethian, "A Fast Level Set Method for Propagating Interfaces," *J. Computer Physics,* vol. 118, no. 2, pp. 269-277, 1995.
[3] S.P. Awate and R.T. Whitaker, "Higher-Order Image Statistics for Unsupervised, Information-Theoretic, Adaptive, Image Filtering," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 44-51, 2005.
[4] S.P. Awate and R.T. Whitaker, "Nonparametric Neighborhood Statistics for MRI Denoising," *Proc. Int'l Conf. Information Processing in Medical Imaging,* 2005.
[5] D. Barash and D. Comaniciu, "A Common Framework for Nonlinear Diffusion, Adaptive Smoothing, Bilateral Filtering and Mean Shift," *Image Vision Computing,* vol. 22, no. 1, pp. 73-81, 2004.
[6] T. Chan, J. Shen, and L. Vese, "Variational PDE Models in Image Processing," *Notice of Am. Math. Soc.,* vol. 50, pp. 14-26, Jan. 2003.
[7] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," *IEEE Trans. Pattern Analysis Machine Intelligence,* vol. 17, no. 8, pp. 790-799, Aug. 1995.
[8] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, and A.C. Evans, "Design and Construction of a Realistic Digital Brain Phantom," *IEEE Trans. Medical Imaging,* vol. 17, no. 3, pp. 463-468, 1998.
[9] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 5, pp. 603-619, May 2002.
[10] T.M. Cover and J.A. Thomas, *Elements of Information Theory.* Wiley, 1991.
[11] V. de Silva and G. Carlsson, "Topological Estimation Using Witness Complexes," *Proc. Symp. Point-Based Graphics,* 2004.
[12] J. DeBonet and P. Viola, "A Nonparametric Multiscale Statistical Model for Natural Images," *Proc. Conf. Advances in Neural Infomation Processing Systems,* pp. 773-779, 1998.
[13] E. Dougherty, *Random Processes for Image and Signal Processing.* Wiley, 1998.
[14] R. Duda, P. Hart, and D. Stork, *Pattern Classification.* Wiley, 2001.
[15] A. Efros and T. Leung, "Texture Synthesis by Nonparametric Sampling," *Proc. Int'l Conf. Computer Vision,* p. 1033, 1999.
[16] M. Everett and J. Mancuso, *Reptiles.* Stanford Univ., 2001.
[17] K. Fukunaga and L. Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," *IEEE Trans. Infomation Theory,* vol. 21, no. 1, pp. 32-40, 1975.
[18] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, pp. 721-741, 1984.
[19] *Unsupervised Adaptive Filtering,* S. Haykin, ed. Wiley, 2000.
[20] J. Huang and D. Mumford, "Statistics of Natural Images and Models," *Proc. Int'l Conf. Computer Vision,* pp. 541-547, 1999.
[21] A. Hyvarinen, P. Hoyer, and E. Oja, "Sparse Code Shrinkage: Denoising of Nongaussian Data by Maximum Likelihood Estimation," *Neural Computation,* vol. 11, no. 7, pp. 1739-1768, 1999.
[22] A. Lee, K. Pedersen, and D. Mumford, "The Nonlinear Statistics of High-Contrast Patches in Natural Images," *Int'l J. Computer Vision,* vol. 54, nos. 1-3, pp. 83-103, 2003.
[23] J. Miller and C. Stewart, "Muse: Robust Surface Fitting Using Unbiased Scale Estimates," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 300-306, June 1996.
[24] D. Mumford and J. Shah, "Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems," *Computation of Pure and Applied Math.,* vol. 42, pp. 577-685, 1989.
[25] K.N. Nordstrom, "Biased Anisotropic Diffusion: A Unified Regularization and Diffusion Approach to Edge Detection," *Image Vision Computing,* vol. 8, no. 4, pp. 318-327, 1990.

[26] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces.* Springer, 2003.
[27] P. Perona and J. Malik, "Scale-Space and Edge Detection Using Anisotropic Diffusion," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 7, pp. 629-639, July 1990.
[28] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A Joint Inter and Intrascale Statistical Model for Bayesian Wavelet Based Image Denoising," *IEEE Trans. Image Processing,* vol. 11, pp. 545-557, 2002.
[29] K. Popat and R. Picard, "Cluster Based Probability Model and Its Application to Image and Texture Processing," *IEEE Trans. Image Processing,* vol. 6, no. 2, pp. 268-284, 1997.
[30] J. Portilla, "Full Blind Denoising through Noise Covariance Estimation Using Gaussian Scale Mixtures in the Wavelet Domain," *Proc. IEEE Int'l Conf. Image Processing,* pp. 1217-1220, Oct. 2004.
[31] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image Denoising Using Scale Mixtures of Gaussians in the Wavelet Domain," *IEEE Trans. Image Processing,* vol. 12, no. 11, pp. 1338-1351, 2003.
[32] *Geometry-Driven Diffusion in Computer Vision,* B.M. Romeny, ed. Kluwer Academic, 1994.
[33] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear Total Variation Based Noise Removal Algorithms," *Physica D,* vol. 60, pp. 259-268, 1992.
[34] D.W. Scott, *Multivariate Density Estimation.* Wiley, 1992.
[35] L. Sendur and I. Selesnick, "Bivariate Shrinkage Functions for Wavelet-Based Denoising Exploiting Interscale Dependency," *IEEE Trans. Signal Processing,* vol. 50, pp. 2744-2756, 2002.
[36] J. Sethian, *Level Set Methods and Fast Marching Methods.* Cambridge Univ. Press, 1999.
[37] C. Shannon, "A Mathematical Theory of Communication," *Bell System Technical J.,* vol. 27, pp. 379-423, July 1948.
[38] B. Silverman, *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 1986.
[39] W. Snyder, Y. Han, G. Bilbro, R. Whitaker, and S. Pizer, "Image Relaxation: Restoration and Feature Extraction," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 6, pp. 620-624, June 1995.
[40] J. Starck, E. Candes, and D. Donoho, "The Curvelet Transform for Image Denoising," *IEEE Trans. Image Processing,* vol. 11, no. 6, 2000.
[41] T. Tasdizen, R. Whitaker, P. Burchard, and S. Osher, "Geometric Surface Processing via Normal Maps," *ACM Trans. Graphics,* 2003.
[42] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," *Proc. Int'l Conf. Computer Vision,* p. 839, 1998.
[43] L.A. Vese and S.J. Osher, "Modeling Textures with Total Variation Minimization and Oscillating Patterns in Image Processing," *J. Scientific Computing,* vol. 19, pp. 553-572, 2003.
[44] P. Viola and W. Wells, "Alignment by Maximization of Mutual Information," *Proc. Int'l Conf. Computer Vision,* pp. 16-23, 1995.
[45] L. Wei and M. Levoy, "Order-Independent Texture Synthesis," Technical Report TR-2002-01, Computer Science Dept., Stanford Univ., 2002.
[46] J. Weickert, *Anisotropic Diffusion in Image Processing.* Teubner-Verlag, 1998.
[47] J. Weickert, "Coherence-Enhancing Diffusion Filtering," *Int'l J. Computer Vision,* vol. 31, pp. 111-127, 1999.
[48] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdu, and M. Weinberger, "Universal Discrete Denoising: Known Channel," HP Labs Technical Report HPL-2003-29, 2003.

Suyash P. Awate received the BE degree in computer engineering from the University of Mumbai, India, in 2001, earning a distinction and a university rank. He was the recipient of the Dhirubhai Ambani Undergraduate Merit Scholarship (1997-2001) during his undergraduate curriculum. He joined the University of Utah in 2001 to pursue doctoral studies and is currently a PhD candidate in the School of Computing and a member of the Scientific Computing and Imaging Institute. His interests include image processing, medical image analysis, computer vision, and computer graphics. He is a student member of the IEEE.

Ross T. Whitaker received the BS degree in electrical engineering and computer science from Princeton University in 1986, earning summa cum laude. From 1986 to 1988, he worked for the Boston Consulting Group, entering the University of North Carolina (UNC) at Chapel Hill in 1989. At UNC, he received the Alumni Scholarship Award and completed the PhD degree in computer science in 1994. From 1994-1996, he worked at the European Computer-Industry Research Centre in Munich, Germany, as a research scientist in the User Interaction and Visualization Group. From 1996-2000, he was an assistant professor in the Department of Electrical Engineering at the University of Tennessee. Since then, he has been at the University of Utah, where he is an associate professor in the School of Computing and a faculty member of the Scientific Computing and Imaging Institute. He is a member of the IEEE Computer Society.