A Robust Method for Edge-Preserving Image Smoothing*

Gang Dong and Kannappan Palaniappan

Department of Biology
University of Massachusetts, Amherst, Massachusetts 01003 USA
Department of Computer Science
University of Missouri-Columbia, Columbia, Missouri 65211 USA

Abstract. Image smoothing is a critical preprocessing step in many image processing tasks. In this paper, a generalized edge-preserving smoothing model is derived from robust statistics theory, and its connections to anisotropic diffusion and bilateral filtering are established. The relationships allow us to derive an improved numerical scheme in the context of a robust estimation process for edge preserving smoothing. Experiments illustrate that the proposed smoothing algorithm is capable of effectively reducing the distracting effects of noise without sacrificing image edge structures. The robust edge-preserving smoothing method performs several dB better in terms of PSNR compared to anisotropic diffusion, bilateral filtering and the Bayes least squares Gaussian scale mixtures a wavelet-based methods for image enhancement.

1 Introduction

Non-linear smoothing techniques continue to be an active area of research in the field of image processing. The basic idea of image smoothing is to exchange the intensity value at each pixel through some linear or nonlinear function of the intensities of the pixel's local neighboring pixels, to generate a new pixel value that is more representative of the local image structure that improves on some image quality such as noise, edge sharpness, visual fidelity, etc. Traditionally, linear smoothing filters, in particular Gaussian filters [1] have been widely adopted because of their mathematical simplicity and numerous techniques to achieve computational efficiency. However, a major drawback of linear filters is that they tend to smooth across edge boundaries thereby destroying important edge structures.

In order to reduce noise within regions while preserving edges between regions, various alternative nonlinear approaches have been developed in the literature. Among them, anisotropic diffusion [2] and its variants [3]-[5] produce both visually impressive and quantitatively superior results. These partial differential equation (PDE) based methods are excellent in the creation of a piecewise constant image with preserved discontinuities from a noisy input image, but being iterative processes they are inherently slow in nature. Recently, the bilateral filter [7]-[9] has been proposed as a simple and non-iterative alternative approach, and is demonstrated to be capable of efficiently generating results similar to those obtained by the iterative PDE-based anisotropic methods.

* This work was partially supported by a U.S National Institute of Health NIBIB award R33 EB00573.

J. Blanc-Talon et al. (Eds.): ACIVS 2008, LNCS 5259, pp. 390–399, 2008.

In this paper, we propose a generalized edge-preserving smoothing model derived from robust statistical methods [10]-[14], by exploiting the fact that the basic common principle behind most edge-preserving smoothing algorithms is the criteria of statistical robustness. Our proposed solution to edge-preserving image smoothing can be viewed as a minimization of a robust estimation problem, and we further show that the anisotropic diffusion and the bilateral filter methods can be derived from the same robust modeling framework. This derivation is important since it paves the way towards a better understanding of the similarities and differences between these two approaches. Our new algorithm, which has excellent edge-preserving qualities, is based on taking advantage of the observation that nonlinear smoothing can be cast as a minimization procedure. In order to avoid local minima issues, which often arise in the context of nonlinear smoothing, a graduated nonconvexity procedure is introduced to calculate a nearly optimal solution. Comparative results are given, to demonstrate the favorable performance of the proposed method.

2 A Robust ESTIMATION Model for IMAGE Smoothing

The task of image smoothing is twofold: all the homogeneous regions that contain noise should be filtered or smoothed to reduce the noise induced artifacts, and the locations of the image intensity edges that define the shape of the representative two-dimensional structures are retained and enhanced in an accurate manner. In general, most methods rely explicitly or implicitly on a specific noise assumption (such as Gaussian, Rayleigh or Exponential noise), by formalizing knowledge about the statistical characteristics of the image sensors and acquisition system. It, however, often happens in practice that an assumed model holds only approximately in that the assumed model describes the majority of pixels, but some pixels may follow a different statistical behavior or no pattern at all. For example, the image intensity noise may be assumed to be normally distributed but the actual distribution may have heavy tails, or the local neighborhood may include significant image structure such as an edge boundary. Such atypical pixels are called outliers, or gross errors with respect to the assumed model.

The methods of robust statistics [10]-[14] provide a reliable way of detecting outliers for the estimation problems in which underlying model or noise assumptions are inexact. Here *robustness* means that the estimation procedure should be insensitive to departures from the underlying assumptions due to outliers in the data. The estimation problem of finding the best fitting smoothly approximating image from a noisy input image can be posed under the framework of robust statistics. Given a robust function $\rho(x)$, we wish to find an image g as the estimate of the input image f by satisfying the following functional criterion,

$$\min_{g} \boldsymbol{E}(g) \tag{1}$$

where

$$\boldsymbol{E}(g) = \sum_{s \in \Omega} \sum_{p \in \eta_s} w(p-s)\rho\left(f_p - g_s, \sigma\right) \tag{2}$$

where s is the pixel position in a discrete, two-dimensional image plane denoted by Ω, η_s represents the neighborhood of pixel s, $w(x)$ is a spatial weighting function, and the constant σ is a scale (variance) parameter.

It shows that through the combination of the two components in (2), the intensity of pixel s is determined based on its neighboring attributes in terms of both the spatial positions and intensity differences. Specifically, the weighting function $w(x)$ gives more weight to the pixels at the center of the neighborhood and less weight to those at the periphery, for which a Gaussian function can be used. The robust error function $\rho(x)$ allows all inliers within the neighborhood to contribute towards smoothing, yet still retain the influence of pixels with significantly different intensities to keep sharp edges. A few other such robust energy functions are reviewed in [5], [6], [13].

In general, there are quite a few numerical methods that could be used in searching for an optimal solution of Eq. (1) and (2). In the following, we shall show that two different minimization procedures lead to two productive smoothers – the anisotropic diffusion and the bilateral filters.

In the general case, the expression in Eq. (1) does not admit closed-form solutions, so iterative methods are necessary. For example, one can apply a gradient descent method or Newton's method, yielding

$$g_s^{t+1} = g_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} w(p-s)\psi(f_p - g_s^t) \tag{3}$$

where $\psi(x) = \rho'(x)$, the first derivative $\rho(x)$, t denotes discrete iterations, λ is a scalar that determines the convergence rate, and $|\eta_s|$ is the number of neighborhood pixels involved. If we replace $\psi(x)$ by $xc(x)$, i.e., if we define the interaction function $c(x)$ by $c(x) = \psi(x)/x$, then (3) becomes

$$g_s^{t+1} = g_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} w(p-s)c(f_p - g_s^t)(f_p - g_s^t) \tag{4}$$

If we update image data f in iteration $t+1$ using the smooth image g_s^t resulting from the previous iteration t, the iterative update procedure in Eq. (4) is equivalent to the discrete formulation of anisotropic diffusion given in [2]. Hence, anisotropic diffusion can be viewed as a gradient descent solution (when the neighboring intensity differences are small) using the evolving sample set. Despite a straightforward iterative implementation, a major drawback of the gradient descent method is its relatively slow convergence rate, especially in the vicinity of the solution. To obtain a superlinear convergence rate, we describe a computational method called *iterative reweighting* that takes special advantage of the characteristics of the problem, and shows that the bilateral filter is indeed an instance of iterative reweighting with the evolving sample set.

The local minimum of Eq. (2), satisfies the following condition, by differentiation,

$$\sum_{p \in \eta_s} w(p-s)\psi\left(f_p - g_s\right) = \sum_{p \in \eta_s} w(p-s)c\left(f_p - g_s\right)\left(f_p - g_s\right) = 0 \tag{5}$$

The corresponding recursive estimate of Eq. (2) is then given by,

$$g_s^{t+1} = \frac{\sum_{p \in \eta_s} w(p-s)c\left(f_p - g_s^t\right)f_p}{\sum_{p \in \eta_s} w(p-s)c\left(f_p - g_s^t\right)} \tag{6}$$

which iteratively expresses the estimate as a weighted mean. We can rewrite Eq. (6) as follows,

$$\begin{aligned} g_s^{t+1} &= g_s^t + \frac{\sum_{p \in \eta_s} w(p-s) c\left(f_p - g_s^t\right)\left(f_p - g_s^t\right)}{\sum_{p \in \eta_s} w(p-s) c\left(f_p - g_s^t\right)} \\ &= g_s^t + \frac{\sum_{p \in \eta_s} w(p-s) \psi\left(f_p - g_s^t\right)}{\sum_{p \in \eta_s} w(p-s) c\left(f_p - g_s^t\right)} \end{aligned} \tag{7}$$

which shows that iterative reweighting has quadratic convergence similar to Newton's method, when the intensity differences are small and the influence function impacts the update. It is clear that the bilateral filter in [7] can be interpreted as a minimization procedure and is equivalent to the first iteration of (6) with the initialization, $g_s^0 = f_s$. This gives additional insight into the visually impressive smoothing properties exhibited by the bilateral filter within homogeneous regions. This also helps to understand the mean shift algorithm [17] as a bilateral filter with adaptive spatial neighborhoods. Additionally, the bilateral filter can also be viewed as anisotropic diffusion with adaptive step sizes. It can be shown that the iterative reweighting algorithm given by Eq. (6) or (7) is stable, that is the sequence $\left\{g_s^t\right\}$ is bounded, and converges to a local minimum for nonincreasing *c(x)* functions [19].

3 A Robust Smoother with Graduated Non-convexity

A nontrivial issue is the selection of an appropriate robust ρ-function in Eq. (2), as this greatly affects the edge-preserving behavior to prevent smoothing across edges. In order to analyze the behavior of a given ρ-function, the robust error function's influence function ψ shown in Eq. (3) plays an important role. The influence function measures the contribution of a neighboring pixel to the intensity estimate of the central pixel of interest. For instance, for the least squares (LS) estimator with $\rho(x) = x^2$, the influence function is $\psi(x) = 2x$. Since the influence of outliers increases linearly with the size of the error, sharp edges are discouraged and blurring occurs across edge-boundaries during the estimation or smoothing process.

To ameliorate the effect of outliers, the squared cost function can be replaced by more robust error models which grow subquadratically for large $|x|$ to reduce the effect of outliers. We restrict our development of a robust error norm to Tukey's biweight function [11] in order to keep the discussion specific (see Fig. 1 for the qualitative shapes of this robust estimator function). As shown in Fig. 1, the biweight function forces the ψ-function to return to zero when $|x|$, or the intensity difference, is greater than a specified positive number σ, and thus is able to alleviate smoothing across edges. Although a convex robust function such as the Huber function could be used, its influence does not drop to zero for gross errors and consequently, smoothing across edges will continue to occur. The redescending Tukey influence function prohibits smoothing across edges. However, the nonconvexity of the Tukey biweight ρ-function results in a nonconvex optimization problem. That is, the gradient descent method in Eq. (3) using the Tukey biweight function may have multiple solutions corresponding to multiple local minima of the cost function, and generally only one of them will correspond to the global minimizer. So given a poor starting value, an iterative computation may converge to a local minimum. Note that anisotropic diffusion and bilateral filtering both suffer from this local minimum problem as well.

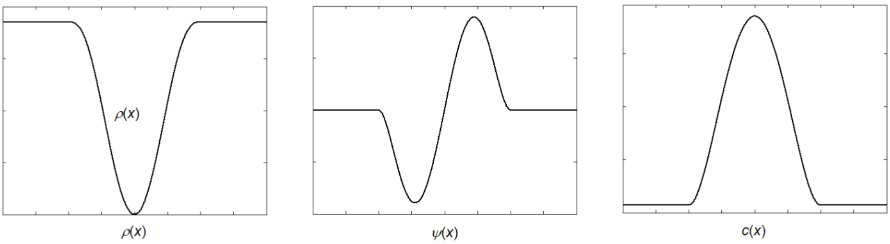

Fig. 1. The qualitative shape of the Tukey biweight function and its corresponding influence and interaction functions respectively

To avoid becoming trapped in a local optimum, stochastic or deterministic annealing is usually used in the literature. Stochastic annealing is appealing because it can minimize an arbitrary energy function. As a consequence of its generality, however, such technique requires exponential time and are extremely slow in practice. We propose a deterministic approach to the minimization of a robust error norm using graduated non-convexity (GNC). The GNC technique by Blake and Zisserman [15] is a discrete version of the continuation technique [16] which transforms a nonconvex optimization problem into a parameterized family of problems. This method first solves a simple convex problem, then 'bootstraps' itself towards a good solution of the nonconvex problem by searching local minima of a sequence of increasingly nonconvex approximate (relaxed) cost functions $\boldsymbol{E}_\gamma$ governed by a control parameter γ.

Specifically, γ is initially set to be a sufficiently large value $\gamma^{(0)}$ so that the first relaxed cost function $\boldsymbol{E}_{\gamma^{(0)}}$ is strictly convex. With such $\gamma^{(0)}$, the unique minimum of the $\boldsymbol{E}_{\gamma^{(0)}}$ is guaranteed by successively using any gradient based method, regardless of initial value. The minimum found in the current phase k is then used as an initial value for the next phase $k+1$ with a lower control value $\gamma^{(k)} < \gamma^{(k+1)}$ for $k = 0, 1, \ldots, K$. When $\gamma^{(k)}$ decreases from $\gamma^{(0)}$ to one, the convexity constraint on $\boldsymbol{E}_{\gamma^{(k)}}$ is relaxed monotonically and the minima progressively approach the desired solution, $\lim_{\gamma\to 1} \boldsymbol{E}_\gamma = \boldsymbol{E}$. In this manner, if we gradually track the global minimum from a high value of, γ to one, it is reasonable to expect that we can approximate the global minimum of $\boldsymbol{E}$ as $\gamma \to 1$.

The GNC algorithm for Tukey's biweight function, can then use the *dilation* approximation technique suggested by [18], admitting a sequence of relaxed functions as follows

$$\rho_\gamma(x,\sigma) = \begin{cases} (1-[1-(x/\gamma\sigma)^2]^3)/3 & |x| \le \gamma\sigma, \\ 1/3 & \text{otherwise.} \end{cases}$$
$$\psi_\gamma(x,\sigma) = \begin{cases} x[1-(x/\gamma\sigma)^2]^2 & |x| \le \gamma\sigma, \\ 0 & \text{otherwise.} \end{cases}$$
$$c_\gamma(x,\sigma) = \begin{cases} [1-(x/\gamma\sigma)^2]^2 & |x| \le \gamma\sigma, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

We can see that there exists γ_0 such that $\rho_\gamma(x,\sigma)$ is convex for any $\gamma > \gamma_0$. The relaxed biweight robust functions are plotted in Fig. 2 as are the corresponding interaction functions for various values of γ.

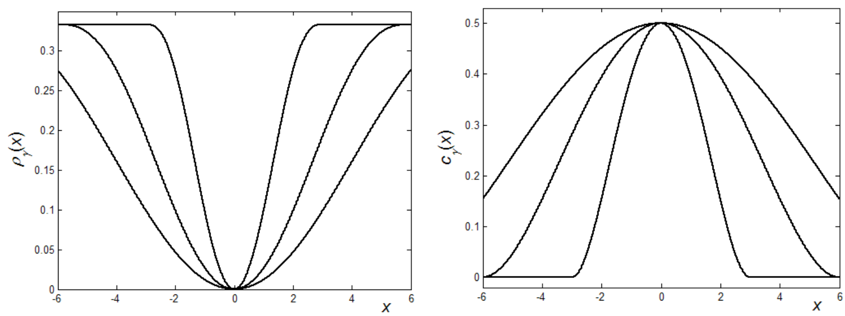

Fig. 2. Graduated nonconvexity (GNC) for (left) Tukey's biweight functions with three different values of γ and (right) their corresponding interaction functions

4 Experimental Results

In this section, we first describe the implementation details of our experiments, and then provide numerical simulation results using a set of standard test images degraded with different levels of Gaussian noise. The results are compared to a few well known methods.

The parameter γ provides a degree of freedom that we introduced in order to allow us to apply GNC methods (varying γ), which controls the relaxation speed. We use a *logarithmic* decrease in the value of γ by letting $\gamma^{(k)} = \gamma^{(0)} + \ln(k+1)(1-\gamma^{(0)})/{}'\ln(K+1)$ for $k = 0, 1, \ldots, K$. A typical value of K for moderate noise levels is $K = 5$. The other choice includes *linear* or *exponential* decrease schemes. A fast initial decrease of γ, as in the logarithmical decrease, is justified by the fact that GNC is particularly sensitive to a change in speed of γ during the late stages of the GNC minimization. We use the Gaussian function for the spatial weighting function $w(x)$. The size of neighborhood η_s is empirically chosen as 9×9 pixels, which occupies more than 90% of the support area of the Gaussian function $w(x)$. Image boundaries are handled by assuming symmetric boundary conditions. Iterative reweighting is adopted due to its favorable convergence property.

We compare our algorithm to three classical adaptive smoothing algorithms: anisotropic diffusion [2], bilateral filter [7], and the Bayes least squares Gaussian scale mixtures (BLS-GSM) by Portilla *et al.* [20] in the wavelet domain, due to the success these three nonlinear algorithms have achieved in the context of edge-preserving smoothing. Gaussian functions are used for the edge stopping functions in the anisotropic diffusion and bilateral filter, as suggested in [2] and [7]. For each method tested, we vary its parameters exhaustively to obtain the best possible results.

Performance was quantitatively evaluated by assessing the amount of smoothing across edges using the peak signal-to-noise ratio (PSNR) between the filtered or smoothed image and the original image (without noise). Larger PSNR values indicate better edge-preserving image smoothing behavior.

We have extensively tested the noise removal capability of the four algorithms using a set of standard test images degraded by Gaussian noise. Fig. 3 shows smoothing results of four adaptive smoothing algorithms on the noisy *Goldhill* image of Fig. 3(b). The original images is severely degraded with Gaussian noise with standard deviation up to σ_n = 30. The anisotropic diffusion and bilateral filter yield similar results; both methods are unable to completely remove the noise and also excessively smooth other regions of the image including across edges. In the result of BLS-GSM, some plateau regions are erroneously smoothed, creating ring-like artificial textured regions. The proposed algorithm produces noticeably better results than any of those obtained using the other methods as shown in Fig. 3 for the Goldhill test image. The visual improvement can be quantitatively confirmed using PSNR. Our algorithm leads to a large PSNR gain from 18.59 to 29.75 dB, while the other three algorithms improve PSNRs to 24.42 dB, 25.01 dB and 26.59 dB, respectively. Table 1 shows the PSNR results for three other test images with the same amount of synthetic noise added. In all cases the proposed robust edge preserving algorithm outperforms the other three competitive methods. Moreover, due to the robustness of Tukey's biweight function our algorithm yields a nearly identical resultant image in a wide range of iteration conditions, while anisotropic diffusion and bilateral filtering have to stop after a certain number of iterations in order to avoid over smoothing leading to a flat image. This provides a practical way to use more flexible termination conditions.

Table 1. PSNR improvement after filtering using four different smoothing algorithms applied to four test images with σ_n=30 or starting PSNR of 18.59

Image	Anisotropic Diffusion	Bilateral	BLS-GSM	Robust proposed
Peppers	28.03	28.17	28.01	33.02
Goldhill	24.42	25.01	26.59	29.75
Barbara	26.59	26.03	27.94	30.43
Flintstones	24.97	24.59	23.59	28.62

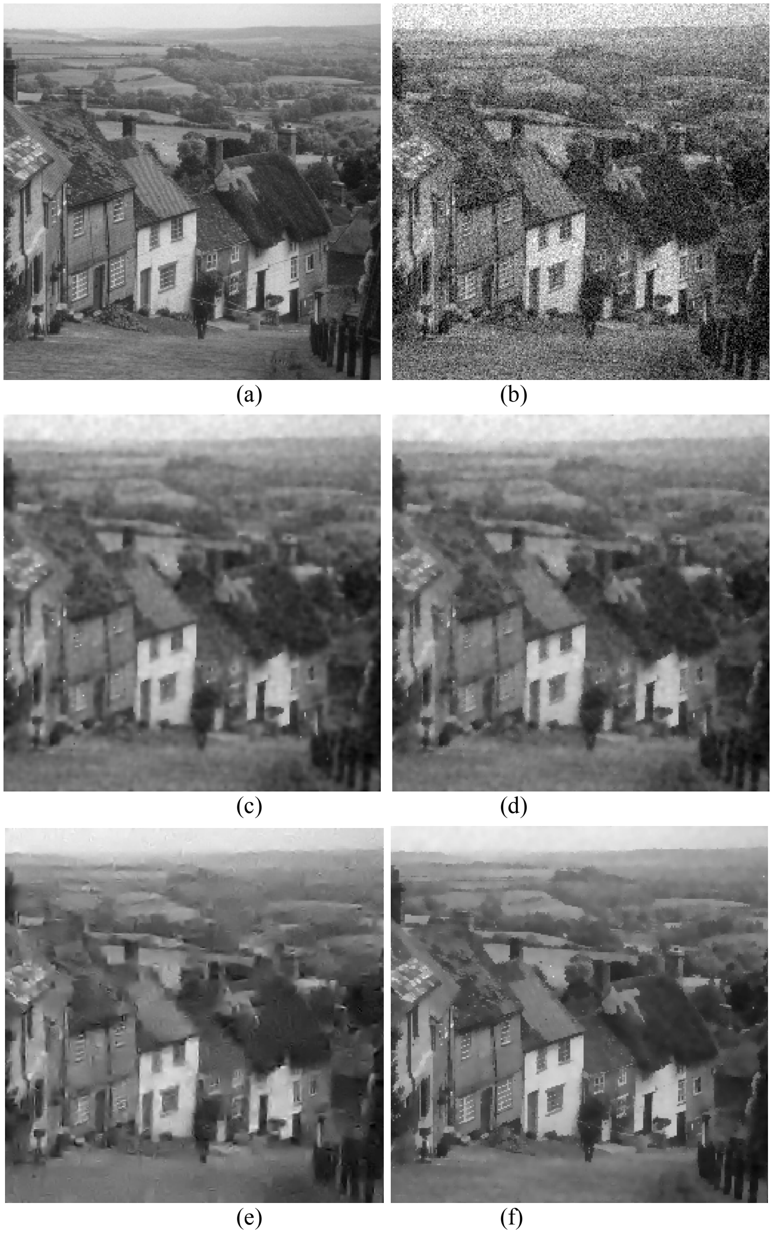

Fig. 3. (a) Original *Goldhill* image; (b) corrupted image (Gaussian noise, $\sigma_n = 30$) used as input image for results in (c)-(f); (c) anisotropic diffusion result, (d) bilateral filter result, (e) BLS-GSM filter result; (f) proposed algorithm result

5 Conclusions

This paper shows a strong relationship between anisotropic diffusion, bilateral filtering, and robust function optimization. We use the insight to develop a new robust formulation for edge-preserving smoothing that when coupled with the GNC numerical optimization scheme overcomes the local minima problems associated with iterative reweighting. Experiments show that the resulting non-linear filter preserves edges and reduces the distracting effects of noise better than other competitive approaches and is several dB better in terms of PSNR compared to anisotropic diffusion, bilateral filtering and the Bayes least squares Gaussian scale mixtures a wavelet-based method for image enhancement.

References

[1] Marr, D.C., Hildreth, E.C.: Theory of edge detection. Proc. Roy. Soc. London B207, 187–212 (1980)
[2] Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Machine Intell. 12, 629–639 (1990)
[3] Weickert, J.: Theoretical foundations of Anisotropic diffusion in image processing and computer vision. Computing 11, 221–236 (1996)
[4] You, Y.L., Xu, W., Tannenbaum, A., Kaveh, M.: Behavioral analysis of anisotropic diffusion in image processing. IEEE Trans. Image Processing 5, 1539–1553 (1996)
[5] Li, S.A.: Discontinuity-adaptive MRF prior and robust statistics: A comparative study. Image and Vision Computing 13, 227–233 (1995)
[6] Black, M.J., Sapiro, G., Marimont, D., Heeger, D.: Robust anisotropic diffusion. IEEE Trans. Image Processing 7, 421–432 (1998)
[7] Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proceeding of the IEEE International Conference on Computer Vision, pp. 59–66 (1998)
[8] Yaroslavsky, L.P.: Digital Picture Processing-An Introduction. Springer, Heidelberg (1985)
[9] Smith, S.M., Brady, J.M.: SUSAN - a new approach to low level image processing. International Journal of Computer Vision 23, 45–78 (1997)
[10] Huber, P.J.: Robust Statistics. John Wiley and Sons, New York (1981)
[11] Tukey, J.W.: Explortary Data Analysis. Addison-Wesley, Reading (1977)
[12] Zhuang, X., Palaniappan, K., Haralick, R.M.: Highly robust statistical methods based on minimum-error Bayesian classification. In: Chen, C.W., Zhang, Y.-Q. (eds.) Visual Information Representation, Communication and Image Processing. Optical Engineering Series, vol. 64, pp. 415–430. Marcel-Dekker, New York (1999)
[13] Nath, S.K., Palaniappan, K.: Adaptive robust structure tensors for orientation estimation and image segmentation. In: Bebis, G., Boyle, R., Koracin, D., Parvin, B. (eds.) ISVC 2005. LNCS, vol. 3804, pp. 445–453. Springer, Heidelberg (2005)
[14] Palaniappan, K., Jiang, H., Baskin, T.I.: Non-rigid motion estimation using the robust tensor method. In: IEEE Computer Vision and Pattern Recognition Workshop on Articulated and Nonrigid Motion, Washington, DC, June 27, pp. 25–33. IEEE Computer Society Press, Los Alamitos (2004)
[15] Blake, A., Zisserman, A.: Visual Reconstruction. MIT Press, Cambridge (1987)

[16] Wasserstrom, E.: Numerical solutions by the continuation method. SIAM Rev. 15, 89–119 (1973)
[17] Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24, 603–619 (2002)
[18] Nikolova, M.: Markovian reconstruction using a GNC approach. IEEE Trans. Image Processing 8, 1204–1220 (1999)
[19] Dong, G., Acton, S.T.: On the convergence of bilateral filter for edge-preserving image smoothing. IEEE Signal Processing Letters 14, 617–620 (2007)
[20] Portilla, J., Strela, V., Wainwright, M., Simoncelli, E.: Image denoising using scale mixtures of Gaussian in the wavelet domain. IEEE Trans. Image Processing 12, 1338–1351 (2003)