Higher-Order Image Statistics for Unsupervised, Information-Theoretic, Adaptive, Image Filtering

Suyash P. Awate and Ross T. Whitaker

UUCS-05-005

School of Computing
University of Utah
Salt Lake City, UT 84112, USA

Abstract

The restoration of images 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. Therefore, these methods typically 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 the joint entropy between them. Thus UINTA automatically discovers the statistical properties of the signal and can thereby restore a wide spectrum of images and applications. This paper describes the formulation required to minimize the joint entropy measure, presents several important practical considerations in estimating image-region statistics, and then presents results on both real and synthetic data.

Higher-Order Image Statistics for Unsupervised, Information-Theoretic, Adaptive, Image Filtering

Suyash P. Awate Ross T. Whitaker
School of Computing, University of Utah, Salt Lake City, Utah 84112
{suyash,whitaker}@cs.utah.edu

Abstract

The restoration of images 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. Therefore, these methods typically 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 the joint entropy between them. Thus UINTA automatically discovers the statistical properties of the signal and can thereby restore a wide spectrum of images and applications. This paper describes the formulation required to minimize the joint entropy measure, presents several important practical considerations in estimating image-region statistics, and then presents results on both real and synthetic data.

1. Introduction

Restoring 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 examples in this paper. However, such degradations can also include correlated noise, spatially varying blurring ("smudging"), areas with loss of contrast, etc.

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 typically 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 algorithms/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. UINTA restores pixels by comparing them to 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 hence the filtering, obtained as a variation of the functional, is nonlinear. UINTA operates without a priori knowledge of the geometric or statistical structure of the signal or degradation, 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 hence *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 all its important free parameters automatically via 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.

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 entropy measure. Entropy optimization entails the estimation of probability densities for the associated random variables. Section 4 describes a nonparametric multivariate density estimation technique and gives some reasons behind UINTA's success in spite of a large number of dimensions involved. Section 5 formulates a gradient-

[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.

descent scheme to optimize the joint entropy measure and discusses several important practical challenges pertaining to statistical estimation and its application to image neighborhoods. Section 6 gives numerous experimental results on real and synthetic images and analyzes UINTA's behavior on the same. Section 7 summarizes the contributions of the paper and presents ideas for further exploration.

2. Related Work

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 that lead to algorithms based on partial differential equations (PDEs), or statistical methods that lead to nonlinear estimation problems. They typically involve high computational costs, entail the tuning of extra free parameters, and enforce strong geometric or statistical assumptions on images.

PDE-based image processing methods became widespread after Perona and Malik's work on *anisotropic diffusion* [16], which smoothed high-frequency noise while sharpening edges. The anisotropic diffusion equation is also the first variation of an image energy that penalizes image gradients with an allowance for outliers [14], and therefore seeks piecewise constant solutions. 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 [13, 20, 19, 26, 27] with a variety of algorithms based on level sets [22, 15].

An alternative to variational models is to construct nonlinear transforms in the frequency domain. The wavelet literature addresses image denoising extensively. The current state-of-the-art methods [18] treat the wavelet coefficients as random variables and model their a priori marginal/joint probability density functions (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 specific statistical models imposed on the coefficients.

Among the statistical approaches to nonlinear filtering, an important class of methods rely on stochastic image models described by Markov random fields (MRFs) [9]. They exploit the equivalence between MRFs and Gibbs distributions to model image statistics with Gibbs distributions. UINTA also exploits the Markov property of the images, but rather than imposing a particular model on the image, it *estimates* the relevant conditional PDFs from the input data and updates pixel intensities to decrease the randomness of these conditional PDFs.

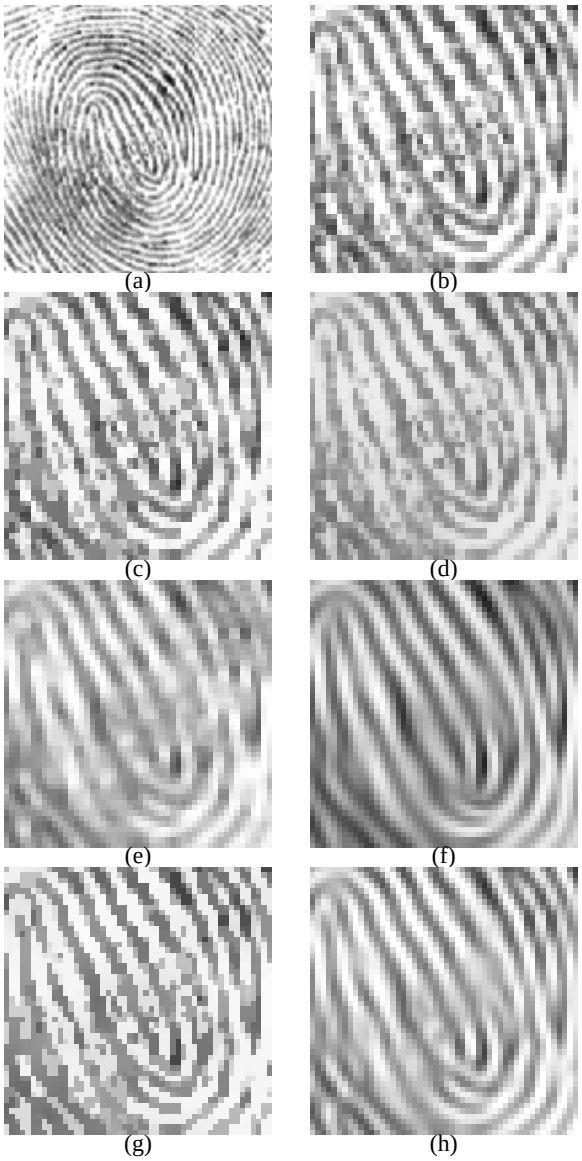

Figure 1. (a) Degraded fingerprint image and (b) a zoomed inset. Zoomed insets of the fingerprint image (grayscale range:0-100) processed with (c) anisotropic diffusion (K=0.45 grayscale values, 99 iterations) (d) bilateral filtering (σ_{domain}=3 pixels, σ_{range}=15 grayscale values) (e) curvature flow (time step=0.2, 5 iterations) (f) coherence enhancing diffusion (σ=0.1 pixels, ρ=2 pixels, α=0.0001, C=0.0001, 15 iterations) (g) unrestricted mean shift filtering [1] (σ_{domain}=2 pixels, σ_{range}=5 grayscale values, 5 iterations) (h) wavelet denoising [18] (σ_{noise}=14 grayscale values).

The literature shows several statistically-based image processing algorithms that do rely on information theory, such as the *mean-shift* procedure [8, 2, 3], which can be said to be a *mode seeking* process on the PDF associated with image intensities. It moves the samples uphill on that PDF to a steady state with all samples converged to the nearest

mode. However, the mean-shift procedure operates only on image intensities (be they scalar or vector valued) and does not account for neighborhood structure in images. This paper shows the mathematical relationship between the mean-shift procedure and entropy reduction and thereby formulates UINTA as a generalization of the mean-shift procedure, which incorporates neighborhoods to reduce the entropy of the associated conditional PDFs.

Figure 1 shows the results of filtering on a degraded fingerprint image using some of the prevalent nonlinear techniques, demonstrating their typical characteristics. We see that the piece-wise smooth image models associated with anisotropic diffusion, bilateral filtering [24], and curvature flow [22] (Figure 1(c)-(e)) are clearly inappropriate for this image. Coherence enhancing diffusion [27] (which is as well suited to this image as virtually any other) does not succeed in retaining or enhancing the light-dark contrast boundaries, and yet it forces elongated structures to grow or connect (Figure 1(f)). An unrestricted mean-shift filtering [1] (Figure 1(g)) on image intensities yields a thresholded image retaining most of the noise. Wavelet denoising (Figure 1(h)), which imposes statistical models on wavelet coefficients and an additive white Gaussian noise model, is unable to get rid of the smudges and excessively smoothes the image.

Popat *et al.*[17] were among the first to use nonparametric Markov sampling in images. They attempt to capture the higher-order nonlinear image statistics via cluster-based nonparametric density estimation and use them for image restoration, image compression, and texture classification. However, their method involves training unlike UINTA which attempts to learn these statistics from the degraded data itself.

Researchers analyzing the statistics of natural image neighborhoods [11, 12, 5] found the multi-dimensional intensity data derived from image neighborhoods to be concentrated in clusters or low-dimensional manifolds exhibiting nontrivial topologies. UINTA also relies on the hypothesis that natural images exhibit some regularity in neighborhood structure, but it discovers this regularity for each image individually in a nonparametric manner.

Weissman *et al.*, [28], propose the DUDE algorithm that addresses the problem of denoising data sequences generated by a discrete source and received over a discrete, memoryless channel. DUDE assigns image values based on the similarity of neighborhoods gathered from image statistics, which resembles the construction of conditional probabilities in the proposed method. However, the DUDE approach is limited to discrete-valued signals whereas the proposed method addresses continuous-valued signals, such as those associated with grayscale images. While the DUDE algorithm is demonstrably effective for removing *replacement noise*, it is less effective in case of additive noise.

3. Joint Entropy Based Image Filtering

This section describes the mathematical formulation of UINTA. It begins with an overview of the random-field image model and the associated notation, concluding with a high-level algorithmic description of UINTA.

3.1. Random Field Image Model

A random field/process [6] 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. We use a shorthand to denote random variables $X(\Omega;t)$ by $X(t)$. We denote a specific realization $X(\omega;t)$ (the intensity at pixel t), as 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 and pixels in N_t are called neighbors of t. We define a random vector $Y(t)=\{X(t)\}_{t\in N_t}$, denoting its realization by $y(t)$, corresponding to the set of intensities at the neighbors of pixel t. We also define a random vector $Z(t)=(X(t),Y(t))$ corresponding to 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). We denote the original (not degraded) image by $X(\omega,T)$ and its associated set of neighborhood intensities by $Y(\omega,T)$ with regions $Z(\omega,T)$. 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 short hand 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 filtering 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. 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 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 also a reweighted gradient descent on the joint entropy $h(\tilde{X},\tilde{Y})$.

This choice of entropy as a measure of goodness follows from several observations. First, the addition of two independent random variables (e.g. a signal and additive noise) increases the entropy [4]. 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, natural images tend to have very low entropy relative to their degraded versions. Therefore, a gradient descent on entropy first affects the randomness in degradations substantially more than the signal. Furthermore, the entropy reduction is limited by an information-based stopping criterion, as described in Section 3.4.

3.3. High-Level Algorithmic Structure

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 region $\hat{z}^m$ in the current image I^m, compute $\partial h(\hat{X}|\hat{Y} = \hat{y}^m)/\partial \hat{x}^m$.
3. Construct a new image I^{m+1}, using finite forward differences on the gradient descent, with intensities $\hat{x}^{m+1} = \hat{x}^m - \lambda \partial h/\partial \hat{x}^m$.
4. Check the stopping criterion (see Section 3.4). If not done, go to Step 2, otherwise I^{m+1} is the output.

3.4. Stopping Criteria

The choice of stopping criteria for this algorithm depend 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. For instance, we have found empirically that the entropy of the noise diminishes faster than that of the signal, and therefore, an effective strategy is to stop when the relative rate of change of the entropy, from one iteration to the next, falls below some threshold. In the case of a known level of additive noise, UINTA iterates until the root-mean-square difference (residual) between input (noisy) and the processed image equals the noise level. We have found empirically that this method is quite effective and we have used this approach in all of the results (for which noise levels are known) in this paper.

4. Nonparametric Density Estimation

Entropy optimization entails the estimation of the PDFs associated with neighborhoods. High-dimensional spaces are notoriously challenging for data analysis (regarded as *the curse of dimensionality* [21, 7]) because they are so sparsely populated. Despite theoretical arguments suggesting that density estimation beyond a few dimensions is impractical, the empirical evidence from literature is more optimistic [21, 17]. The results in this paper confirm that observation. One advantage for UINTA is that the random vector $\tilde{Z} \equiv (\tilde{X}, \tilde{Y})$ comprises random variables with identical PDFs, via stationarity, which lends itself to more accurate density estimates [21, 23]. Further, UINTA relies on the neighborhoods in natural images having a lower-dimensional topology in the multi-dimensional feature space [12, 5].

For image regions comprising n pixels, UINTA employs Parzen-windowing [7] with an n-dimensional Gaussian, $G(\tilde{z}, \Psi_n)$, as the kernel where Ψ_n is the $n \times n$ covariance matrix. Having no a priori information on the structure of the PDFs, we choose an isotropic Gaussian of standard deviation σ, i.e. Ψ_n=$\sigma^2 I$, where I is the $n \times n$ identity matrix. Using optimal values of the Parzen-window parameters is critical to UINTA's success, and Section 5.2 gives strategies for the same.

Entropy is the expectation of negative log-probability, and therefore we can approximate it with the sample mean [25]. Thus, for a stationary ergodic process

$$h(\tilde{Z}) \approx \frac{-1}{|T|} \sum_{t_i \in T} \log \left(\frac{1}{|A_i|} \sum_{t_j \in A_i} G_n(\tilde{z}_i - \tilde{z}_j, \Psi_n) \right) \quad (1)$$

where z_j is shorthand for $z(t_j)$. The samples in set A_i are, typically, a small random fraction of those in T. This significantly reduces the computational cost for the entropy estimation, from $O(|T|^2)$ to $O(|A_i||T|)$, and produces a stochastic approximation for entropy. The latter leads to a stochastic-gradient algorithm [10] that effectively overcomes the effects of spurious local maxima introduced in the Parzen-window density estimate [25].

5. Entropy Minimization of 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 via A_i. The gradient descent is

$$\frac{\partial \tilde{x}_i}{\partial \tau} = -\frac{\partial h(\tilde{X}|\tilde{Y} = \tilde{y}_i)}{\partial \tilde{x}_i} \approx \frac{1}{|T|} \frac{\partial \log P(\tilde{z}_i)}{\partial \tilde{x}_i} = \quad (2)$$

$$\frac{-1}{|T|} \frac{\partial \tilde{z}_i}{\partial \tilde{x}_i} \sum_{t_j \in A_i} \frac{G_n(\tilde{z}_i - \tilde{z}_j, \Psi_n)}{\sum_{t_k \in A_i} G_n(\tilde{z}_i - \tilde{z}_k, \Psi_n)} \Psi_n^{-1}(\tilde{z}_i - \tilde{z}_j) \quad (3)$$

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

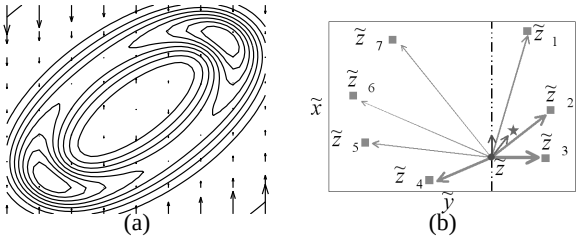

Figure 2. (a) An example of a 2D PDF contour plot with the forces (vertical arrows) that reduce entropy of the conditional PDFs $P(\tilde{X}|\tilde{Y}=\tilde{y})$, (see (2)). (b) Attractive forces (arrows: width $\equiv$ force magnitude) act on a sample ($\tilde{z}$:circle) towards other samples ($\tilde{z}_j$:squares) in the set A_i (see (3)). The resultant force acts towards the weighted mean (star), and the sample $\tilde{z}$ moves based on its projection (vertical arrow).

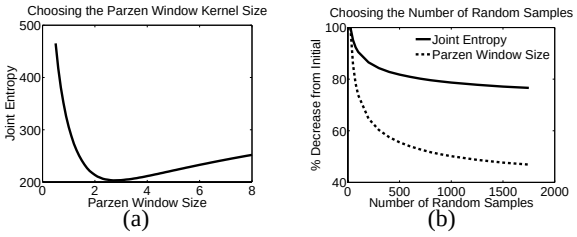

Figure 3. (a) $h(X,Y)$ versus σ (for the *Lena* image in Figure 6(a), $|A_i| = 1000$). (b) $h(X,Y)$ and σ (for the *Lena* image), are almost unaffected for $|A_i| > 1000$. To give smoother curves, each measurement, for a particular $|A_i|$, was averaged over 3 different random sets A_i.

the center pixel intensity $\tilde{x}_i$, and τ is a dummy variable for time. Figure 2 elucidates this process.

If we choose a $\sigma > 0$ (and finite), the entropy for a finite set of samples is always bounded. Because we perform a (projected) gradient descent on a bounded energy function, the process converges (for sufficiently small time steps). Indeed, analysis of simple examples shows the existence of nontrivial steady states, e.g. an image which is a discrete sampling of a linear function. Empirical evidence shows that the filtering algorithm does sometimes converge to interesting results. However, for many applications, convergence to a fixed point is not the goal; as with many other iterative filtering strategies, several iterations of the gradient descent are sufficient for acceptable restoration.

5.1. Relationship to the Mean-Shift Procedure

A gradient descent on the entropy $h(\tilde{X})$ of the grayscale pixel intensities using finite forward differences, $\tilde{x}_i^{m+1} = \tilde{x}_i^m - \lambda \partial h(\tilde{X})/\partial \tilde{x}_i^m$, with a time step $\lambda = |T|\sigma^2$ and $A_i = T, \forall i$ gives exactly the mean-shift update proposed by Fukunaga [8]. The UINTA updates have the same form, except that the weights are influenced not only by the distances/similarities between intensities $\tilde{x}_i$ but also by the distances/similarities between the neighborhoods $\tilde{y}_i$. That is, pixels in the image with similar neighborhoods have a relatively larger impact on the weighted mean (see (3)) that drives the updates of the center pixels. Note that in UINTA, as well as in [8], the PDFs on which the samples climb get updated after every iteration, unlike the approach in [3].

5.2. Implementation Issues

The UINTA algorithm, as presented in previous sections, presents a number of significant engineering questions which are crucial for its effectiveness. This section discusses some of these issues and the proposed solutions.

The first issue is the selection of the scale or size of the Parzen window: the Gaussian kernel standard deviation σ. The Parzen-window density estimate, using finitely many samples, shows a great deal of sensitivity for different values of σ [7]. Thus, the particular choice of σ, and thereby Ψ_n, is a crucial factor that determines the behavior of the entire process of entropy optimization. Furthermore, this choice is related to the sample size $|A_i|$ in the stochastic approximation. For a particular choice of $|A_i|$, we propose to use the σ that minimizes the joint entropy, which we will call the *optimal* scale for a data set. We determine this automatically at each iteration in UINTA processing (Figure 3(a)). Our experiments show that for sufficiently large $|A_i|$ the entropy estimates and the optimal scale are virtually constant, and thus $|A_i|$ can also be generated directly from the input data (Figure 3(b)).

In practice, image statistics are not stationary for most images and are more accurately modeled as piecewise stationary ergodic. Thus, while processing pixel t, the set A_i should comprise pixels that are nearby t. To achieve this, we choose a unique set of pixels, for each pixel t, at random from a Gaussian distribution on the image coordinates centered at t with standard deviation $\beta = 40$. This strategy gives consistently better results than uniform sampling, and we have found that it performs well for any choice of β that encompasses more than several hundred pixels.

Another issue is the shape of the image neighborhoods. Square neighborhoods produce anisotropic artifacts. To obtain isotropic filtering results we propose a feature space metric that controls the influence of each neighborhood pixel so that the resulting *mask* is more isotropic. In this way, directions in the feature space corresponding to corners of neighborhoods collapse so that they do not influence the filtering. A similar strategy enables us to handle image boundaries without distorting the image statistics. That is, pixels at image boundaries rely on the statistics in lower-dimensional subspaces corresponding to the set of neighborhood pixels lying within the image.

The algorithmic complexity of UINTA is significant:

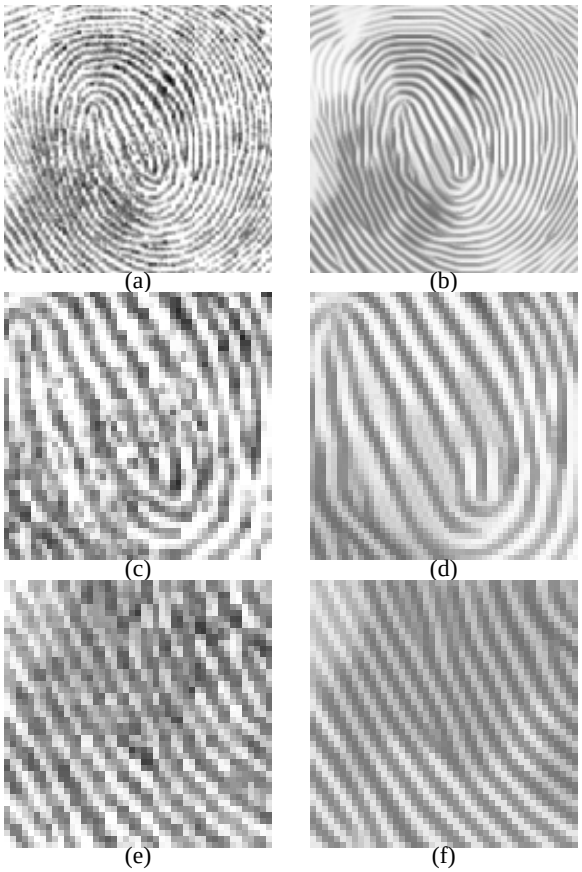

Figure 4. (a) Degraded fingerprint. (b) UINTA filtered image. (c),(e) and (d),(f) show zoomed insets of (a) and (b), respectively.

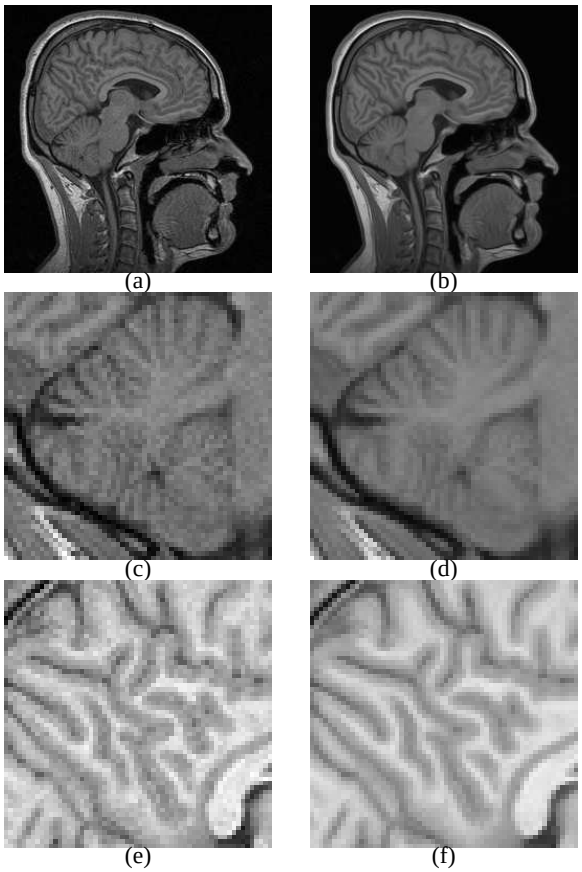

Figure 5. (a) Noisy MRI head. (b) UINTA filtered image. (c),(e) and (d),(f) show zoomed insets of (a) and (b), respectively.

$O(|A_i||T|E^D)$ where D is the image dimension and E is the extent of the neighborhood along a dimension. This is exponential in E, and the current results are limited to 2D images. The literature suggests some potential improvements (e.g. [29]). However, the purpose of this paper is to introduce the theory and methodology—algorithmic improvements are the subject of future work.

6. Results

This Section gives the results of UINTA filtering on real and synthetic images and analyzes the behavior of UINTA on the same. The noise in the synthetic-image examples is additive, zero-mean, independent, and Gaussian. Because of interactions of neighborhoods from one iteration to the next, the time step $\lambda = |T|\sigma^2$ can lead to oscillations in the results. We have found that a time step of $\lambda = |T|\sigma^2/\sqrt{|N_t|}$, where $|N_t|$ denotes neighborhood size, alleviates this effect. UINTA takes roughly 2 minutes per iteration, for a 256^2 image, with standard hardware.

The fingerprint image in Figure 4 shows an example where the degradation involves smudges (blurring) and is clearly not additive noise. UINTA enhances the contrast of the light and dark lines without significant shrinkage. The results are noticeably better than any of those obtained using other methods shown in Figure 1. UINTA performs a kind of multidimensional classification of image neighborhoods—therefore some features in the top-left are lost because they resemble the background more than the ridges. As a stopping criterion, UINTA uses the relative change in entropy as described in Section 3.4.

Figure 5 shows the results of processing an MRI image of a human head. This shows UINTA's ability to adapt to a variety of grayscale features in real images approximated by piecewise stationary models. Figure 6 gives an example of restoring the *Lena* image corrupted with independent and identically distributed (i.i.d.) additive Gaussian noise. The wavelet denoising technique (Figure 6(d)) does yield a lower root mean squared error (RMSE). However, it also introduces ringing-like artifacts in smooth facial regions.

Figure 7 shows UINTA filtering on a noisy image of hand-drawn curves. 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 come

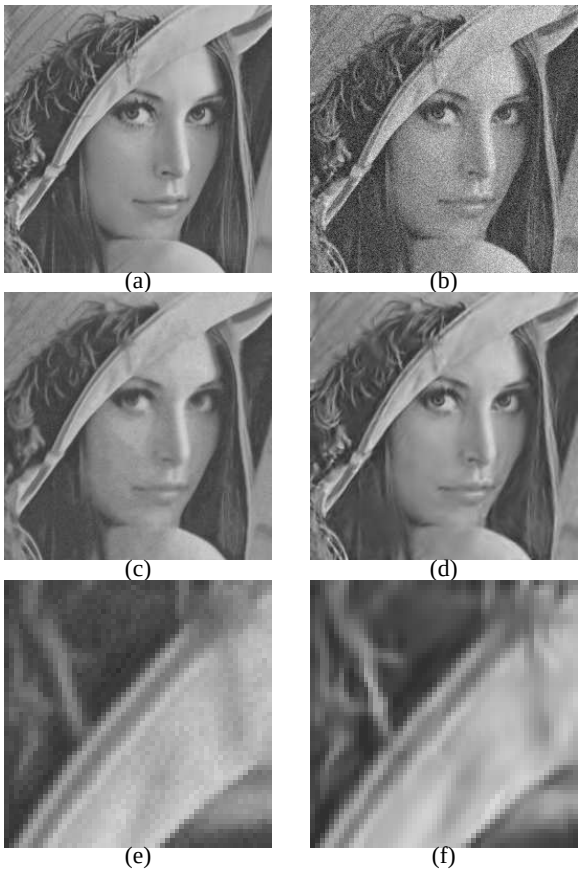

Figure 6. (a) *Lena* image (intensity range 0:100). (b) Noisy image (RMSE 10). (c) UINTA filtered image (RMSE 4.6). (d) Wavelet denoised [18] image (RMSE 3.6). (e) and (f) show zoomed insets of (c) and (d) respectively.

too close, exhibit a very sharp bend, or when the noise introduces ambiguous gaps. The wavelet processed image depicts significant artifacts around the edges.

Figure 8 shows a real image of a building facade that exhibits a certain degree of redundancy. UINTA is able to exploit this property to perform comparably (RMSE) with the wavelet denoiser and with fewer visual artifacts.

UINTA is not designed with a particular noise model in mind. We have observed empirically that it performs best with i.i.d. additive noise. However, experiments with correlated noise on natural images show that UINTA performs better, qualitatively and quantitatively, as compared to wavelet denoisers. How to incorporate specific noise models within the UINTA framework is an area of future research.

7. Conclusions and Discussion

UINTA is a novel, unsupervised, information-theoretic, adaptive filter that improves the predictability of pixel intensities from the intensities in the 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 imposing strong models. Because it is nonparametric, it can adapt to the statistics of the input image, and thereby applies quite readily to new

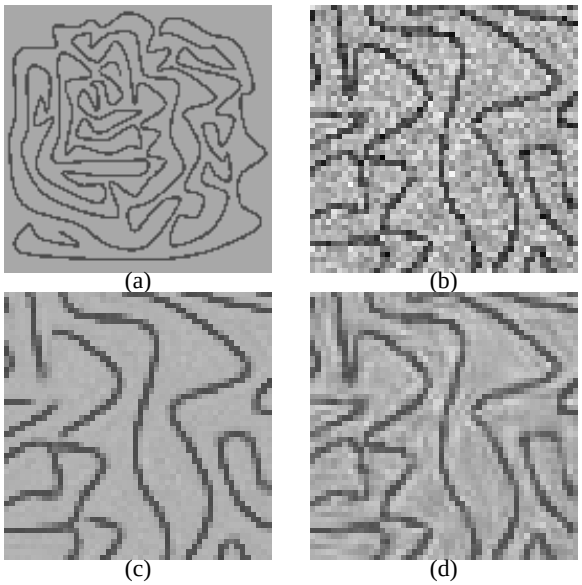

Figure 7. (a) Hand drawn curves (intensity range $0 : 100$). Zoomed insets of the (b) noisy image (RMSE 25), (c) UINTA filtered image (RMSE 15.4), and (d) wavelet denoised [18] image (RMSE 16).

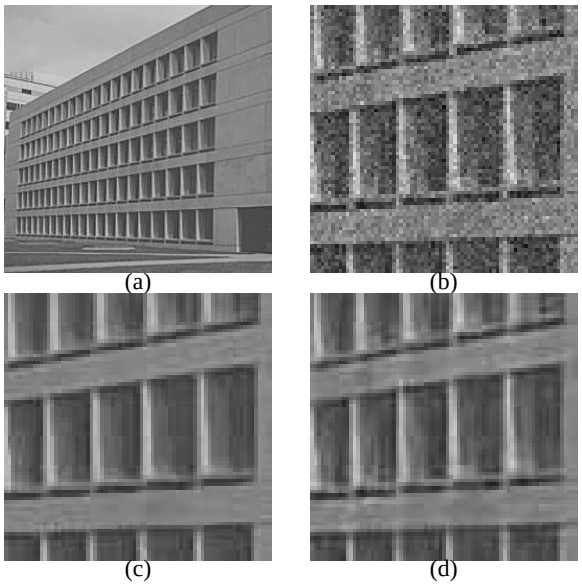

Figure 8. (a) Building facade (intensity range $0 : 100$). Zoomed insets of the (b) noisy image (RMSE 10), (c) UINTA filtered image (RMSE 4.5), and (d) wavelet denoised [18] image (RMSE 4.4).

applications with very 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 generalizes the mean-shift classification algorithm [2] by conditioning the distribution based on the pixel neighborhood. Results show that the statistics of image neighborhoods are sufficiently regular for reliable image denoising.

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 and/or noise. The computation times for the implementation are impractical for most applications, and improving the computational scheme is an important area of future work.

The empirical results in this paper have significant implications. They show that it is possible to construct non-parametric density estimations in the very high dimensional spaces of image neighborhoods. The UINTA formulation also generalizes in several different ways. All of the mathematics, statistics, and engineering in this paper is appropriate for higher-dimensional image domains and vector-valued data. 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 the NSF grant EIA0313268 and the NSF CAREER grant CCR0092065.

References

[1] D. Barash and D. Comaniciu. A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift. *Image Vision Comput.*, 22(1):73–81, 2004.

[2] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(8):790–799, 1995.

[3] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, 2002.

[4] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley, 1991.

[5] V. de Silva and G. Carlsson. Topological estimation using witness complexes. *Symp. on Point-Based Graphics*, 2004.

[6] E. Dougherty. *Random Processes for Image and Signal Processing*. Wiley, 1998.

[7] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley, 2001.

[8] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Info. Theory*, 21(1):32–40, 1975.

[9] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6:721–741, 1984.

[10] S. Haykin, editor. *Unsupervised Adaptive Filtering*. Wiley, 2000.

[11] J. Huang and D. Mumford. Statistics of natural images and models. In *Int. Conf. Comp. Vision*, pages 541–547, 1999.

[12] A. Lee, K. Pedersen, and D. Mumford. The nonlinear statistics of high-contrast patches in natural images. *Int. J. Comput. Vision*, 54(1-3):83–103, 2003.

[13] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Com. of Pure and Applied Math.*, 42:577–685, 1989.

[14] K. Nordstrom. Biased anisotropic diffusion: a unified regularization and diffusion approach to edge detection. *Image Vision Comput.*, 8(4):318–327, 1990.

[15] S. Osher and R. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2003.

[16] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, July 1990.

[17] K. Popat and R. Picard. Cluster based probability model and its application to image and texture processing. *IEEE Trans. Image Processing*, 6(2):268–284, 1997.

[18] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. Imag. Proc.*, 12(11):1338–1351, 2003.

[19] B. Romeny, editor. *Geometry-Driven Diffusion in Computer Vision*. Kluwer Academic Publishers, 1994.

[20] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[21] D. Scott. *Multivariate Density Estimation*. Wiley, 1992.

[22] J. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge Univ. Press, 1999.

[23] B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.

[24] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. Int. Conf. Comp. Vision*, page 839. IEEE Computer Society, 1998.

[25] P. Viola and W. Wells. Alignment by maximization of mutual information. In *Int. Conf. Comp. Vision*, pages 16–23, 1995.

[26] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner-Verlag, 1998.

[27] J. Weickert. Coherence-enhancing diffusion filtering. *Int. J. Comp. Vis.*, 31:111–127, 1999.

[28] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdu, and M. Weinberger. Universal discrete denoising: Known channel. *HP Labs Tech. Report HPL-2003-29*, 2003.

[29] C. Yang, R. Duraiswami, N. Gumerov, and L. Davis. Improved fast gauss transform and efficient kernel density estimation. In *Proc. Int. Conf. Comp. Vision*, pages 464–471, 2003.