Shock Filters Based on Implicit Cluster Separation

Vinay P. Namboodiri Subhasis Chaudhuri*
Department of Electrical Engineering
Indian Institute of Technology, Bombay
Mumbai, India. 400076.
{vinaypn, sc} @ ee.iitb.ac.in

Abstract

One of the classic problems in low level vision is image restoration. An important contribution toward this effort has been the development of shock filters *by Osher and Rudin [1]. It performs image de-blurring using hyperbolic partial differential equations. In this paper we relate the notion of cluster separation from the field of pattern recognition to the shock filter formulation. A kind of shock filter is proposed based on the idea of gradient based separation of clusters. The proposed formulation is general enough as it can allow various models of density functions in the cluster separation process. The efficacy of the method is demonstrated through various examples.*

1. Introduction

The problem that is addressed in this paper is one of deblurring an image $Y(x)$ that has been blurred by a blurring kernel $h(x)$ representing some physical process. This problem is modeled by the convolution relation:

$$Y(x) = \int U(t)h(x-t)dt \tag{1}$$

where x can denote a $2D$ space in which case $U(x)$ might represent an image. As is normally assumed the function $h(x)$ has the properties that it is non-negative, and the integral of the function $h(x)$ is unity.

This problem has been investigated in many ways like Wiener filtering, deconvolution and shock filters. The disadvantages of other methods and the advantages of the shock filter based inverse diffusion approach have been convincingly argued by Osher and Rudin in their paper [1]. The problem of inverse diffusion is also of interest in the sense of scale space theory ([2] and [3]). Most theories on scale space are based on the notion of forward diffusion using heat equation or, as in the case of Perona-Malik [4], some non-homogeneous models of forward diffusion. Similar to the various models of forward scale spaces it would also be of theoretical and practical interest to have backward scale space theories based on inverse diffusion wherein one tries to go to the appropriate finer levels of scales.

The process of inverse diffusion has been suggested by using shock filters of Osher and Rudin [1]. The basic idea is to formulate a non-linear hyperbolic partial differential equation and to solve it within the constraints of conservation laws and by implying the total variation to be constant. Here we approach the problem from the viewpoint of cluster separation. Consider a piecewise continuous image which has been blurred. The process of blurring can then be thought of as mixing of various clusters where each piecewise continuous segment constitutes a cluster. The solution process can then be envisaged as one of separation of clusters. A method for clustering in pattern recognition has been the one based on gradient based estimation of kernel density [5]. This method was largely forgotten till recently [6]. It has also been used lately by Comaniciu and Meer [7] for discontinuity preserving smoothing and image segmentation. The method of gradient based clustering is adapted for the problem of gradient based cluster separation. We show that the resultant formulation is a kind of shock filter and further relate it to the classical shock filters. This method compares well with the original shock filter as well as the other shock filters developed recently.

The rest of the paper is organized as follows. In the next section we discuss the related work done in this area. In section 3 we consider cluster separation based on non-parametric gradient density estimates. In section 4 the shock filter formulation is described.In section 5 we relate the shock filter to the gradient based cluster separation and show how it is in essence a kind of shock filter. In section 6 the practical issues are discussed. In section 7 we present the experimental results. We conclude in section 8.

2. Related Work

The shock filter was first proposed by Kramer and Bruckner [8]. It is based on the idea to use a dilation process near a maximum and an erosion process around a minimum. The decision whether a pixel belongs to the influence zone of a

*Financial assistance under the *Swarnajayanti* Fellowship is gratefully acknowledged.

maximum or a minimum is made on the basis of its Laplacian. The term *shock filter* was first introduced by Rudin [9]. The experimental shock filter by Rudin was based on a modification of the nonlinear Burgers' equation. This model was further improved by Osher and Rudin in [1], where the total variation preserving computational approach and the theoretical basis for the same was developed. Further, a modification was suggested by Alvarez and Mazorra [10] where they incorporated a smoothing kernel in the model. The relationship of these methods to the Kramer-Bruckner filter became evident later ([11], [12])

The recent work in this field includes work by Kimmel *et al.* [13], Weickert [14], Gilboa *et al.* ([15], [16], [17]) and Remaki and Cheriet [18]. In [13], Kimmel *et al.* have developed a shock filter based on a geometric framework and the inverse diffusion is carried out along the edge. In [14], Weickert describes a coherence enhancing shock filter where the shock filter is steered with the orientation information. In [15] by Gilboa *et al.*, the authors have modified the diffusion coefficient in the Perona-Malik formulation [4] and they use a diffusion coefficient which switches adaptively between forward and backward diffusion process. In [16], Gilboa *et al.* extend the work done in [15] and define a triple-well potential based diffusion process which is an energy minimizer flow aimed at reducing oscillations among three low energy states. In [17], Gilboa *et al.* suggest complex shock filters based on the complex diffusion process where the diffusion coefficient lies in the complex domain. In [18], the authors consider the problem of shock filters in the framework of generalized functions and propose shock filters where the speed of shock propagation is also controlled.

3. Gradient Based Cluster Separation

A technique for clustering a set of points is to explicitly move the points in the direction of the gradient of the kernel density estimates ([5], [7]). Since the true probability density function or even its form is not known, non-parametric techniques are used to obtain estimates of the density gradient [19]. The approach is to obtain a differentiable nonparametric estimate of the probability density function and then its gradient is computed.

Let $X_1, X_2, ...X_N$ be a set of N independent and identically distributed n-dimensional random vectors in the d-dimensional feature space $\mathbb{R}^d$ and a symmetric positive definite $d \times d$ bandwidth matrix $\mathbf{H}$ ([7], [19]). The kernel density estimators have the form

$$\hat{f}(X) = \frac{1}{N}\sum_{j=1}^{N} k_{\mathbf{H}}(X - X_j), \tag{2}$$

where $k(X)$ is a bounded kernel function with compact support satisfying

$$\lim_{\|X\|\to\infty} \|X\|^d k(X) = 0 \text{ and } \int_{R^d} k(X)dX = 1.$$

A fully parameterized $\mathbf{H}$ increases the complexity of the estimation and, in practice, the bandwidth matrix $\mathbf{H}$ is chosen to be the identity matrix $\mathbf{H} = h^2\mathbf{I}$. Then the kernel density estimator takes the form

$$\hat{f}(X) = \frac{1}{Nh^d}\sum_{j=1}^{N} k\left(\frac{X - X_j}{h}\right). \tag{3}$$

A differentiable kernel function is used and then the density gradient is estimated as gradient of eqn.(3). This gives the density gradient estimate as

$$\nabla\hat{f}(X) = (Nh^d)^{-1}\sum_{j=1}^{N} \nabla_x k(h^{-1}(X - X_j)) \tag{4}$$

$$= (Nh^{d+1})^{-1}\sum_{j=1}^{N}(X - X_j)\nabla k(h^{-1}(X - X_j)). \tag{5}$$

Eqn.(4) is the general form of the density gradient estimate. If one uses the Gaussian probability density kernel function and uses the gradient of the Gaussian kernel then the resulting estimate of the density gradient is

$$\nabla\hat{f}(X) = (N)^{-1}\sum_{i=1}^{N}(X_i - X)(2\pi)^{-n/2}h^{-(n+2)} \\ .\exp\left[-(X - X_i)^T\left(\frac{X - X_i}{2h^2}\right)\right]. \tag{6}$$

In [5], the authors point out how eqn(6) is essentially a weighted measure of the mean shift of the observations about the point X. In order to move the values, the estimate of mean shift of the normalized gradient is used. The mean shift of the normalized gradient is

$$\frac{\nabla\hat{f}(X)}{\hat{f}(X)} = \nabla \ln \hat{f}(X). \tag{7}$$

The method for gradient based clustering then is essentially a recursive algorithm to transform each observation according to the clustering algorithm

$$X_j^{i+1} = X_j^i + a\nabla \ln \hat{f}(X_j^i). \tag{8}$$

Here a is a constant which determines the rate of convergence of the clusters.

Figure 1 illustrates the process of clustering that occurs as a result of eqn.(8). The cluster values when moved in the direction of the density gradient, come closer to each other and, finally move to the mean of the cluster. This process

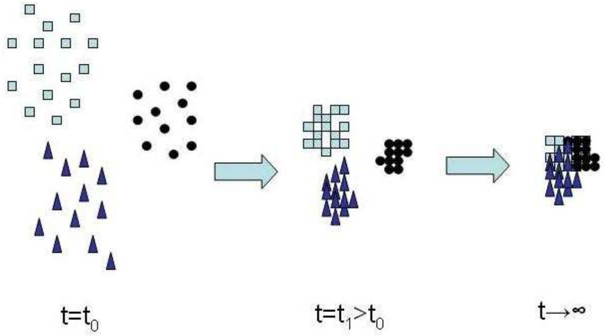


Figure 1: Illustration of the forward diffusion process in terms of image features. (a) At $t = t_0$, (when there is no blurring), the feature clusters are well separated. (b) At $t = t_1 > t_0$, due to blurring the features move closer to each other for each cluster and (c) at sufficient blurring, the feature clusters merge when they are indistinguishable.

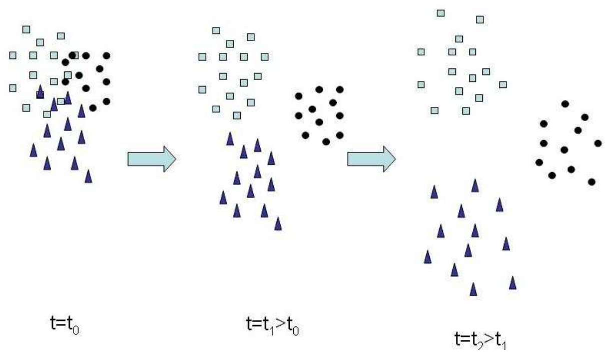


Figure 2: Illustration of inverse diffusion in the feature space (which has some blurring). As we do the inverse diffusion, the feature space separate out by moving farther away from each other. This is a divergent process.

is similar to the forward diffusion where the image diffuses toward the average gray scale. When the image is already blurred, the points belonging to various clusters are mixed together and they need to be separated. Therefore we modify eqn.(8) and move the values of the various points away in the direction of the respective density gradient estimated using the non-parametric density gradient estimation technique. The figure 2 illustrates the resultant difference in the movement of the points.

This is a kind of inverse diffusion process as the cluster separation process when done for an image results in sharpening of the line fields. The equation for the gradient based cluster separation algorithm takes the form

$$X_j^{i+1} = X_j^i - a\nabla \ln \hat{f}(X_j^i) \tag{9}$$

The process of cluster separation is done here without explicit delineation of the values into clusters or segmentation in the image space. Hence this is a kind of implicit cluster separation process. The process yields the deblurred image as output corresponding to the clusters being separated. However the cluster separation process is divergent, and the process is controlled by incorporating a factor based on cluster separation distance. This is discussed in section 6.

4. Shock Filters

The shock filters [1] are based on the scalar conservation law

$$u_t + f(u)_x = 0 \tag{10}$$

which is solved for $-\infty < x < \infty$ $(x \in R^1)$, $t > 0$ with initial data $u(x, 0) = u_0(x)$. If $f'' \neq 0$, then the solution generally develops discontinuities even for very smooth $u_0(x)$. In finite time, the characteristics of the solution generally intersect and shocks develop, i.e. the solution becomes a weak solution. The solution, with the shocks is obtained through a single, globally defined algorithm. The numerical approximation is done by using grid approximation ($x_i = ih, t^n = n\triangle t$). The shock-capturing approximation is given by the following equation ([1])

$$u_i^{n+1} = u_i^n - \frac{\triangle t}{h}(g_{i+1/2}^n - g_{i-1/2}^n), \tag{11}$$

where,

$$g_{i+1/2}^n = g(u_{i-k}^n, \ldots, u_{i+k+1}^n) \tag{12}$$

is the numerical flux approximating $f(u)$ and g is Lipschitz continuous. The scalar conservation form alone is not sufficient to formulate a solution scheme. In [1], Osher and Rudin have formulated a scheme using preservation of total variation as a constraint. The form of the function is

$$u_t = -|u_x| F(u_{xx}) \tag{13}$$

in one dimension. The authors have actually taken F to be a constant c, i.e. the eqn(13) has the form

$$u_t = -|u_x| c \tag{14}$$

In two dimensions the form is

$$u_t = -\sqrt{u_x^2 + u_y^2} F(L(u)) \tag{15}$$

where they have taken $L(u)$ as

$$L(u) = u_{xx}u_x^2 + 2.u_{xy}u_xu_y + u_{yy}u_y^2 \tag{16}$$

and $F(u)$ is normalized.

The Osher-Rudin model has been extended by Alvarez-Mazorra [10] and they have proposed the following model

$$u_t + F(G_\sigma * u_{xx}, G_\sigma * u_x)u_x = 0 \tag{17}$$

where $G_\sigma(.)$ is a family of smoothing kernels which depends on a parameter σ (for instance, a family of Gaussians), $F(.,.)$ is a function which satisfies:

$$F(p,q)pq \geq 0 \text{ for any } p,q \in R. \tag{18}$$

5. Relation to Shock Filters and Generalization

The proposed formulation of gradient separation in eqn(9) is of the conservation form (eqn10). Consider the eqn(11) and eqn(9). The function g in eqn(11) is taken to be the function $a\nabla_x \ln \hat{f}(X_j)$ in eqn(11). This is so because the function $a\nabla_x \ln \hat{f}(X_j)$ satisfies the Lipschitz condition. The Lipschitz condition states that

$$||f(u_2) - f(u_1)|| < L||u_2 - u_1|| \tag{19}$$

for some finite number L in some region R of f. The Lipschitz condition is implied if the function $f(u_2)$ has finite partial derivatives [20]. Since the kernel function k is chosen to be differentiable, the gradient based cluster separation function satisfies the Lipschitz condition.

The current formulation shows the most similarity to the Alvarez-Mazorra [10] formulation. The principal way in which the model differs from the Alvarez-Mazorra model is that instead of considering u_x, that is the gradient of u, the gradient of $k(u)$ is considered. This is quite significant as this increases the robustness of the value considered. In regions which are flat shaded the gradient is zero. In areas where the edges have been blurred, the direction of the density gradient indicates the direction of the dominant edge and helps in restoring the image to its deblurred form. The use of density gradient in the Alvarez-Mazorra model would extend it. In a similar manner, the use of density gradient estimates instead of u_x can be used in generalizing the other models of shock filters as well.

The current formulation is not total variation preserving as eqn(9) does not have a variation preserving term as is present in eqn(15). Hence it does not exhibit nice convergence and stability properties exhibited in eqn(15). However, in order to stabilize the inverse diffusion, a modification done to the recursive formulation is to incorporate a criteria for stabilizing the diffusion based on the value of the gradient density function indicating the cluster separation. In case of a uniform region the density estimate is zero and the diffusion is stable. In case of non-zero density gradient, the value of the density function increases in the most rapid direction. When the value increases beyond a threshold indicating that the clusters are separated fully, then the diffusion is saturated and further diffusion is not done.

6. Implementation Issues

The formulation presented in the eqn(9) is quite general enough as various forms of the kernel density functions can be incorporated. One of the main sources for blurred images is capturing of real images using a finite aperture lens where the object is not in focus. In such a case the point spread blur function can be modeled as a Gaussian [21]. Even if the original density function is not explicitly known, the non-parametric kernel function generally provides a good enough approximation to the unknown underlying model. Another criterion to be considered is the parameter h in eqn(4). Currently since the application is not a general cluster separation of N dimensional points, but that of de-blurring of an image, an eight neighborhood based support suffices. This implicitly limits the variance of the Gaussian kernel used for the density function. However this is not a factor due to the recursive nature of the algorithm and the commutative property of the Gaussian function.

A factor which has to be considered in the implementation of the algorithm is the threshold to be used for saturating the inverse diffusion. This is taken as a relative percentage of the initial density gradient estimate. Practically we have found that a factor of 75% proves adequate for indicating the cluster separation, i.e. we set $a = 0$ whenever

$$|X_j^{i+1} - X_j^i| \geq 0.75|X_j^i| \quad \forall j. \tag{20}$$

At this stage, the diffusion would not be Lipschitz continuous, however, that does not affect the diffusion since the diffusion is stopped at that time. In case of homogeneous regions the density gradient estimate is approximately zero and in those regions not much sharpening happens. This is taken into account and the pixels in the uniform regions do not determine the stopping criterion of the inverse diffusion. Another stopping criterion could be one based on the relative cluster separation, and a form of thresholding can be used on the cluster separation metric to determine the point for stopping the inverse diffusion.

7. Experimental Results

We first compare the performance of our method with that of Osher-Rudin filter. Here the Lena image has been blurred with a Gaussian blur with zero mean and variance 3.0. The results are shown in the fig. 3. The fig. 3(a) shows the input image which is blurred. The fig. 3(b) shows the result of the Osher-Rudin model. The Osher-Rudin model appears to be an "impressionistic" output of the original. This has been pointed out by the authors in their paper [1]. Figure 3(c) and 3(d) show the results obtained from the Alvarez-Mazorra method [10] and the Gilboa *et al.* complex diffusion shock filter [17]. The results obtained using these shock filters are not impressive. This is particularly because they are primarily designed to handle noisy blurred images. In case there is no noise and only de-blurring needs to be done, then they do not perform well. The fig. 3(e) shows the result of using Lucy-Richardson algorithm which is a standard deconvolution algorithm available in Matlab. The exact point spread

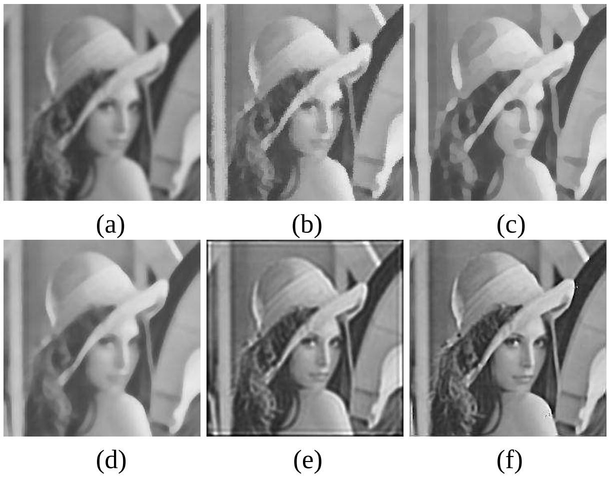


Figure 3: Result of inverse diffusion of (a) the blurred Lena image using (b) Osher-Rudin shock filter, (c) Alvarez-Mazorra shock filter, (d) Gilboa *et al.* complex shock filter , (e) Lucy-Richardson deconvolution and (f) gradient based cluster separation shock filter.

function was given as input and yet one can observe certain ringing effects in the result. The fig. 3(f) shows the result of our method. Our method is able to successfully restore most of the blurred edges to their original form. This is primarily because the density gradient determines a better estimate of the gradient direction in which the image has to be restored compared to the original gradient being computed.

Next we consider the case where a space varying Gaussian blur is applied to the Lena image. We have applied a radially varying Gaussian blur with the variance ranging from 1.0 in the center to 2.0 at the boundaries in a radially symmetric manner. The result of de-blurring by our method is shown in fig. 4. This shows that the method can also handle inhomogeneous blur adequately. The next experiment was done using real data set where a defocused image of a ball was captured. The result of de-blurring is shown in fig. 5(b). In this case a few disturbances can be noticed. This is primarily a result of the quantization inherent when the data set is stored using 8 bits. As a result the de-blurring process generates a few anomalies due to spurious shocks being generated. We suggest the use of a higher accuracy of 16 bits while storing the real data set to avoid the anomalies.

In the last case we consider a case where the blurring model is not Gaussian. A spatial averaging blur model is considered for a satellite image. This is a practical aspect particular to a kind of satellite imagery. Here we can observe in the fig. 6 that the method is able to successfully decipher the underlying details in different blur models as well. This is achieved using a Gaussian kernel density function itself while computing the density gradient. This shows

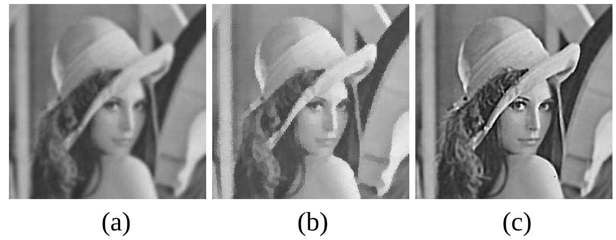


Figure 4: Result of space varying blind deconvolution with (a) input image using (b) Osher-Rudin shock filter, and (c) gradient based cluster separation process.

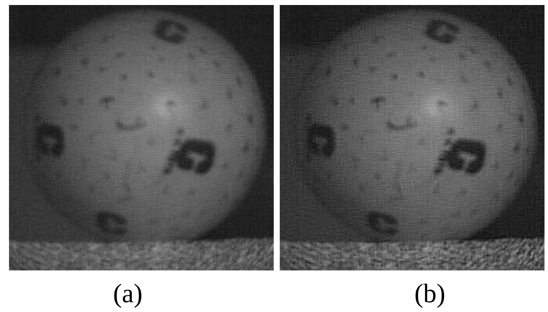


Figure 5: Performance of space varying blind deconvolution for a real aperture image where (a) is the input image and (b) the result using gradient based cluster separation process.

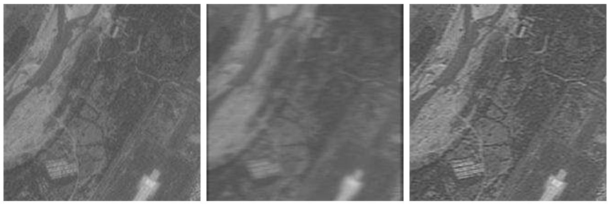
the resilience of the method for de-blurring general kinds of blur.

8. Conclusion

In this paper we have developed a shock filter based on the implicit cluster separation. An interesting insight that has been obtained is that the gradient of the density estimate determines the dominant edge direction in a better way while doing the inverse diffusion. Thus instead of using the gradient at each point, one can use the gradient of the density estimate and obtain more generalized variants of the existing shock filters. We also demonstrate the need for delimiting the cluster separation process to prevent the inverse diffusion from diverging. The method has been experimentally compared with several existing shock filters.

References

[1] S. Osher and L. I. Rudin, "Feature-oriented image enhancement using shock filters," *SIAM Journal of Numerical Analysis*, vol. 27, no. 4, pp. 919–940, August 1990.



(a) (b) (c)

Figure 6: Illustration with a 1-D blur kernel where (a) is the original image which has been blurred as shown in (b) in the horizontal direction, and in (c) the image is de-blurred using the proposed technique.

[2] A. P. Witkin, "Scale-Space Filtering," in *Proc. of the 4th International Joint Conference on Artificial Intelligence*, Kalsruhe, West Germany, 1983, pp. 1019–1022.

[3] J. J. Koenderink and A. J. van Doorn, "Dynamic shape," *Biological Cybernetics*, vol. 53, pp. 383–396, 1986.

[4] P. Perona and J. Malik, "Scale Space and Edge Detection using Anisotropic Diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[5] K. Fukunaga and L. D. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, January 1975.

[6] Y. Cheng, "Mean shift, mode seeking and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, August 1995.

[7] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[8] H. P. Kramer and J. B. Bruckner, "Iterations of a nonlinear transformation for enhancement of digital images," *Pattern Recognition*, vol. 7, no. 1-2, pp. 53–58, 1975.

[9] L. I. Rudin, "Images, numerical analysis of singularities and shock filters," Tech. Rep. TR 5250:87, Computer Science Department, California Institute of Technology, December 1987, Ph.D. Thesis.

[10] L. Alvarez and L. Mazorra, "Signal and image restoration using shock filters and anisotropic diffusion," *SIAM Journal of Numerical Analysis*, vol. 31, no. 2, pp. 590–605, April 1994.

[11] F. Guichard and J.-M. Morel, "A note on two classical shock filters and their asymptotics," in *Scale-Space and Morphology in Computer Vision, volume 2106 of Lecture Notes in Computer Science*, M. Kerckhove, Ed., 2001, pp. 75–84.

[12] J. J. Gerbrands J. G. M. Schavemaker, M. J. T. Reinders and E. Backer, "Image sharpening by morphological filtering," *Pattern Recognition*, vol. 33, pp. 997–1012, 2000.

[13] R. Malladi R. Kimmel and N. Sochen, "Images as embedded maps and minimal surfaces: Movies, color, texture and volumetric medical images," *International Journal of Computer Vision*, vol. 39, no. 2, pp. 111–129, 2000.

[14] J. Weickert, "Coherence enhancing shock filters," in *Pattern Recognition, Lecture Notes in Computer Science, Vol. 2781, Springer, Berlin, Proceedings of DAGM 2003*, 2003, pp. 1–8, in Magdeburg, Germany -September 10-12 2003.

[15] N. Sochen G. Gilboa and Y. Y. Zeevi, "Forward-and-backward diffusion processes for adaptive image enhancement and denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 7, pp. 689–703, July 2002.

[16] N. Sochen G. Gilboa and Y. Y. Zeevi, "Image sharpening by flows based on triple well potentials," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 121–131, 2004.

[17] N. Sochen G. Gilboa and Y. Y. Zeevi, "Image enhancement and denoising by complex diffusion processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 1020–1036, 2004.

[18] L. Remaki and M. Cheriet, "Numerical schemes of shock filter models for image enhancement and restoration," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 2, pp. 129–143, 2003.

[19] P. E. Hart R. O. Duda and D. G. Stork, *Pattern Classification, 2nd ed.*, Wiley Interscience, New York, NY, 2001.

[20] A. H. Barr B. V. Herzen and H. R. Zatz, "Geometric collisions for time-dependent parametric surfaces," *SIGGRAPH Comput. Graph.*, vol. 24, no. 4, pp. 39–48, 1990.

[21] S. Chaudhuri and A. N. Rajagopalan, *Depth From Defocus: A Real Aperture Imaging Approach*, Springer Verlag, 1999.