UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Variational Restoration of Non-Flat Image Features: Models and Algorithms

Tony Chan
Jianhong Shen

June 1999

CAM Report 99-20

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90095-1555

http://www.math.ucla.edu/applied/cam/index.html

UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Variational Restoration of Non-Flat Image Features: Models and Algorithms

Tony Chan
Jianhong Shen

June 1999

CAM Report 99-20

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90095-1555

http://www.math.ucla.edu/applied/cam/index.html

VARIATIONAL RESTORATION OF NON-FLAT IMAGE FEATURES: MODELS AND ALGORITHMS

TONY CHAN AND JIANHONG SHEN

COMPUTATIONAL AND APPLIED MATHEMATICS
DEPARTMENT OF MATHEMATICS
LOS ANGELES, CA 90095-1555
chan, jhshen@math.ucla.edu

ABSTRACT. We develop both mathematical models and computational algorithms for variational denoising and restoration of non-flat image features. Non-flat image features are those that live on Riemannian manifolds, instead of Euclidean spaces. Familiar examples include the *orientation* feature (from optical flows or gradient flows) that lives on the unit circle $\mathbf{S}^1$, the *alignment* feature (from fingerprint waves or certain texture images) that lives on the real projective line $\mathbb{RP}^1$, and the *chromaticity* feature (from color images) that lives on the unit sphere $\mathbf{S}^2$. In this paper, we apply the variational method to the restoration of non-flat image features. General mathematical models for both continuous image domains and discrete domains (or *graphs*) are constructed. Riemannian objects such as *metric, distance* and *Levi-Civita connection* play important roles. We also discuss efficient computational algorithms for the non-linear restoration equations. Numerical results support both the models and algorithms developed in this paper. The mathematical framework can also be applied to the restoration of general non-flat data outside the scope of image processing and computer vision.

Date: June 10, 1999. Find typos? Why wait? Email "jhshen@math.ucla.edu."

1991 *Mathematics Subject Classification.* Primary: 9412; Secondary: 65H10, 49M37.

Key words and phrases. Variational model, total variation, denoising and restoration, non-flat features, Riemannian manifold, metric and distance, orientation, alignment, chromaticity.

Research supported by grants from NSF under grant number DMS-9626755 and from ONR under N00014-96-1-0277.

1. Introduction

In image analysis, restoration and denoising by variational methods are typically applied to gray level image functions or vector-valued (color) image functions (see Mumford and Shah [10], Rudin, Osher and Fatemi [15], and the monographs by Morel and Solimini [9], and Weickert [21]). Such functions are "flat" or affine. By flatness or affinity, we mean a function that takes values in a linear space, such as the real line, or the linear RGB space.

There are some important image features that do not live in flat spaces, however. One example is the *orientation feature*, which comes from optical flows or gradient flows (of gray images). One can easily see that in a video that records the rigid motion of an object against a static background, the orientation of the motion stores one half of the whole valuable information (and *speed* is the other half). The second example is the *alignment feature* which is commonly found in many texture images. For example, for fingerprints, the alignment pattern of fingerprint "waves" is crucial for identification tasks (see Perona [12]). *The chromaticity feature* of color images is another example of non-flat features (see Tang, Sapiro and Caselles [19, 18]). For a linear RGB color image, the pixel value $\mathbf{I}$ is a vector in $\mathbb{R}^3$. The magnitude $\|\mathbf{I}\|$ measures the brightness, while its trace on the unit sphere $\mathfrak{f} = \mathbf{I}/\|\mathbf{I}\|$ contains the color saturation information, and is called the *chromaticity feature* in Tang, Sapiro and Caselles [18]. It lives on the unit sphere $\mathbf{S}^2$ in the linear RGB space.

The mathematical object for modeling non-flat features is differential manifolds. Perceptual sensitivity to a particular type of feature is measured by Riemannian metrics. Therefore, our models shall be inevitably based on Riemannian manifolds. The complete picture has been summarized in Figure 1. The first step is taken by image processing or video processing

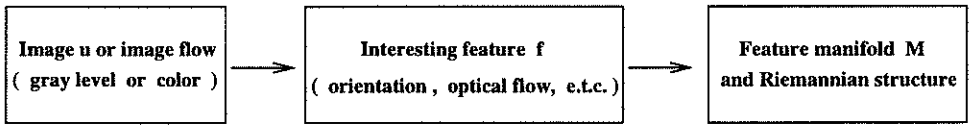

FIGURE 1. The flow chart of model setup.

researchers. Meaningful features are often defined according to our visual system or the type of task we attempt to carry out. The second step leads to mathematical analysis. Once the feature is described, the manifold stands

there. But a truly perceptually meaningful Riemannian metric depends on our visual system. For example, orientations in 2-D images live on the unit circle. It is quite reasonable to assume that generally our visual system does not prefer any single angle ($\pi/4$, say). Thus we can use the standard Riemannian structure (rotationally invariant) for orientations. Sometimes the converse is true: familiar manifolds usually have their natural (related to certain transform groups on symmetry) Riemannian metrics, and one can directly start with them for convenience.

The starting point of this paper is the last box. That is, we assume that a perceptually meaningful Riemannian structure is already defined.

As for flat features of an image, variational method is one of the most efficient tools for analyzing non-flat features. This is primarily due to its denoising and restoration effects. The nonflat features obtained initially are often noisy due to common factors like

(1) the noise contaminating original images, or left there even after a denoising pre-processing;
(2) regional destruction of raw images (due to the aging of a film, say), or insufficiency of resolution in raw images;
(3) inefficiency of the algorithm that extracts the non-flat features.

As in the classical literature, properly designed variational models can (i) attenuate noise, (ii) preserve intrinsic singularities, (iii) enhance the "edges" of non-flat features. For general features, an "edge" means a segment where the feature distribution undergoes abrupt changes. For the orientation feature from optical flows, for example, the edges may be identical to the real physical boundaries of moving bodies.

Perona has taken the pioneering step in [12] by studying a variational model for orientation diffusion with applications to fingerprint analysis. His work defines the right framework for further deeper mathematical and computational study on the subject. Our present work is its followup and extends it both theoretically and computationally. (Note: By the time we completed the initial version of this paper, we found surprisingly (and happily) that Sapiro and his colleagues were also developing the diffusion theory for directions (corresponding to the feature manifold $\mathbf{S}^m$) independently (see [18, 19]), almost at the same time. We shamelessly admit that some ideas on chromaticity presented in this final version are impossible without benefiting from this inspiring team.)

There are two key mathematical questions upon studying variational restoration of non-flat features. First, it is not as obvious as in Euclidean spaces to come up with a feasible energy functional for non-flat features. Perona [12] offered some bright physical intuitions for the orientation feature, but a general mathematical framework is still missing, which may delay many potentially important applications for other non-flat features. Secondly, as one can easily expect, the associated Euler-Lagrange equations or steepest-descent (diffusion) equations are generally non-linear because of the non-flatness of feature manifolds and the intrinsic non-linearity of their classical counterparts (the TV model, for instance). To get our mathematical models truly work for applications, it is important to develop efficient (simple, stable, or fast) numerical algorithms for these non-linear equations.

Answering these two fundamental questions is the main task of our present paper. We present a systematic way for constructing the energy functions for general non-flat data or features. We interpret Perona's original model in this general framework and discuss its generalizations. The mathematical framework also extends the classical *total variational* (TV) model. Numerical evidence shows that the TV model is more successful than the L^2 model for restoring non-flat features which have intrinsic singularities. We also design efficient numerical algorithms to solve the non-linear restoration equations. Numerical results highlight the success of both our mathematical framework and computational algorithms.

The organization of the paper is as follows. Section 2 constructs the mathematical foundation for the work of both Perona and ours. We study the restoration of non-flat features in a general and abstract way. The starting point is the construction of the energy functional on a Riemannian feature manifold. In section 3, we discuss restoration models specifically for orientations and alignments, two non-flat image features frequently encountered in gray level image analysis. Construction of energy functionals is through the new concept—Locally Riemannian Distance. The latter half of the section is devoted to developing numerical algorithms that can solve efficiently the nonlinear equations associated with the orientation and alignment features. Section 4 studies the computational issues for the chromaticity restoration on $\mathbf{S}^2$. A simple two-step filtering algorithm on the sphere is established for the discrete models. Numerical examples are gathered in section 5, where we discuss applications in fingerprint images, optical flows, and color images.

Whereas our work was initially motivated by image processing, we believe that the mathematical framework and algorithms can find their applications in the restoration of general non-flat data in other fields as well.

2. Variational Models and Restoration Equations

Since non-flat features do not live in a flat space, the energy functionals and associated restoration equations depend on the "geometry" of the non-flatness. The geometry is described by Riemannian manifolds and Riemannian metrics. These are the basic objects that we work with in this section. We shall establish the most general models for the restoration of non-flat features.

In the two subsections, we study models for both continuous image domains and discrete image domains (or graphs, more generally). The latter has the advantages of easier implementation in computers and friendly accessibility to the widest readers.

2.1. Models for continuous image domains.

Let Ω be an image domain in $\mathbb{R}^2$. For the purpose of image analysis, Ω can be a square, a disk, or other regular domains.

Suppose that the non-flat feature of interest lives on an m-dimensional Riemannian manifold M. We shall call M the *feature manifold*. For examples, the unit circle $\mathbf{S}^1$ is the feature manifold for orientation (or arrows with unit length), while the real projective line $\mathbb{RP}^1$ is the feature manifold for alignment (or line segments with unit length).

An image is a real function $u : \Omega \to \mathbb{R}^n$ ($n = 1$ or 3, practically). The collection of "all" images is denoted by the symbol $(\Omega \to \mathbb{R}^n)$. One can put the class labels such as L^p or BV (bounded variation) in front of it. But since we do not study the theoretical sides (i.e., existence and uniqueness) of the model equations in the present paper, the symbol allows more freedom. Moreover, the regularity of most images from natural scenes is usually poor since they belong to the statistical category in a large degree.

A map $\mathfrak{f} : \Omega \to M$ is called a *feature distribution* and $(\Omega \to M)$ denotes the collection of all feature distributions of interest. An M-feature of images is a map

$$F : (\Omega \to \mathbb{R}^n) \longrightarrow (\Omega \to M),$$

such that for any image u, $\mathfrak{f} = F_u$ is a feature distribution. That is, for any pixel $\boldsymbol{p} = (x, y) \in \Omega$, $\mathfrak{f}(\boldsymbol{p}) = F_u(\boldsymbol{p})$ is a (feature) point on M.

It is often true in applications, as outlined in the introduction section, that the M-feature distribution $\mathfrak{f} = F_u$ associated to a given image u is often noisy. Our major goal is to construct denoising and restoration models using the classical tool of variational method.

In practice, a feature distribution $\mathfrak{f} \in (\Omega \to M)$ is typically a map without overall smoothness. However, when we constructing a restortation model, as well practiced in the case of affine features, $\mathfrak{f}(\boldsymbol{p})$ is always assumed to have certain degree of regularities. For instance, we can assume that $\mathfrak{f}(\boldsymbol{p})$ belongs to the Sobolev space $W^{1,2}(\Omega, M)$. In numerical computation, it is fortunately true that models based on the good regularity assumption often work effectively for irregular images as well. Next section discusses discrete models, for which the regularity of data causes no problem in both theory and computation.

Under this setup, $\partial_x \mathfrak{f}(\boldsymbol{p})$ and $\partial_y \mathfrak{f}(\boldsymbol{p})$ are two tangent vectors in the tangent space $T_{\mathfrak{f}(\boldsymbol{p})}M$. Let $\langle \cdot, \cdot \rangle_{\mathfrak{f}(\boldsymbol{p})}$ denote the Riemannian inner product at $\mathfrak{f}(\boldsymbol{p})$, and $\|\cdot\|_{\mathfrak{f}(\boldsymbol{p})}$ the induced norm in $T_{\mathfrak{f}(\boldsymbol{p})}M$. We shall omit the location subscript if it is clear or insignificant in the context.

We now define the *strength* function

$$e(\mathfrak{f}; \boldsymbol{p}) = \sqrt{\|\partial_x \mathfrak{f}(\boldsymbol{p})\|^2 + \|\partial_y \mathfrak{f}(\boldsymbol{p})\|^2}. \tag{1}$$

To have a better understanding of its meaning, we introduce the correlation matrix

$$C_{\mathfrak{f}}(\boldsymbol{p}) = \begin{bmatrix} \langle\, \partial_x \mathfrak{f},\, \partial_x \mathfrak{f}\, \rangle & \langle\, \partial_x \mathfrak{f},\, \partial_y \mathfrak{f}\, \rangle \\ \langle\, \partial_y \mathfrak{f},\, \partial_x \mathfrak{f}\, \rangle & \langle\, \partial_y \mathfrak{f},\, \partial_y \mathfrak{f}\, \rangle \end{bmatrix}.$$

Then

$$e(\mathfrak{f}; \boldsymbol{p}) = \sqrt{\operatorname{trace}(C_{\mathfrak{f}}(\boldsymbol{p}))} = \sqrt{\sigma_1^2 + \sigma_2^2},$$

where σ_1 and σ_2 are the singular values of the m (the dimension of M) by 2 "matrix" $[\partial_x \mathfrak{f}, \partial_y \mathfrak{f}]$ (which is well-defined under any local orthonormal frame on M). A similar formulation has also been proposed in Sapiro [17, 16] when $\mathfrak{f}$ is a vector valued distribution.

The strength function defined in this way is rotationally invariant. That is, for any orthogonal transform in the image domain

$$\mathbf{R}: \quad \mathbb{R}^2 \to \mathbb{R}^2,$$

we have

$$e(\mathfrak{f}(\mathbf{R}^{-1}\cdot); \mathbf{R}\boldsymbol{p}) = e(\mathfrak{f}(\cdot); \boldsymbol{p}).$$

The reason is that the corresponding correlation matrices are orthogonally similar to each other. This property is meaningful in real life since we do not want the results to depend on the direction in which an image is displayed before us.

The L^2 *total energy* of a feature distribution $\mathfrak{f}$ is

$$\mathcal{E}(\mathfrak{f}) = \frac{1}{2}\int_\Omega e^2(\mathfrak{f};\boldsymbol{p})\,d\boldsymbol{p}.$$

Inspired by the affine case where Rudin and Osher [14], Rudin, Osher and Fatemi [15] first introduced the total variational (TV) model for image restoration, we define the *total variation* to be the L^1 total energy

$$\mathcal{E}^{\mathrm{TV}}(\mathfrak{f}) = \int_\Omega e(\mathfrak{f};\boldsymbol{p})\,d\boldsymbol{p}.$$

Let d denote the metric on M induced by its Riemannian structure. That is, for any two points $\mathfrak{f}$ and $\mathfrak{g} \in M$,

$$d(\mathfrak{f},\mathfrak{g}) = \inf_\Gamma \text{ length}(\Gamma),$$

allowing Γ to go over all piecewise smooth paths that link $\mathfrak{f}$ and $\mathfrak{g}$. (We shall also allow other kind of distances later. See the discrete model in the next subsection.)

Given a noisy feature distribution $\mathfrak{f}^{(0)}(\boldsymbol{p})$, the classical restoration model associated with the L^2 energy is to solve

$$\min_{\mathfrak{f}} \mathcal{E}(\mathfrak{f},\lambda),$$

where the fitted L^2 energy

$$\begin{aligned}\mathcal{E}(\mathfrak{f};\lambda) &= \mathcal{E}(\mathfrak{f}) + \frac{\lambda}{2}\int_\Omega d^2(\mathfrak{f}^{(0)},\mathfrak{f})\,d\boldsymbol{p}\\ &= \frac{1}{2}\int_\Omega e^2(\mathfrak{f};\boldsymbol{p})\,d\boldsymbol{p} + \frac{\lambda}{2}\int_\Omega d^2(\mathfrak{f}^{(0)},\mathfrak{f})\,d\boldsymbol{p}.\end{aligned}$$

As in the affine case, the first term regularizes the restoration solution and the second one is a fitting constraint.

For the TV restoration, the associated *fitted TV energy* is

$$\mathcal{E}^{\mathrm{TV}}(\mathfrak{f};\lambda) = \int_\Omega e(\mathfrak{f};\boldsymbol{p})\,d\boldsymbol{p} + \frac{\lambda}{2}\int_\Omega d^2(\mathfrak{f}^{(0)},\mathfrak{f})\,d\boldsymbol{p}.$$

The constant λ above is related to the Lagrange multiplier for the associated constrained optimization problem—

$$\text{Minimize} \quad \mathcal{E}(\mathfrak{f}) \ (\text{or } \mathcal{E}^{\mathrm{TV}}(\mathfrak{f})) \quad \text{subject to} \quad \frac{1}{|\Omega|}\int_\Omega d^2(\mathfrak{f}^{(0)},\mathfrak{f})d\boldsymbol{p} = \sigma^2,$$

where $|\Omega|$ is the area of Ω and σ the standard deviation of the noise. In the flat case, the connection between the constrained and unconstrained

formulations was discussed theoretically by Chambolle and Lions [2] and numerically by Blomgren and Chan [1]. In image analysis, λ is often conveniently fixed or chosen *a priori*, or estimated by the gradient-projection method (Rudin, Osher and Fatemi [15]).

To solve the above variational problems, we study the associated Euler-Lagrange equations. For the fitted L^2 energy $\mathcal{E}(\mathfrak{f};\lambda)$, the Euler-Lagrange equation can be shown to be

$$-\partial_x^*(\partial_x \mathfrak{f}) - \partial_y^*(\partial_y \mathfrak{f})) + \frac{\lambda}{2}\mathrm{grad}_{\mathfrak{f}} d^2(\mathfrak{f}^{(0)}, \mathfrak{f}) = 0,$$

where, at any pixel $\boldsymbol{p} = (x, y) \in \Omega$, $\mathrm{grad}_{\mathfrak{f}} d^2(\mathfrak{f}^{(0)}, \mathfrak{f})$ denotes the gradient vector of the scalar function $d^2(\mathfrak{f}^{(0)}(\boldsymbol{p}), \cdot)$ on M; and ∂_x^* and ∂_y^* denote covariant derivatives acting on vector fields defined on $\mathfrak{f}(\Omega)$. Specifically, let ∇ be the Levi-Civita connection (see Helgason [7]) on the feature Riemannian manifold M. Then the covariant derivatives ∂_x^* and ∂_y^* are given explicitly by

$$\partial_x^* = \nabla_{\partial_x \mathfrak{f}}, \qquad \partial_y^* = \nabla_{\partial_y \mathfrak{f}}.$$

They can legally act on any vector field X that is defined along $\mathfrak{f}(\Omega)$ due to the locality property of covariant derivatives. For instance, if M is a domain in $\mathbb{R}^m$, then

$$\partial_x^* X = \frac{\partial}{\partial x} X(\mathfrak{f}(x, y)).$$

Generally, when M is isometrically embedded in Euclidean space $\mathbb{R}^N$ ($N \geq m$), a vector field X on M is also one in $\mathbb{R}^N$. Since the flat derivative in $\mathbb{R}^N$

$$\frac{\partial}{\partial x} X(\mathfrak{f}(x, y))$$

usually overshoots from the tangent space of M, one needs an extra step of projection for the covariant derivative—

$$\partial_x^* X = \Pi_{\mathfrak{f}} \frac{\partial}{\partial x} X(\mathfrak{f}(x, y)), \tag{2}$$

where $\Pi_{\mathfrak{f}}$ is the orthogonal projection from $T_{\mathfrak{f}}\mathbb{R}^N$ onto the tangent space $T_{\mathfrak{f}}M$.

The Euler-Lagrange equation can be solved by the infinitesimal steepest descent method (or time-marching method). This leads to the *fitted diffusion equation*

$$\frac{\partial \mathfrak{f}}{\partial t} = \partial_x^*(\partial_x \mathfrak{f}) + \partial_y^*(\partial_y \mathfrak{f})) - \frac{\lambda}{2}\mathrm{grad}_{\mathfrak{f}} d^2(\mathfrak{f}^{(0)}, \mathfrak{f}),$$

with the Neumann adiabatic boundary condition plus the initial condition

$$\mathfrak{f}\,|_{t=0} = \mathfrak{f}^0,$$

where, $\mathfrak{f}^0$ can be the given raw feature distribution $\mathfrak{f}^{(0)}$, for convenience. All the terms in the diffusion equation are in the tangent space $T_{\mathfrak{f}(\boldsymbol{p})}M$. As the marching time increases, $\mathfrak{f}$ evolves to the equilibrium distribution of the functional.

Similarly, for the TV model, the Euler-Lagrange equation becomes

$$-\partial_x^*\left(\frac{\partial_x \mathfrak{f}}{e(\mathfrak{f};\boldsymbol{p})}\right) - \partial_y^*\left(\frac{\partial_y \mathfrak{f}}{e(\mathfrak{f};\boldsymbol{p})}\right) + \frac{\lambda}{2}\mathrm{grad}_{\mathfrak{f}} d^2(\mathfrak{f}^{(0)},\mathfrak{f}) = 0.$$

Steepest descent leads to a forced anisotropic diffusion equation

$$\frac{\partial \mathfrak{f}}{\partial t} = \partial_x^*\left(\frac{\partial_x \mathfrak{f}}{e(\mathfrak{f};\boldsymbol{p})}\right) + \partial_y^*\left(\frac{\partial_y \mathfrak{f}}{e(\mathfrak{f};\boldsymbol{p})}\right) - \frac{\lambda}{2}\mathrm{grad}_{\mathfrak{f}} d^2(\mathfrak{f}^{(0)},\mathfrak{f}).$$

Remark 1. Let us understand the first two terms on the right-hand side of the last equation. Assume the image domain is 1-dimensional (i.e. a real interval $[a,b]$) and parameterized by x. Then $\mathfrak{f}(x)$ is a *path* on the feature manifold, and minimization of

$$\int_a^b \|\mathfrak{f}_x\| dx$$

is to reduce the total path length. This viewpoint connects our problem to the much familiar object — *geodesic*. A geodesic is *self-parallel* in the well-known sense that if $\boldsymbol{\tau}$ is the **unit** tangent vector, then

$$\nabla_{\boldsymbol{\tau}}\boldsymbol{\tau} = 0.$$

This condition also means that it is locally a "straight" line on the manifold. Now it is easy to see that

$$\boldsymbol{\tau} = \frac{\mathfrak{f}_x}{\|\mathfrak{f}_x\|} = \frac{\mathfrak{f}_x}{e(\mathfrak{f};\boldsymbol{p})},$$

and the linearity of the Levi-Civita connection leads to

$$\left(\frac{d}{dx}\right)^*\left(\frac{\mathfrak{f}_x}{e(\mathfrak{f};\boldsymbol{p})}\right) = \|\mathfrak{f}_x\|\nabla_{\boldsymbol{\tau}}\boldsymbol{\tau}.$$

This gives us a better geometric sense about the two "TV-terms" appearing in the last two equations.

EXAMPLE 1. Feature manifold M embedded in $\mathbb{R}^N$.

When the feature manifold M can be isometrically embedded in Euclidean space $\mathbb{R}^N$ (such as the unit circle in $\mathbb{R}^2$ and the unit sphere $\mathbf{S}^m$ in $\mathbb{R}^{m+1}$), the Levi-Civita connection is simply the Euclidean derivatives followed by a

projection as pointed out above. In this case, a feature point $\mathfrak{f}$ can be seen as an ordinary vector in $\mathbb{R}^N$, and the definition of the strength function

$$e(\mathfrak{f};\boldsymbol{p}) = \sqrt{\|\mathfrak{f}_x(\boldsymbol{p})\|^2 + \|\mathfrak{f}_y(\boldsymbol{p})\|^2}$$

now only involves the ordinary differentiations of vectors and is exactly

$$\|\nabla\mathfrak{f}\| = \sqrt{\|\nabla\mathfrak{f}_1\|^2 + \|\nabla\mathfrak{f}_2\|^2 + \cdots + \|\nabla\mathfrak{f}_N\|^2},$$

if $\mathfrak{f} = (\mathfrak{f}_1, \mathfrak{f}_2, \cdots, \mathfrak{f}_N) \in \mathbb{R}^N$. Moreover, the diminishing flow fields F of the energy functionals can be computed explicitly. When $M = \mathbf{S}^m$ is embedded in $\mathbb{R}^{m+1}$, this work has been carried out recently in Tang, Sapiro and Caselles [19], and was also worked out independently by Osher [11]. For a general embedded feature manifold, for example, the diminishing flow field

$$F = \partial_x^*(\partial_x\mathfrak{f}) + \partial_y^*(\partial_y\mathfrak{f})$$

for the L^2 energy $\mathcal{E}$ is simply given by

$$F = \Pi_{\mathfrak{f}}\,\Delta\mathfrak{f}, \tag{3}$$

where, Δ is the flat Laplacian on Ω, and $\Pi_{\mathfrak{f}}$ is the orthogonal projection from $T_{\mathfrak{f}}\mathbb{R}^N$ onto $T_{\mathfrak{f}}M$. Spheres $\mathbf{S}^m$ in $\mathbb{R}^{m+1}$ are the most interesting to image processing. For the sphere, it was shown by Tang, Sapiro and Caselles [19] and Osher [11] that the diminishing flow for the L^2 energy is

$$F(\mathfrak{f}) = \Delta\mathfrak{f} + \|\nabla\mathfrak{f}\|^2\mathfrak{f}, \tag{4}$$

and for the TV energy,

$$F^{\mathrm{TV}}(\mathfrak{f}) = \nabla\cdot\left(\frac{\nabla\mathfrak{f}}{\|\nabla\mathfrak{f}\|}\right) + \|\nabla\mathfrak{f}\|\,\mathfrak{f}. \tag{5}$$

Osher [11] established these formulae by the method of Lagrange multiplier analysis, and Tang, Sapiro and Caselles [19] proved them by working out explicitly the variational problem. Here we give a simpler geometric proof based on the special form of the Levi-Civita connection (2).

Proof of Eq. (4) *and* (5) (see Figure 2). Take the TV diminishing flow Eq. (5) for example. For sphere $\mathbf{S}^m$, the projection at $\mathfrak{f}$ is

$$\Pi_{\mathfrak{f}}\mathbf{v} = \mathbf{v} - (\mathbf{v}\cdot\mathfrak{f})\mathfrak{f}.$$

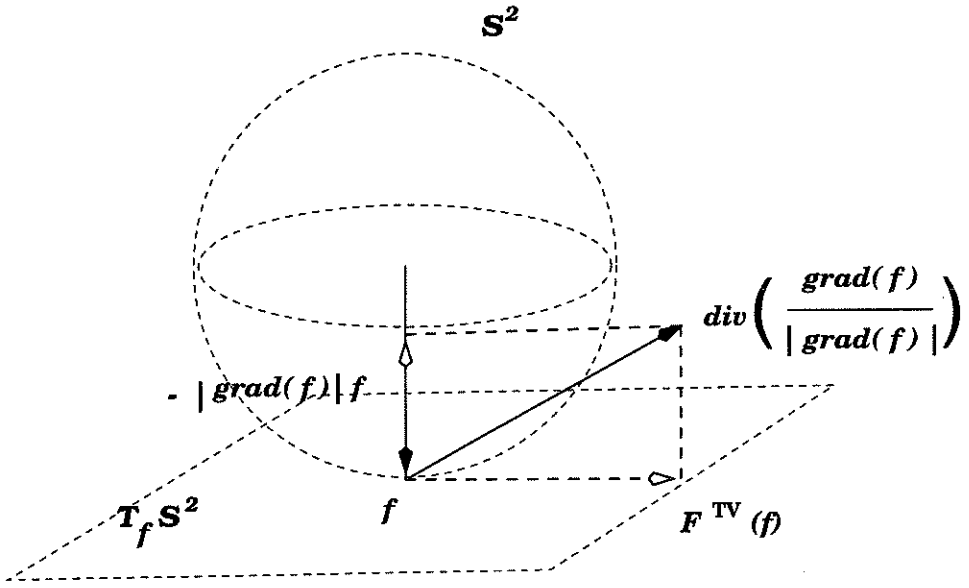

FIGURE 2. The TV diminishing flow.

Hence for any distribution $\mathfrak{f} : \Omega \to \mathbf{S}^m$,

$$\begin{aligned} F^{\mathrm{TV}}(\mathfrak{f}) &= \Pi_{\mathfrak{f}} \nabla \cdot \frac{\nabla \mathfrak{f}}{\|\nabla \mathfrak{f}\|} \\ &= \nabla \cdot \frac{\nabla \mathfrak{f}}{\|\nabla \mathfrak{f}\|} - \left(\nabla \cdot \frac{\nabla \mathfrak{f}}{\|\nabla \mathfrak{f}\|} \cdot \mathfrak{f} \right) \mathfrak{f} \\ &= \nabla \cdot \frac{\nabla \mathfrak{f}}{\|\nabla \mathfrak{f}\|} - \left(\nabla \cdot \frac{\mathfrak{f} \cdot \nabla \mathfrak{f}}{\|\nabla \mathfrak{f}\|} - \|\nabla \mathfrak{f}\| \right) \mathfrak{f} \\ &= \nabla \cdot \frac{\nabla \mathfrak{f}}{\|\nabla \mathfrak{f}\|} + \|\nabla \mathfrak{f}\| \, \mathfrak{f}, \end{aligned}$$

since $\mathfrak{f} \cdot \nabla \mathfrak{f} = \nabla \mathfrak{f}^2 / 2 = 0$. ■

This continuous model provides a universal language for restoring non-flat features. However, the model assumes that the readers have some knowledge of modern differential geometry, concepts such as the Levi-Civita connection. It is unavoidable in this continuous formulation since the variational formulation is global. Fortunately, there does exist a way that one can circumvent this technical difficulty. We now develop a simpler model for "digital" image domains, namely graphs, where differentiation of vector fields (i.e. Levi-Civita connection) can be avoided. Besides, the discrete models have their own computational advantages as discussed later in the paper.

2.2. **Models for discrete image domains.** Since numerically it is always necessary to lay down a finite discrete grid to approximate the continuous image domain Ω, it is practically convenient and important to study directly

non-flat features over a discrete grid, or more generally, a graph. This is the main task of this section. Perona took the first step in [12].

Let $G[\Omega_n]$ be an undirectional graph over a finite set of nodes Ω_n ("n" stands for "numerical"), and G is the dictionary of edges. We write $\alpha \sim \beta$ if nodes α and β are linked by an edge. For any node $\alpha \in \Omega_n$, denote by N_α all its neighbors

$$\{\beta \in \Omega_n : \ \beta \sim \alpha\}.$$

We start with the definition of the concept of LRD, which shall play an important role in our models. *A locally Riemannian distance (LRD) is a piecewise smooth continuous function* d_l

$$d_l : \quad M \times M \to \mathbb{R}^+ = [0, \infty),$$

such that

(1) For any $\mathfrak{f}, \mathfrak{g} \in M$, $d_l(\mathfrak{f}, \mathfrak{g}) = d_l(\mathfrak{g}, \mathfrak{f})$.
(2) $d_l(\mathfrak{f}, \mathfrak{g}) = 0$, *if and only if* $\mathfrak{f} = \mathfrak{g}$.
(3) $d_l(\mathfrak{f}, \mathfrak{g}) + d_l(\mathfrak{f}, \mathfrak{h}) \geq d_l(\mathfrak{g}, \mathfrak{h})$.
(4) $d_l(\mathfrak{f}, \mathfrak{g}) = d(\mathfrak{f}, \mathfrak{g}) + O(d^2(\mathfrak{f}, \mathfrak{g}))$ *as* $d(\mathfrak{f}, \mathfrak{g}) \to 0$.

Here d *is the induced Riemannian distance defined in the previous section.*

Remark 2. Conditions (1)–(3) are the basic ingredients of a distance, which authorize d_l to "legally" measure the amplitude of noise. Condition (4) restricts our attention to only those distances that are locally Riemannian. The motivation is as follows. Suppose the feature Riemannian manifold is locally flat (i.e. locally isometric to Euclidean space $\mathbb{R}^m$). Then condition (4) ensures that the "non-flat Laplacian" (see the remarks at the end of this section) is locally exactly the ordinary Laplacian. In this way, many classical analytical results may be transplanted easily.

EXAMPLE 2. Perona's distance.

Consider the unit circle $\mathbf{S}^1$ as the concerned feature manifold. For any two points $\mathfrak{f} = e^{i\theta_1}$ and $\mathfrak{g} = e^{i\theta_2}$, define

$$d_{\text{perona}}(\mathfrak{f}, \mathfrak{g}) = \sqrt{2(1 - \cos(\theta_1 - \theta_2))}.$$

Symmetry is immediate. Now if $d_{\text{perona}}(\mathfrak{f}, \mathfrak{g}) = 0$. Then $\cos(\theta_1 - \theta_2) = 1$, which implies that $\mathfrak{f} = \mathfrak{g}$. It is also clear that the leading term of $d_{\text{perona}}(\mathfrak{f}, \mathfrak{g})$ is the Riemannian distance when $\mathfrak{f}$ and $\mathfrak{g}$ are close. Therefore, to show d_{perona}

is a locally Riemannian distance, it suffices to verify the triangular relation. The proof follows. Formula

$$\sin(\theta_1 + \theta_2) = \sin\theta_1 \cos\theta_2 + \sin\theta_2 \cos\theta_1$$

leads to

$$|\sin(\theta_1 + \theta_2)| \leq |\sin(\theta_1)| + |\sin(\theta_2)|.$$

Noticing that

$$|\sin\theta| = \sqrt{\frac{1 - \cos 2\theta}{2}},$$

we obtain

$$\sqrt{1 - \cos(2\theta_1 + 2\theta_2)} \leq \sqrt{1 - \cos 2\theta_1} + \sqrt{1 - \cos 2\theta_2}.$$

Now that θ_1 and θ_2 are arbitrary, this last inequality holds even without the factor 2, which eventually leads to (by noticing that $\theta_3 - \theta_2 = (\theta_1 - \theta_2) + (\theta_3 - \theta_1)$)

$$\sqrt{1 - \cos(\theta_3 - \theta_2)} \leq \sqrt{1 - \cos(\theta_1 - \theta_2)} + \sqrt{1 - \cos(\theta_3 - \theta_1)}.$$

It is the triangular relation for d_{perona}!

It is the distance that Perona has proposed in his work [12] (though he did not formulate it in this way).

As an example of a non-LRD, consider

$$d_{\text{line}}(\mathfrak{f}, \mathfrak{g}) = \sqrt{\frac{1 - \cos(2(\theta_1 - \theta_2))}{2}}.$$

It is not a local distance for $\mathbf{S}^1$ since

$$d_{\text{line}}(1, -1) = 0.$$

For the real projective line $\mathbb{RP}^1$, however, d_{line} is an LRD.

For a given LRD d_l, we define the associated *strength* of a given feature distribution $\mathfrak{f}$ to be

$$e(\mathfrak{f}; \alpha) = [\sum_{\beta \in N_\alpha} d_l^2(\mathfrak{f}_\beta, \mathfrak{f}_\alpha)]^{\frac{1}{2}},$$

which generalizes the continuous strength function in Eq. (1) or $\|\nabla \mathfrak{f}\|$ when $\mathfrak{f}$ is a flat function. Next, we can define the l^2 total energy

$$\mathcal{E}(\mathfrak{f}) = \sum_{\alpha \in \Omega_n} \frac{1}{2} e^2(\mathfrak{f}; \alpha);$$

and the *total variation* (TV)

$$\mathcal{E}^{\mathrm{TV}}(\mathfrak{f}) = \sum_{\alpha\in\Omega_n} e(\mathfrak{f};\alpha).$$

For a given feature distribution $\mathfrak{f}^{(0)}$ and some weight constant λ, define the *fitted* l^2 total energy

$$\mathcal{E}(\mathfrak{f};\lambda) = \mathcal{E}(\mathfrak{f}) + \lambda \sum_{\alpha\in\Omega_n} \frac{1}{2} d_l^2(\mathfrak{f}_\alpha^{(0)}, \mathfrak{f}_\alpha),$$

and the *fitted* TV total energy

$$\mathcal{E}^{\mathrm{TV}}(\mathfrak{f};\lambda) = \mathcal{E}^{\mathrm{TV}}(\mathfrak{f}) + \lambda \sum_{\alpha\in\Omega_n} \frac{1}{2} d_l^2(\mathfrak{f}_\alpha^{(0)}, \mathfrak{f}_\alpha).$$

To establish the diffusion equations, we first compute the gradient fields for the total energies. For the l^2 total energy,

$$\frac{\partial \mathcal{E}(\mathfrak{f})}{\partial \mathfrak{f}_\alpha} = \sum_{\beta\in N_\alpha} \frac{\partial}{\partial \mathfrak{f}_\alpha} d_l^2(\mathfrak{f}_\alpha, \mathfrak{f}_\beta); \tag{6}$$

for the TV total energy,

$$\frac{\partial \mathcal{E}^{\mathrm{TV}}(\mathfrak{f})}{\partial \mathfrak{f}_\alpha} = \sum_{\beta\in N_\alpha} \left[\frac{\partial}{\partial \mathfrak{f}_\alpha} d_l^2(\mathfrak{f}_\alpha, \mathfrak{f}_\beta)\right] \left(\frac{1/e(\mathfrak{f};\alpha) + 1/e(\mathfrak{f};\beta)}{2}\right). \tag{7}$$

Hence the corresponding fitted diffusion equations are

$$\frac{d\mathfrak{f}_\alpha}{dt} = -\sum_{\beta\in N_\alpha} \frac{\partial}{\partial \mathfrak{f}_\alpha} d_l^2(\mathfrak{f}_\alpha, \mathfrak{f}_\beta) - \frac{\lambda}{2} \frac{\partial}{\partial \mathfrak{f}_\alpha} d_l^2(\mathfrak{f}_\alpha^{(0)}, \mathfrak{f}_\alpha); \tag{8}$$

$$\frac{d\mathfrak{f}_\alpha}{dt} = -\sum_{\beta\in N_\alpha} \left[\frac{\partial}{\partial \mathfrak{f}_\alpha} d_l^2(\mathfrak{f}_\alpha, \mathfrak{f}_\beta)\right] \left(\frac{1/e(\mathfrak{f};\alpha) + 1/e(\mathfrak{f};\beta)}{2}\right) - \frac{\lambda}{2} \frac{\partial}{\partial \mathfrak{f}_\alpha} d_l^2(\mathfrak{f}_\alpha^{(0)}, \mathfrak{f}_\alpha). \tag{9}$$

Notice that here our *starting point* is the locally Riemannian distance d_l. In his work on orientation diffusion, Perona [12] started with a choice of energy, which is originally inspired by physical considerations. For general non-flat feature manifolds, d_l (and the special case of $d_l = d$, the Riemannian distance) seems to be more fundamental.

As an example, we now apply the discrete models to the restoration of *chromaticity*. The continuous diffusion model without the fitting constraint was first discussed in Tang, Sapiro and Caselles [18]. Our discrete model has the following advantages: 1) unlike the continuous model, one does not need to choose a numerical discretization scheme (often complicated) for the spatial derivatives; 2) the discrete model rigorously diminishes the energy function; while in the continuous model, the numerical PDE's only

asymptotically (i.e. as the step size of time or space tends to zero) diminish the continuous energy functionals.

EXAMPLE 3. Restoration of the chromaticity feature on $\mathbf{S}^2$.

The 2-dimensional sphere $\mathbf{S}^2$ models the *chromaticity* feature of color images. Let $\mathbf{I}(\boldsymbol{p})$ denote the linear RGB vector value at pixel $\boldsymbol{p}$. Then $\mathbf{I}(\boldsymbol{p})$ can be perceptually separated into the *brightness component* $B = \|\mathbf{I}(\boldsymbol{p})\|$, and the *chromaticity component*

$$\mathfrak{f}(\boldsymbol{p}) = \frac{\mathbf{I}(\boldsymbol{p})}{B}.$$

$\mathfrak{f}(\boldsymbol{p})$ records the saturation degree of colors. We now apply the discrete models developed above to chromaticity restoration. Take the TV model (9) for example.

Embed $\mathbf{S}^2$ in $\mathbb{R}^3$ and take d_l to be the embedded Euclidean distance, i.e.

$$d_l(\mathfrak{f},\mathfrak{g}) := \|\mathfrak{f}-\mathfrak{g}\|_{\mathbb{R}^3} = \sqrt{(\mathfrak{f}-\mathfrak{g})^2}, \qquad \text{for any} \quad \mathfrak{f},\mathfrak{g} \in \mathbf{S}^2.$$

Let $\nabla_{\mathfrak{f}}$ denote the gradient for $\mathfrak{f} \in \mathbb{R}^3$ and $\partial/\partial\mathfrak{f}$ the gradient on $\mathbf{S}^2$. Then for any scalar function $G(\mathfrak{f})$ on $\mathbb{R}^3$,

$$\frac{\partial}{\partial \mathfrak{f}} G(\mathfrak{f}) = \Pi_{\mathfrak{f}} \nabla_{\mathfrak{f}} G(\mathfrak{f}).$$

Therefore, for any fixed feature point $\mathfrak{g} \in \mathbf{S}^2$,

$$\begin{aligned}\frac{\partial}{\partial \mathfrak{f}} d_l^2(\mathfrak{f},\mathfrak{g}) &= \Pi_{\mathfrak{f}} \nabla_{\mathfrak{f}} (\mathfrak{f}-\mathfrak{g})^2 \\ &= 2\Pi_{\mathfrak{f}}(\mathfrak{f}-\mathfrak{g}) = -2\Pi_{\mathfrak{f}}(\mathfrak{g}) \\ &= -2(\mathfrak{g} - (\mathfrak{g}\cdot\mathfrak{f})\mathfrak{f}).\end{aligned}$$

The TV model (9) is explicitly given by

$$\frac{d\mathfrak{f}_\alpha}{dt} = \sum_{\beta \in N_\alpha} \Pi_{\mathfrak{f}_\alpha}(\mathfrak{f}_\beta) \left(\frac{1}{e(\mathfrak{f};\alpha)} + \frac{1}{e(\mathfrak{f};\beta)} \right) + \lambda \Pi_{\mathfrak{f}_\alpha}(\mathfrak{f}_\alpha^{(0)}). \tag{10}$$

There are no spatial derivatives in the last equation! In computation, one only needs to take care of the time discretization, which is easy to handle with the *geodesic marching scheme* we propose in Section 4. It is easily seen that the fitted diminishing flow on the right-hand side is indeed on the tangent plane at $\mathfrak{f}_\alpha$, which is crucial to prevent $\mathfrak{f}_\alpha$ from wandering away from the feature manifold infinitesimally. This completes our example. We will return to it later in Section 4.

More generally, suppose the feature manifold M is isometrically embedded in $\mathbb{R}^N$. Let $\Pi_{\mathfrak{f}}$ denote the orthogonal projection from $T_{\mathfrak{f}}\mathbb{R}^N$ onto the tangent

plane $T_{\mathsf{f}}M$. If we choose d_l to be the straight line distance in $\mathbb{R}^N$ (but restricted on M), then the TV restoration equation is given by

$$\frac{d\mathsf{f}_\alpha}{dt} = \sum_{\beta\in N_\alpha} \Pi_{\mathsf{f}_\alpha}(\mathsf{f}_\beta - \mathsf{f}_\alpha)\left(\frac{1}{e(\mathsf{f};\alpha)} + \frac{1}{e(\mathsf{f};\beta)}\right) + \lambda\Pi_{\mathsf{f}_\alpha}(\mathsf{f}_\alpha^{(0)} - \mathsf{f}_\alpha). \tag{11}$$

Especially, if M is the level set of a non-singular (i.e. with non-zero gradient everywhere) function $\phi(\mathsf{f})$ in R^{m+1}: $\phi = 0$, then

$$\Pi_{\mathsf{f}}\mathbf{v} = \mathbf{v} - (\mathbf{v}\cdot\phi_{\mathsf{f}})\frac{\phi_{\mathsf{f}}}{\|\phi_{\mathsf{f}}\|^2},$$

where ϕ_{f} denotes the gradient of ϕ and $\mathbf{v} \in T_{\mathsf{f}}\mathbb{R}^{m+1}$. In the continuous case, Osher [11] first proposed to consider the level-set feature manifolds (and indeed, both $\mathbf{S}^1$ and $\mathbf{S}^2$ are level-set manifolds).

Some comments are in order now.

Remark 3.

1. *Non-flat Laplacian.* Suppose the feature manifold M is in fact a flat domain in $\mathbb{R}^m$. Take d_l to be the natural Euclidean distance

$$d(\mathsf{f}_\alpha, \mathsf{f}_\beta) = \|\mathsf{f}_\alpha - \mathsf{f}_\beta\|, \qquad \text{in } \mathbb{R}^m.$$

Then

$$\frac{\partial d_l^2(\mathsf{f}_\alpha, \mathsf{f}_\beta)}{\partial \mathsf{f}_\alpha} = 2(\mathsf{f}_\alpha - \mathsf{f}_\beta),$$

and

$$-\frac{\partial \mathcal{E}(\mathsf{f})}{\partial \mathsf{f}_\alpha} = 2\sum_{\beta\in N_\alpha}(\mathsf{f}_\beta - \mathsf{f}_\alpha),$$

which is the flat graph Laplacian acting on f (up to a constant). Graph Laplacians unlock many combinatorial secrets of a graph, a topic which has been recently well studied in the *spectral graph theory* (see Chung [4], for example).

Therefore, formula (6) can be viewed as a "non-flat" Laplacian.

2. *Nonlinear Adaptive Filter.* The right-hand side of formula (7) generalizes the curvature term frequently appearing in image analysis (Rudin and Osher [14]; Rudin, Osher and Fatemi [15]) —

$$\nabla\left(\frac{\nabla u}{|\nabla u|}\right),$$

which leads to the total variational anisotropic diffusion of a gray level image u, and is a special but the most commonly applied case of Perona and Malik's general concept of anisotropic diffusion [13]. For a general graph, formula (9) defines a non-linear adaptive diffusion. The feature

diffuses faster near a node where the strength e is smaller, or equivalently, where the feature distribution is smoother. It adaptively slows down when the distribution gets rougher and the local strength grows. This adaptivity is particularly useful for preserving local characteristic architectures of a given feature distribution, such as the singularities of fingerprint "waves," and feature "edges."

3. *The Graph Curvature Term.* Suppose that M is isometrically embedded in $\mathbb{R}^N$, and d_l is the restricted Euclidean distance. Denote by $K_\alpha(\mathsf{f})$ the flat flow field in $\mathbb{R}^N$

$$\sum_{\beta\in N_\alpha}(\mathsf{f}_\beta-\mathsf{f}_\alpha)\left(\frac{1}{e(\mathsf{f};\alpha)}+\frac{1}{e(\mathsf{f};\beta)}\right).$$

Then the TV restoration equation (11) reads

$$\frac{\partial \mathsf{f}_\alpha}{\partial t}=\Pi_{\mathsf{f}_\alpha}K_\alpha(\mathsf{f})+\lambda\Pi_{\mathsf{f}_\alpha}(\mathsf{f}_\alpha^{(0)}-\mathsf{f}_\alpha).$$

We now show that $K_\alpha(\mathsf{f})$ is exactly the graph version for the classical curvature term. Fix a node α. Let $\vec{e}$ denote any of the directed edges starting from α. Define the directional derivative of a flat feature f along $\vec{e}$ by

$$\frac{\partial \mathsf{f}}{\partial \vec{e}}=\frac{\partial \mathsf{f}}{\partial \vec{e}}\Big|_\alpha=\frac{\partial \mathsf{f}}{\partial \vec{e}}\Big|_\beta:=\mathsf{f}_\beta-\mathsf{f}_\alpha,$$

where β is the tail node of $\vec{e}$. Then the strength function is exactly

$$e(\mathsf{f};\alpha)=\left[\sum_{\vec{e}\sim\alpha}\left(\frac{\partial \mathsf{f}}{\partial \vec{e}}\right)^2\right]^{\frac{1}{2}},$$

which generalizes the magnitude of the ordinary gradient. Here $\vec{e}\sim\alpha$ includes all directed edges starting from α. Under this setup, it is easy to see that

$$K_\alpha(\mathsf{f})=\sum_{\vec{e}\sim\alpha}\frac{\partial}{\partial \vec{e}}\,\frac{1}{e}\,\frac{\partial \mathsf{f}}{\partial \vec{e}}\Big|_\alpha,$$

which is exactly in the form of graph divergence and gradient!

3. Restoration of Orientation and Alignment: Equations and Algorithms

In this section, based on the general mathematical models constructed above, we study the restoration of two classes of non-flat features frequently encountered in gray level image analysis, namely, the orientation feature and

the alignment feature. For the time being, we shall only discuss the discrete models.

3.1. **Restoration equations for orientation.** Orientation lives on the unit circle $\mathbf{S}^1$, which can be naturally parameterized by the angle parameter θ. Define the sawtooth function on $\mathbb{R}$:

$$D(x) = \begin{cases} |x| & |x| \leq \pi \\ 2\pi \text{ periodic extension} & \text{for general } x. \end{cases}$$

Then the Riemannian distance d is given by

$$d(e^{i\theta_1}, e^{\theta_2}) = D(\theta_1 - \theta_2),$$

or simply, $d(\theta_1, \theta_2) = D(\theta_1 - \theta_2)$, which measures the length of the shortest arc on $\mathbf{S}^1$ that links the two feature points. Recall that in the previous example, we discussed Perona's locally Riemannian distance

$$d_{\text{perona}}(\theta_1, \theta_2) = \sqrt{2(1 - \cos(\theta_1 - \theta_2))} = 2\left|\sin\frac{\theta_1 - \theta_2}{2}\right|.$$

The fact that it is an LRD is better seen geometrically through Figure 3, which shows that d_{perona} measures the Euclidean distance of the two feature points $e^{i\theta_1}$ and $e^{i\theta_2}$ in $\mathbb{R}^2$ (instead of measuring along $\mathbf{S}^1$).

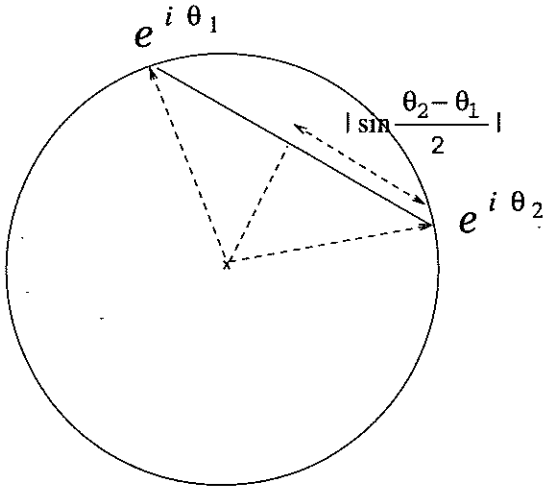

FIGURE 3. The Perona distance.

Remark 4. The Perona distance belongs to a wider class of *isometrically embedded distances (IED).* An IED, denoted by d_{ie}, is the result of an isometric embedding of $\mathbf{S}^1$ in $\mathbb{R}^2$ (see Figure 4). Precisely, this means to establish a smooth (C^2, say) injective mapping

$$F : \mathbf{S}^1 \to \mathbb{R}^2,$$

which satisfies the isometry condition

$$\|dF\|_{\mathbb{R}^2} = \|ds\|_{\mathbb{S}^1},$$

for any infinitesimal segment ds on the circle and its image dF. The IED associated with F is defined by

$$d_{\text{ie}}(\mathfrak{f}, \mathfrak{g}) = \|F(f) - F(g)\|_{\mathbb{R}^2}, \tag{12}$$

for any $\mathfrak{f}, \mathfrak{g}$ in $\mathbb{S}^1$. d_{ie} is an LRD because of the isometry condition. Such a distance is biased, or not rotationally invariant. It can be useful in occasions where we do know *a priori* that certain directions are more robust or important than the others.

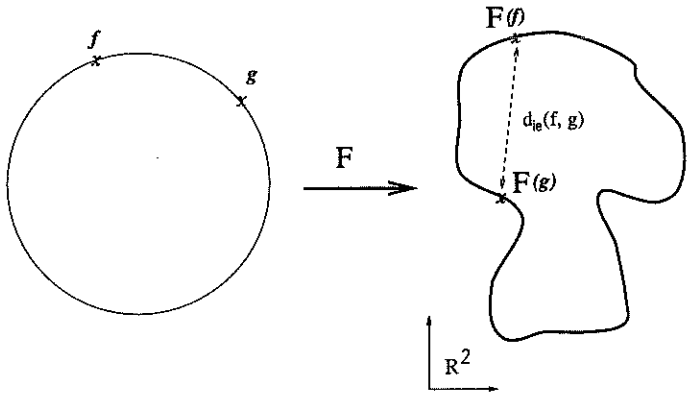

FIGURE 4. The isometrically embedded distance (IED).

One advantage of d_{perona} over D is the smoothness of d^2_{perona}, on which we take derivatives in the restoration formulae Eq.(6–9). Remarkably, d_{perona} is also rotationally invariant. In what follows, we shall only discuss this distance function.

We now derive the parametric restoration equations for d_{perona}. Parameterize the circle by $\theta \to e^{i\theta}$. Then the tangent vector $\partial/\partial\theta$ (or equivalently, $ie^{i\theta}$) is an orthonormal basis for the tangent space at each point. All vectors appearing in the diffusion equations are expressible using this basis. Assume $\mathfrak{f}_\alpha = e^{i\theta_\alpha}$. Then

$$\frac{\partial \mathfrak{f}_\alpha}{\partial t} = \frac{\partial \theta_\alpha}{\partial t}\frac{\partial}{\partial \theta_\alpha};$$

$$\frac{\partial d^2_{\text{perona}}(\mathfrak{f}_\alpha, \mathfrak{f}_\beta)}{\partial \mathfrak{f}_\alpha} = 2\frac{\partial(1-\cos(\theta_\alpha - \theta_\beta))}{\partial \theta_\alpha}\frac{\partial}{\partial \theta_\alpha} = 2\sin(\theta_\alpha - \theta_\beta)\frac{\partial}{\partial \theta_\alpha};$$

$$\frac{\partial d^2_{\text{perona}}(\mathfrak{f}_\alpha, \mathfrak{f}^{(0)}_\alpha)}{\partial \mathfrak{f}_\alpha} = 2\frac{\partial(1-\cos(\theta_\alpha - \theta^{(0)}_\alpha))}{\partial \theta_\alpha}\frac{\partial}{\partial \theta_\alpha} = 2\sin(\theta_\alpha - \theta^{(0)}_\alpha)\frac{\partial}{\partial \theta_\alpha}.$$

Then the fitted l^2 diffusion equation (8) becomes

$$\frac{d\theta_\alpha}{dt} = \sum_{\beta \in N_\alpha} 2\sin(\theta_\beta - \theta_\alpha) + \lambda \sin(\theta_\alpha^{(0)} - \theta_\alpha), \tag{13}$$

which is exactly the equation that Perona first studied in [12]. Similarly, the fitted TV diffusion equation is

$$\frac{d\theta_\alpha}{dt} = \sum_{\beta \in N_\alpha} \sin(\theta_\beta - \theta_\alpha)\left(\frac{1}{e(\theta;\alpha)} + \frac{1}{e(\theta;\beta)}\right) + \lambda \sin(\theta_\alpha^{(0)} - \theta_\alpha), \tag{14}$$

where, according to the previous section, the strength is

$$e(\theta;\alpha) = [\sum_{\beta \in N_\alpha} 2(1 - \cos(\theta_\beta - \theta_\alpha))]^{\frac{1}{2}}$$

Remark 5. Notice that the parametric equations (13) and (14) are equivalent to equations (8) and (9). This is one of the main advantages of the discrete (graph) model — it is less sensitive to the discontinuities in the raw data $\theta^{(0)}$. In fact, a graph is "blind" to discontinuities since there is no infinitesimal distance. For the continuous model, the parametric equations (i.e. on θ) are not equivalent to the corresponding embedded equations (i.e. on $\mathfrak{f}$, treated as a vector in $\mathbb{R}^2$), when the given raw data $\theta^{(0)}$ has discontinuities.

Equations (13) and (14) generalize classical flat equations. To see it, consider a smooth orientation distribution θ_α, that is,

$$\theta_\beta - \theta_\alpha = O(\epsilon),$$

for all $\beta \in N_\alpha$ and some small characteristic parameter ϵ. Then

$$\sum_{\beta \in N_\alpha} \sin(\theta_\beta - \theta_\alpha) = \sum_{\beta \in N_\alpha} (\theta_\beta - \theta_\alpha) + O(\epsilon^3).$$

The leading term is exactly the flat Laplacian on θ. This is generally true for any choice of locally Riemannian distances because of the last condition in the definition list.

Similarly, the strength function can be shown to be

$$e(\theta;\alpha) = [\sum_{\beta \in N_\alpha} (\theta_\beta - \theta_\alpha)^2]^{\frac{1}{2}}(1 + O(\epsilon^2)).$$

For example, for a typical internal cross node $\alpha = (i,j)$ in image processing, its neighbors are

$$\beta_u = (i, j-1), \quad \beta_d = (i, j+1), \quad \beta_l = (i-1, j), \quad \beta_r = (i+1, j).$$

Thus the leading term of the strength function is simply the finite difference approximation to the magnitude of the gradient

$$\sqrt{2}h\|\nabla\theta\| = \sqrt{2}h\left[\left(\frac{\partial\theta}{\partial x}\right)^2 + \left(\frac{\partial\theta}{\partial y}\right)^2\right]^{\frac{1}{2}},$$

where h is the discretization step size.

3.2. **Restoration equations for alignment.** For alignment, direction of a line segment is unimportant (see Figure 5). This amounts to saying that the feature manifold is the real projective line $\mathbb{RP}^1$.

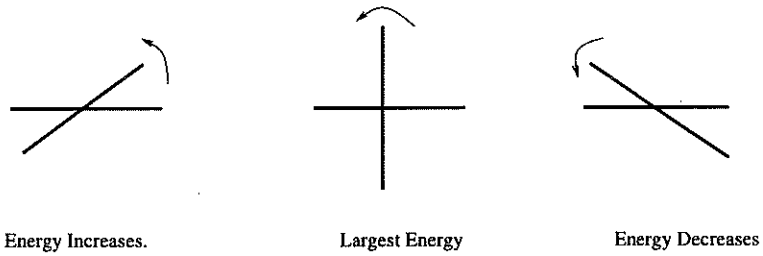

FIGURE 5. The alignment feature.

A line segment in the direction of θ is denoted by l_θ. Then l_θ and $l_{\theta+\pi}$ denote the same line segment. A one-to-one correspondence between $\mathbb{RP}^1$ and the unit circle $\mathbf{S}^1$ can be established by representing each line segment l_θ by a point $e^{i2\theta}$. This embedding technique has already been practiced in image processing (see Perona [12], and Granlund and Knuttson [6]).

The Riemannian metric on $\mathbb{RP}^1$ is given by

$$d(l_{\theta_1}, l_{\theta_2}) = |\theta_1 - \theta_2|,$$

whenever $|\theta_1 - \theta_2| < \pi/2$.

The line distance mentioned in Section 2 is given by

$$d_{\text{line}}(l_{\theta_1}, l_{\theta_2}) = \sqrt{\frac{1 - \cos(2(\theta_1 - \theta_2))}{2}} = |\sin(\theta_1 - \theta_2)|.$$

d_{line} is in fact the pull-back of d_{perona} under the circle representation of $\mathbb{RP}^1$. If fact, any LRD d_l on the circle (an IED, especially) can be pulled back to $\mathbb{RP}^1$ by defining

$$d_l^{\mathbb{RP}^1}(l_{\theta_1}, l_{\theta_2}) = \frac{1}{2}d_l(e^{i2\theta_1}, e^{i2\theta_2}).$$

In what follows, we shall only work with d_{line}.

Similar to the orientation diffusion, we obtain the fitted l^2 diffusion equations for d_{line}

$$\frac{d\theta_\alpha}{dt} = \sum_{\beta \in N_\alpha} \sin(2\theta_\beta - 2\theta_\alpha) + \frac{\lambda}{2} \sin(2\theta_\alpha^{(0)} - 2\theta_\alpha). \tag{15}$$

The fitted TV diffusion equation is

$$\frac{d\theta_\alpha}{dt} = \sum_{\beta \in N_\alpha} \frac{1}{2} \sin(2\theta_\beta - 2\theta_\alpha) \left(\frac{1}{e(\theta;\alpha)} + \frac{1}{e(\theta;\beta)} \right) + \frac{\lambda}{2} \sin(2\theta_\alpha^{(0)} - 2\theta_\alpha), \tag{16}$$

where, the local strength is given by

$$e(\theta;\alpha) = \left[\sum_{\beta \in N_\alpha} \frac{1 - \cos(2\theta_\beta - 2\theta_\alpha)}{2} \right]^{\frac{1}{2}}.$$

We shall only discuss these two equations in the following. Our later numerical experiment shows that the alignment restoration model is particularly useful for fingerprint processing (also see Perona [12]).

3.3. **Algorithms.** The parametric restoration equations (13–16) are all non-linear. The good news from the discrete model is that the steady equations are all algebraic. Various existing solvers for non-linear algebraic equations can thus make their contributions, though careful modifications should be made.

In this section, we discuss three numerical approaches. We first study the l^2 restoration equations in details, and then point out how to make suitable adjustments for the TV restoration equations. Let Ω_n denote the graph image domain. For the purpose of image processing, one can be happy with the standard rectangular grid, where each (internal) node $\alpha = (i, j)$ has four neighboring nodes $\beta = (i, j \pm 1), (i \pm 1, j)$ such that $\beta \sim \alpha$. Considering that our models and algorithms may be applicable even out of the scope of image processing, in the following, we still state the algorithms using the abstract notation for nodes: $\alpha, \beta, \cdots$.

Recall from the previous subsections that we try to solve directly the system of steady restoration equations

$$2 \sum_{\beta \sim \alpha} \sin(\theta_\beta - \theta_\alpha) + \lambda \sin(\phi_\alpha - \theta_\alpha) = 0, \qquad \alpha \in \Omega_n. \tag{17}$$

Here ϕ is the given raw (noisy) angle distribution. The associated total energy is given by

$$\mathcal{E}(\theta;\lambda) = 2\sum_{\beta\sim\gamma}(1-\cos(\theta_\beta-\theta_\gamma)) + \lambda\sum_{\alpha\in\Omega_n}(1-\cos(\phi_\alpha-\theta_\alpha)),$$

where $\beta\sim\gamma$ denotes all distinct edges.

3.3.1. *The non-linear Gauss-Seidel method.* We first propose a nonlinear Gauss-Seidel method for solving it numerically. It proceeds as follows. Choose an initial guess θ^0 ($\theta^0=\phi$, say) and assign a linear ordering for all nodes:

$$\cdots<\beta<\alpha<\gamma<\cdots.$$

For convenience, for each node α, define

$$N_\alpha^- := \{\beta\in N_\alpha : \beta<\alpha\}, \qquad N_\alpha^+ := \{\beta\in N_\alpha : \beta>\alpha\}.$$

Let k denote the global updating clock and α a certain node. Suppose that in the beginning of step $k|\alpha$, we have the following data available

$$\theta_\beta^k : \beta<\alpha; \quad \theta_\alpha^{k-1}; \quad \theta_\beta^{k-1} : \beta>\alpha.$$

We try to update θ_α^{k-1} to θ_α^k by the end of step $k|\alpha$.

Set $x=\theta_\alpha^k$. According to the restoration equation (17), we demand that

$$2\sum_{\beta\in N_\alpha^+}\sin(\theta_\beta^{k-1}-x) + 2\sum_{\beta\in N_\alpha^-}\sin(\theta_\beta^k-x) + \lambda\sin(\phi_\alpha-\theta_\alpha^{k-1}) = 0. \tag{18}$$

Define

$$C = 2\sum_{\beta\in N_\alpha^+}\cos(\theta_\beta^{k-1}) + 2\sum_{\beta\in N_\alpha^-}\cos(\theta_\beta^k),$$
$$S = 2\sum_{\beta\in N_\alpha^+}\sin(\theta_\beta^{k-1}) + 2\sum_{\beta\in N_\alpha^-}\sin(\theta_\beta^k).$$

Then

$$S\cos x - C\sin x = -\lambda\sin(\phi_\alpha-\theta_\alpha^{k-1}), \tag{19}$$

Define

$$r=\sqrt{C^2+S^2}, \quad \omega=\angle(C,S).$$

Here $\angle(a,b)$ is the angle in the direction of (a,b). Then Eq. (19) simplifies to

$$\sin(\omega-x) = -\frac{\lambda}{r}\sin(\phi_\alpha-\theta_\alpha^{k-1}),$$

which gives (mod 2π)

$$x = \omega + \arcsin\left[\frac{\lambda}{r}\sin(\phi_\alpha - \theta_\alpha^{k-1})\right],$$

or

$$x = \omega - \arcsin\left[\frac{\lambda}{r}\sin(\phi_\alpha - \theta_\alpha^{k-1})\right] - \pi.$$

Here we take the principle range of arcsin to be $[-\pi/2, \pi/2]$. To determine which one we should choose, consider the updated orientation distribution

$$\theta' = (\theta_\beta^k : \beta < \alpha, \quad x, \quad \theta_\beta^{k-1} : \beta > \alpha).$$

We require that the total l^2 energy

$$\mathcal{E}(\theta') = 2\sum_{\beta\sim\gamma}(1 - \cos(\theta'_\beta - \theta'_\gamma))$$

can be as small as possible. This implies that x should be chosen so that

$$2\sum_{\beta\in N_\alpha^-}(1 - \cos(\theta_\beta^k - x)) + 2\sum_{\beta\in N_\alpha^+}(1 - \cos(\theta_\beta^{k-1} - x))$$

is the smallest, or equivalently,

$$2\sum_{\beta\in N_\alpha^-}\cos(\theta_\beta^k - x) + 2\sum_{\beta\in N_\alpha^+}\cos(\theta_\beta^{k-1} - x)$$

can be the biggest. This amounts to maximizing

$$C\cos x + S\sin x$$

under the constraint (19). Hence x is given by

$$\theta_\alpha^k = x = \omega + \arcsin\left[\frac{\lambda}{r}\sin(\phi_\alpha - \theta_\alpha^{k-1})\right]. \tag{20}$$

This completes the $k|\alpha$–step of the nonlinear Gauss-Seidel updating.

Remark 6. Careful readers may have noticed that in the updating equation (18), we do not replace θ_α^{k-1} by x in the fitting term. This is one important lesson we learned from numerical experiments. If one permits that replacement, then the algorithm tends to giving a constant solution.

Gauss-Seidel is simple, local and reliable, but with a relatively slow convergence compared with the other schemes we propose below.

3.3.2. *Linearized iteration method (LIM).* Let $\operatorname{sinc}(x)$ denote the function $\sin x/x$. Then Eq. (17) can be written as

$$\sum_{\beta\sim\alpha}(\theta_\beta-\theta_\alpha)2\operatorname{sinc}(\theta_\beta-\theta_\alpha)+\lambda(\phi_\alpha-\theta_\alpha)\operatorname{sinc}(\phi_\alpha-\theta_\alpha)=0. \tag{21}$$

The *linearized iteration method* generates the iteration as follows. Take a guess θ^0 (taken to be ϕ, say) for the initial step 0. Suppose in the beginning of step k, θ^{k-1} is available. We update θ^{k-1} to θ^k by solving a linear algebraic system. Define

$$a_{\beta\gamma}=2\operatorname{sinc}(\theta_\beta^{k-1}-\theta_\gamma^{k-1}),\quad \text{if}\quad \beta\sim\gamma; \tag{22}$$

$$c_\alpha=\lambda\operatorname{sinc}(\phi_\alpha-\theta_\alpha^{k-1}),\quad \text{for all}\quad \alpha\in\Omega_n. \tag{23}$$

Based on Eq (21), for each α, we require that

$$\sum_{\beta\sim\alpha}a_{\alpha\beta}(\theta_\beta^k-\theta_\alpha^k)+c_\alpha(\phi_\alpha-\theta_\alpha^k)=0.$$

This leads to an ordinary linear system

$$A\theta^k=\mathbf{b}$$

with $\mathbf{b}=(c_\alpha\phi_\alpha:\alpha\in\Omega_n)^{\mathrm{T}}$ and $A=(A_{\alpha\beta}:\alpha,\beta\in\Omega_n)$:

$$A_{\alpha\alpha}=c_\alpha+\sum_{\beta\sim\alpha}a_{\alpha\beta},$$

$$A_{\alpha\beta}=\begin{cases}-a_{\alpha\beta}, & \text{if}\quad \beta\in N_\alpha,\\ 0, & \text{else.}\end{cases}$$

Under this linearization, the non-linear updating can be achieved using any numerical linear algebra solver. LIM converges geometrically, faster than the nonlinear Gauss-Seidel scheme above.

3.3.3. *The time marching method.* In this subsection, we approach the steady distribution by time marching according to the discrete diffusion equation.

Recall that the fitted diffusion model using l^2 energy is given by

$$\frac{d\theta_\alpha}{dt}=2\sum_{\beta\sim\alpha}\sin(\theta_\beta-\theta_\alpha)+\lambda\sin(\phi_\alpha-\theta_\alpha),\qquad \alpha\in\Omega_n. \tag{24}$$

Take a marching step size h and set

$$\theta_\alpha^k=\theta_\alpha(kh).$$

Then the system of differential equations (24) can be solved using various numerical methods.

Introduce the non-flat Perona Laplacian symbol Δ'

$$\Delta'\psi_\alpha = \sum_{\beta\sim\alpha} \sin(\psi_\beta - \psi_\alpha).$$

The simplest single-step method that ignores the non-linearity of the diffusion equation is the Euler method

$$\theta_\alpha^{k+1} = \theta_\alpha^k + h2\Delta'\theta_\alpha^k + h\lambda \sin(\phi_\alpha - \theta_\alpha^k),$$

for each step k. In practice, this is implemented by choosing two small positive free parameters a and b:

$$\theta_\alpha^{k+1} = \theta_\alpha^k + a\Delta'\theta_\alpha^k + b \sin(\phi_\alpha - \theta_\alpha^k).$$

Numerically, Euler method suffers from low accuracy and slow convergence. The major advantage is its simplicity and easy implementation, especially for such a non-linear problem. It can also be easily parallelized. Moreover, from the practical point of view, the exact steady distribution is not so unique and critical for human vision. Numerical experiments show that *near-steady distributions* are usually already good enough for human vision and recognition of characteristic patterns residing in a given raw (noisy) feature distribution. This is the good news Euler method.

From the numerical point of view, one can also try other high order single-step methods such as Runge-Kutta 2. But both methods require a small time step to ensure stability. In order to achieve faster marching by allowing larger step size h, we discuss the implicit trapezoidal scheme (or *second-order Adams-Moulton method*, see Golub and Ortega [5]) combined with the idea of linearization in the preceding subsection. The major advantage of such a scheme is its stability and permission for a large marching step size.

Recall that for a system of equations like

$$\frac{d\theta_\alpha}{dt} = f_\alpha(\theta), \quad \alpha \in \Omega,$$

the implicit trapezoidal schemes at step k is given by

$$\theta_\alpha^{k+1} = \theta_\alpha^k + \frac{h}{2}[f_\alpha(\theta^k) + f_\alpha(\theta^{k+1})].$$

Applying it to our problem, we have

$$\theta_\alpha^{k+1} = \theta_\alpha^k + [h\Delta'\theta_\alpha^k + \frac{h\lambda}{2}\sin(\phi_\alpha - \theta_\alpha^k)] + [h\Delta'\theta_\alpha^{k+1} + \frac{h\lambda}{2}\sin(\phi_\alpha - \theta_\alpha^{k+1})],$$

or

$$\theta_\alpha^{k+1} - [h\Delta'\theta_\alpha^{k+1} + \frac{h\lambda}{2}\sin(\phi_\alpha - \theta_\alpha^{k+1})] = \theta_\alpha^k + [h\Delta'\theta_\alpha^k + \frac{h\lambda}{2}\sin(\phi_\alpha - \theta_\alpha^k)].$$

Next, similar to the previous method, we linearize the left hand side by the substitution

$$X^{k+1}\,\mathrm{sinc}X^k \Leftarrow \sin X^{k+1}, \tag{25}$$

for any expression X^{k+1} involving the current unknown θ^{k+1}. Define

$$a_{\beta\gamma} = h\,\mathrm{sinc}(\theta^k_\beta - \theta^k_\gamma), \quad \text{if} \quad \beta \sim \gamma;$$

$$c_\alpha = \frac{h\lambda}{2}\,\mathrm{sinc}(\phi_\alpha - \theta^k_\alpha), \qquad \alpha \in \Omega_n;$$

$$b_\alpha = \theta^k_\alpha + h\Delta'\theta^k_\alpha + \frac{h\lambda}{2}\sin(\phi_\alpha - \theta^k_\alpha) + c_\alpha\phi_\alpha, \quad \alpha \in \Omega_n.$$

Then the above implicit scheme leads to a linear system of equations $A\theta^{k+1} = \mathbf{b}$ with $\mathbf{b} = (b_\alpha : \alpha \in \Omega_n)^{\mathrm{T}}$ and $A = (A_{\alpha\beta} : \alpha, \beta \in \Omega_n)$:

$$A_{\alpha\alpha} = c_\alpha + \sum_{\beta\sim\alpha} a_{\alpha\beta},$$

$$A_{\alpha\beta} = \begin{cases} -a_{\alpha\beta}, & \text{if} \quad \beta \in N_\alpha, \\ 0, & \text{else.} \end{cases}$$

This linear system is bandlimited, symmetric and positive definite for sufficiently small h and can thus be solved easily using the standard numerical linear algebra solvers.

3.3.4. *Algorithms for the TV restoration equations.* With the algorithms ready for the fitted l^2 energy, it is easy to modify each one separately for the fitted TV energy. We summarize them briefly below.

The non-linear Gauss-Seidel method. The modification is done by computing the local strengths e's explicitly without involving x. Therefore, Eq. (19) holds with a new pair of constants (C, S). The discussion there is also valid. For the flat case, the idea appeared in Chan and Vese [3].

The linearized iteration method (LIM). The sin–sinc linearization technique extends to the TV model by modifying the constant $a_{\alpha\beta}$ to

$$a_{\alpha\beta} = \mathrm{sinc}(\theta^{k-1}_\beta - \theta^{k-1}_\alpha)\left[\frac{1}{e(\theta^{k-1};\alpha)} + \frac{1}{e(\theta^{k-1};\beta)}\right].$$

This generalizes the algorithm for the classical flat case

$$0 = -\nabla\cdot\left(\frac{\nabla u}{|\nabla u|}\right) + \lambda(u - u^0),$$

where Vogel and Oman [20] proposed the linear iteration scheme

$$0 = -\nabla\cdot\left(\frac{\nabla u^k}{|\nabla u^{k-1}|}\right) + \lambda(u^k - u^0).$$

The time marching method. The modification is done by modifying (25) to

$$X^{k+1}\frac{\operatorname{sinc}X^k}{e(\theta^k;\bullet)} \Leftarrow \frac{\sin X^{k+1}}{e(\theta^{k+1};\bullet)}.$$

Remark 7.

1. In practice, the local strength function e is usually modified to

$$\dot{e}_a = \sqrt{a^2 + e^2},$$

for some small parameter a ($a = 10^{-4}$, for example). This modification avoids the zero denominator in all the above TV formulae; On the other hand, it produces an intermediate effect between the TV energy and the l^2 energy. If a is large enough, then this modified TV model gets closer to the l^2 energy. This regularization technique is commonly practiced in the classical literature (see Marquina and Osher [8], for example).

2. For the time marching method, as Marquina and Osher [8] recently proposed for the continuous case, instead of solving the direct infinitesimal steepest descent equation (16)

$$\frac{\partial\theta_\alpha}{\partial t} = \sum_{\beta\in N_\alpha}\frac{1}{2}\sin(2\theta_\beta - 2\theta_\alpha)\left(\frac{1}{e_a(\theta;\alpha)} + \frac{1}{e_a(\theta;\beta)}\right) + \frac{\lambda}{2}\sin(2\theta_\alpha^{(0)} - 2\theta_\alpha),$$

one can solve its preconditioned version

$$\frac{\partial\theta_\alpha}{\partial t} = \sum_{\beta\in N_\alpha}\frac{1}{2}\sin(2\theta_\beta - 2\theta_\alpha)\left(1 + \frac{e_a(\theta;\alpha)}{e_a(\theta;\beta)}\right) + \frac{\lambda}{2}e_a(\theta;\alpha)\sin(2\theta_\alpha^{(0)} - 2\theta_\alpha).$$

Since the "diagonal" multiplier $e_a(\theta;\alpha)$ is non-zero, the preconditioned equation still marches to the original steady solution. However, from the numerical point of view, this modification dramatically improves the marching speed and numerical stability.

4. Restoration of Chromaticity: Algorithms and Discussions

In this section, we discuss the restoration models for the chromaticity feature living on the unit sphere S^2 and the related computational issues. We shall discuss the computational issues only for the discrete models developed in Section 2. For the work on the continuous pure diffusion model (i.e. without the fitting term), see Tang, Sapiro and Caselles [18, 19].

4.1. **The continuous models.** Let $\mathfrak{f} : \Omega \to \mathbf{S}^2$ be a given raw noisy chromaticity distribution. In Section 2, we have discussed the continuous restoration model corresponding to the L^2 and TV regularization energies:

$$\mathfrak{f}_t = \Delta \mathfrak{f} + \|\nabla \mathfrak{f}\|^2 \mathfrak{f} + \lambda \Pi_{\mathfrak{f}}(\mathfrak{f}^{(0)} - \mathfrak{f}), \qquad (L^2)$$

$$\mathfrak{f}_t = \nabla \cdot \left(\frac{\nabla \mathfrak{f}}{\|\nabla \mathfrak{f}\|} \right) + \|\nabla \mathfrak{f}\| \mathfrak{f} + \lambda \Pi_{\mathfrak{f}}(\mathfrak{f}^{(0)} - \mathfrak{f}). \qquad (\mathrm{TV})$$

Here we have chosen the distance function for the fitting term to be the restricted Euclidean distance in $\mathbb{R}^3$ (not the geodesic distance on $\mathbf{S}^2$). The orthogonal projection is given by

$$\Pi_{\mathfrak{f}} : T_{\mathfrak{f}}\mathbb{R}^3 \to T_{\mathfrak{f}}\mathbf{S}^2, \qquad \Pi_{\mathfrak{f}}\mathbf{v} := \mathbf{v} - (\mathbf{v} \cdot \mathfrak{f})\mathfrak{f}.$$

Since $\Pi_{\mathfrak{f}}\mathfrak{f} = 0$, the two equations simplify to

$$\mathfrak{f}_t = \Delta \mathfrak{f} + \|\nabla \mathfrak{f}\|^2 \mathfrak{f} + \lambda \Pi_{\mathfrak{f}} \mathfrak{f}^{(0)}, \tag{26}$$

$$\mathfrak{f}_t = \nabla \cdot \left(\frac{\nabla \mathfrak{f}}{\|\nabla \mathfrak{f}\|} \right) + \|\nabla \mathfrak{f}\| \mathfrak{f} + \lambda \Pi_{\mathfrak{f}} \mathfrak{f}^{(0)}. \tag{27}$$

One can start the evolutions with the initial condition

$$\mathfrak{f}\big|_{t=0} = \mathfrak{f}^{(0)},$$

for instance.

There are two practical problems with these continuous models. Take the TV model for instance. First, if a numerical scheme is sought for Eq. (27), we must take a special care for the diminishing flow term

$$F^{\mathrm{TV}}(\mathfrak{f}) \doteq \nabla \cdot \left(\frac{\nabla \mathfrak{f}}{\|\nabla \mathfrak{f}\|} \right) + \|\nabla \mathfrak{f}\| \, \mathfrak{f},$$

to make sure that the numerical $F^{\mathrm{TV}}(\mathfrak{f})$ does lie on the tangent plane $T_{\mathfrak{f}}S^2$. If the distribution $\mathfrak{f}$ is smooth, this tangency condition causes no serious problem since it is in any case satisfied to the leading order. But our task is to denoise and restore feature distributions with poor regularities! We have to permit discontinuities and singularities. In this realistic situation, typically, the numerical diminishing flow F^{TV} overshoots the tangent plane $T_{\mathfrak{f}}\mathbf{S}^2$. This can of course be fixed by artificially projecting it back onto the tangent plane. But by doing so, we introduce a new degree of uncertainty regarding the diminishing effect of the algorithm.

Besides this overshooting problem, another issue is that we lack a rigorous explanation for why the numerical evolution truly minimizes the continuous energy functional $\mathcal{E}^{\mathrm{TV}}(\mathfrak{f}; \lambda)$ on $\mathbf{S}^2$. The gap between continuous energy

functionals and numerical PDE's cannot be smoothed away easily because of the discretization error.

These are the intrinsic problems with the continuous model. They disppear in the discrete formulation, since the latter starts with discrete energies and results in algebraic equations (on $\mathbf{S}^2$), which need no numerical discretization for spatial derivatives.

4.2. **The discrete models.** Let $\mathfrak{f}^{(0)} : \Omega_n \to \mathbf{S}^2$ be the given raw discrete chromaticity distribution. By taking d_l to be the restricted Euclidean distance, we have established in Section 2 the discrete restoration equations corresponding to the l^2 and TV energies:

$$\frac{d\mathfrak{f}_\alpha}{dt} = 2\sum_{\beta\in N_\alpha} \Pi_{\mathfrak{f}_\alpha}\mathfrak{f}_\beta + \lambda\Pi_{\mathfrak{f}_\alpha}\mathfrak{f}_\alpha^{(0)}, \qquad \alpha\in\Omega_n; \qquad (l^2)$$

$$\frac{d\mathfrak{f}_\alpha}{dt} = 2\sum_{\beta\in N_\alpha} \Pi_{\mathfrak{f}_\alpha}\mathfrak{f}_\beta\left(\frac{1}{e(\mathfrak{f};\alpha)}+\frac{1}{e(\mathfrak{f};\beta)}\right) + \lambda\Pi_{\mathfrak{f}_\alpha}\mathfrak{f}_\alpha^{(0)}, \qquad \alpha\in\Omega_n. \qquad (\mathrm{TV})$$

Define the weight

$$w_{\alpha\beta}(\mathfrak{f}) = \frac{1}{e(\mathfrak{f};\alpha)}+\frac{1}{e(\mathfrak{f};\beta)},$$

for any $\beta\sim\alpha$. Then $w_{\alpha\beta}=w_{\beta\alpha}$. Rewrite the above equations using the linearity of the projection operator

$$\frac{d\mathfrak{f}_\alpha}{dt} = \Pi_{\mathfrak{f}_\alpha}\left(\sum_{\beta\in N_\alpha} 2\mathfrak{f}_\beta + \lambda\mathfrak{f}_\alpha^{(0)}\right) =: F_\alpha(\mathfrak{f}), \qquad \alpha\in\Omega_n;$$

$$\frac{d\mathfrak{f}_\alpha}{dt} = \Pi_{\mathfrak{f}_\alpha}\left(\sum_{\beta\in N_\alpha} w_{\alpha\beta}\mathfrak{f}_\beta + \lambda\mathfrak{f}_\alpha^{(0)}\right) =: F_\alpha^{\mathrm{TV}}(\mathfrak{f}), \qquad \alpha\in\Omega_n.$$

The steady restoration equations are thus given by

$$F_\alpha(\mathfrak{f}) = \Pi_{\mathfrak{f}_\alpha}\left(\sum_{\beta\in N_\alpha} 2\mathfrak{f}_\beta + \lambda\mathfrak{f}_\alpha^{(0)}\right) = 0, \qquad \alpha\in\Omega_n; \qquad (28)$$

$$F_\alpha^{\mathrm{TV}}(\mathfrak{f}) = \Pi_{\mathfrak{f}_\alpha}\left(\sum_{\beta\in N_\alpha} w_{\alpha\beta}\mathfrak{f}_\beta + \lambda\mathfrak{f}_\alpha^{(0)}\right) = 0, \qquad \alpha\in\Omega_n. \qquad (29)$$

Here 0 denotes the zero vector in $T_{\mathfrak{f}}\mathbf{S}^2$. These are "algebraic" equations on the chromaticity sphere $\mathbf{S}^2$.

Now we discuss the numerical implementations of both the marching equations and the steady algebraic equations.

4.2.1. *Geodesic marching for the discrete evolution equation.* Take the TV model

$$\frac{df_\alpha}{dt} = F_\alpha^{\mathrm{TV}}(\mathfrak{f}), \qquad \alpha \in \Omega_n,$$

for example. Set $\mathfrak{f}_\alpha^n = \mathfrak{f}_\alpha(n\Delta t)$. The classical Euler method is a flat marching

$$\tilde{\mathfrak{f}}_\alpha^n = \mathfrak{f}_\alpha^{n-1} + \Delta t F_\alpha^{\mathrm{TV}}(\mathfrak{f}^{n-1}),$$

followed by a projection (see Figure 6)

$$\mathfrak{f}_\alpha^n = \frac{\tilde{\mathfrak{f}}_\alpha^n}{\|\tilde{\mathfrak{f}}_\alpha^n\|}.$$

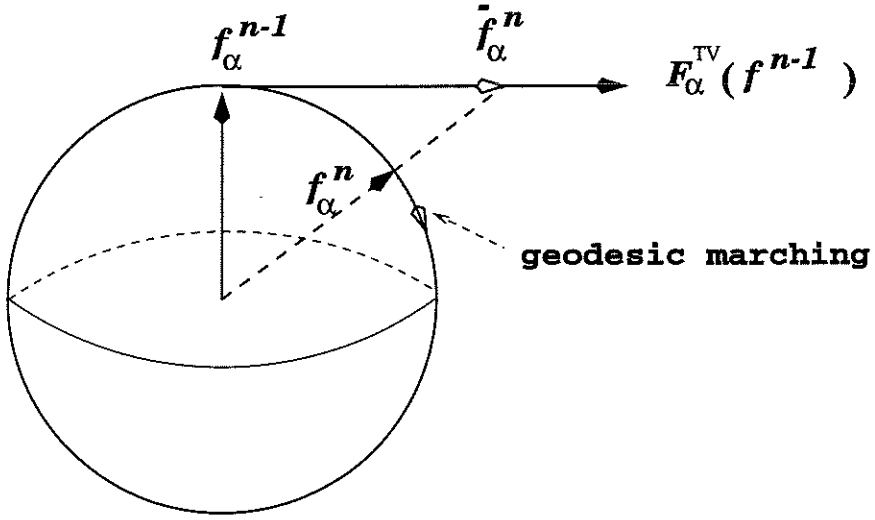

FIGURE 6. A flat marching followed by a projection, and the geodesic marching.

To compress these two steps into a single one, we propose the *geodesic marching*:

(30)

$$\mathfrak{f}_\alpha^n = \cos\left[\Delta t\|F_\alpha^{\mathrm{TV}}(\mathfrak{f}^{n-1})\|\right]\ \mathfrak{f}_\alpha^{n-1} + \sin\left[\Delta t\|F_\alpha^{\mathrm{TV}}(\mathfrak{f}^{n-1})\|\right]\ \frac{F_\alpha^{\mathrm{TV}}(\mathfrak{f}^{n-1})}{\|F_\alpha^{\mathrm{TV}}(\mathfrak{f}^{n-1})\|}.$$

It means that $\mathfrak{f}_\alpha^{n-1}$ marches for distance $\Delta t\|F_\alpha^{\mathrm{TV}}(\mathfrak{f}^{n-1})\|$ along the big circle (i.e. the geodesic) determined by $\mathfrak{f}_\alpha^{n-1}$ and the flow vector $F_\alpha^{\mathrm{TV}}(\mathfrak{f}^{n-1})$. Therefore, the marching always stays on the sphere. It is easy to see that for small Δt, the above two-step classical scheme is equivalent to the geodesic marching up to the leading order. Precisely, the latter always marches a distance $O(\Delta t^2)$ farther than the former (see Figure 6). The geodesic marching scheme is intrinsic for the geometry of the feature manifold, and thus has the most general meaning for numerical restorations of non-flat features.

4.2.2. *Fixed-point iteration, filtering, and the Convex Cone Theorem.* Again take the system of TV equations

$$F_\alpha^{\mathrm{TV}}(\mathsf{f}) = \Pi_{\mathsf{f}_\alpha}\left(\sum_{\beta\in N_\alpha} w_{\alpha\beta}\mathsf{f}_\beta + \lambda \mathsf{f}_\alpha^{(0)}\right) = 0, \quad \alpha \in \Omega_n,$$

for example. It can be interpreted as: for the optimal restoration f, at each node α,

$$\textstyle\sum_{\beta\in N_\alpha} w_{\alpha\beta}\,\mathsf{f}_\beta + \lambda \mathsf{f}_\alpha^{(0)} \quad \text{is} \quad \textit{parallel} \quad \text{to} \quad \mathsf{f}_\alpha.$$

This explanation inspires the iteration scheme:

$$\tilde{\mathsf{f}}_\alpha^n = \sum_{\beta\in N_\alpha} w_{\alpha\beta}\,(\mathsf{f}^{n-1})\mathsf{f}_\beta^{n-1} + \lambda \mathsf{f}_\alpha^{(0)};$$
$$\mathsf{f}_\alpha^n = \tilde{\mathsf{f}}_\alpha^n / \|\tilde{\mathsf{f}}_\alpha^n\|.$$

Or, to stabilize the algorithm (and even lead to new insights), define

$$h_{\alpha\beta}(\mathsf{f}) = \frac{w_{\alpha\beta}(\mathsf{f})}{\sum_{\gamma\in N_\alpha} w_{\alpha\gamma}(\mathsf{f}) + \lambda}, \qquad h_{\alpha\alpha}(\mathsf{f}) = \frac{\lambda}{\sum_{\gamma\in N_\alpha} w_{\alpha\gamma}(\mathsf{f}) + \lambda}.$$

We consider h as an adaptive lowpass filter. Then the above algorithm stabilizes to

$$\tilde{\mathsf{f}}_\alpha^n = \sum_{\beta\in N_\alpha} h_{\alpha\beta}^{n-1}\mathsf{f}_\beta^{n-1} + + h_{\alpha\alpha}^{n-1}\mathsf{f}_\alpha^{(0)}; \tag{31}$$

$$\mathsf{f}_\alpha^n = \tilde{\mathsf{f}}_\alpha^n / \|\tilde{\mathsf{f}}_\alpha^n\|. \tag{32}$$

Here $h_{\alpha\beta}^{n-1} = h_{\alpha\beta}(\mathsf{f}^{n-1})$.

This filtering formulation immediately rewards us with one important property. A region Γ on $\mathbf{S}^2$ is said to be a *convex cone* if its spanning cone $\tilde{\Gamma}$ in $\mathbb{R}^3$

$$\tilde{\Gamma} = \mathbb{R}^+ \times \Gamma = \{r \times \mathbf{v} : r \geq 0, \mathbf{v} \in \Gamma\}$$

is convex.

Theorem 1 (Convex Cone Theorem). *Suppose the raw data $\mathsf{f}_\alpha^{(0)}$'s belong to a convex cone Γ. If we start the above iterations using $\mathsf{f}^{(0)}$ as the initial data, then for any step n, all f_α^n's belong to Γ.*

The simple proof is left for our readers. Since RGB values are nonnegative, the chromaticity feature lives on the positive cone (one eighth of the sphere), which is convex. Therefore, the above fixed point iteration scheme keeps new feature points from wandering away from the positive cone on the sphere. This another advantage of the discrete model.

The above fixed-point iteration is similar to the Gauss-Jacobi method in numerical linear algebra. In the same fashion, we can construct the spherical Gauss-Seidel-like iteration. Assign a linear ordering for all the pixel nodes:

$$\cdots < \beta < \alpha < \gamma < \cdots .$$

Let n denote the Gauss-Seidel clock and α a pixel under consideration. Suppose in the beginning of step $n|\alpha$, we have available the data

$$(\mathbf{f}_\beta^n \,:\, \beta < \alpha; \quad \mathbf{f}_\alpha^{n-1}; \quad \mathbf{f}_\beta^{n-1} \,:\, \beta > \alpha). \tag{33}$$

Within step $n|\alpha$, we update $\mathbf{f}_\alpha^{n-1}$ to $\mathbf{f}_\alpha^n$ by

$$\tilde{\mathbf{f}}_\alpha^n = \sum_{\beta \in N_\alpha^-} h'_{\alpha\beta} \mathbf{f}_\beta^n + \sum_{\beta \in N_\alpha^+} h'_{\alpha\beta} \mathbf{f}_\beta^{n-1} + h'_{\alpha\alpha} \mathbf{f}_\alpha^{(0)};$$
$$\mathbf{f}_\alpha^n = \tilde{\mathbf{f}}_\alpha^n / \|\tilde{\mathbf{f}}_\alpha^n\|.$$

Here the lowpass filter coefficients $(h'_{\alpha\beta})$ are defined similar to the previous algorithm, but for the distribution given by (33). The Convex Cone Theorem still holds for this non-linear Gauss-Seidel method.

5. Numerical Examples and Applications

5.1. **Denoising by the l^2 model: 1-D example (Figure 7).** Figure 7 displays the restoration effect of the l^2 model for a noisy orientation distribution. The first subplot shows the angle distributions, and the rest three are the corresponding orientation distributions. In the first subplot, the solid line represents the clean angle distribution. Then uniform random noise is added to produce the dotted dash line. The restoration with the *a priori* Lagrange multiplier $\lambda = 0.01$ is given by the dash line. Small λ leads to smooth interiors. However, the resulting weak effect of the fitting term may smear orientation "edges," as clearly observable from the figure.

5.2. **Restoration by the TV model: 1-D example (Figure 8).** Similar to Figure 7, we apply the TV model (14) to restore the noisy orientation distribution. We have chosen $\lambda = 1$, and the small parameter a discussed in Section 3.3.4 to be 0.0001.

Compared to the l^2 model, the TV model clearly works better. The restoration of orientation "edges" is more faithful. This is generally true when the underlying clean feature distribution undergoes abrupt changes.

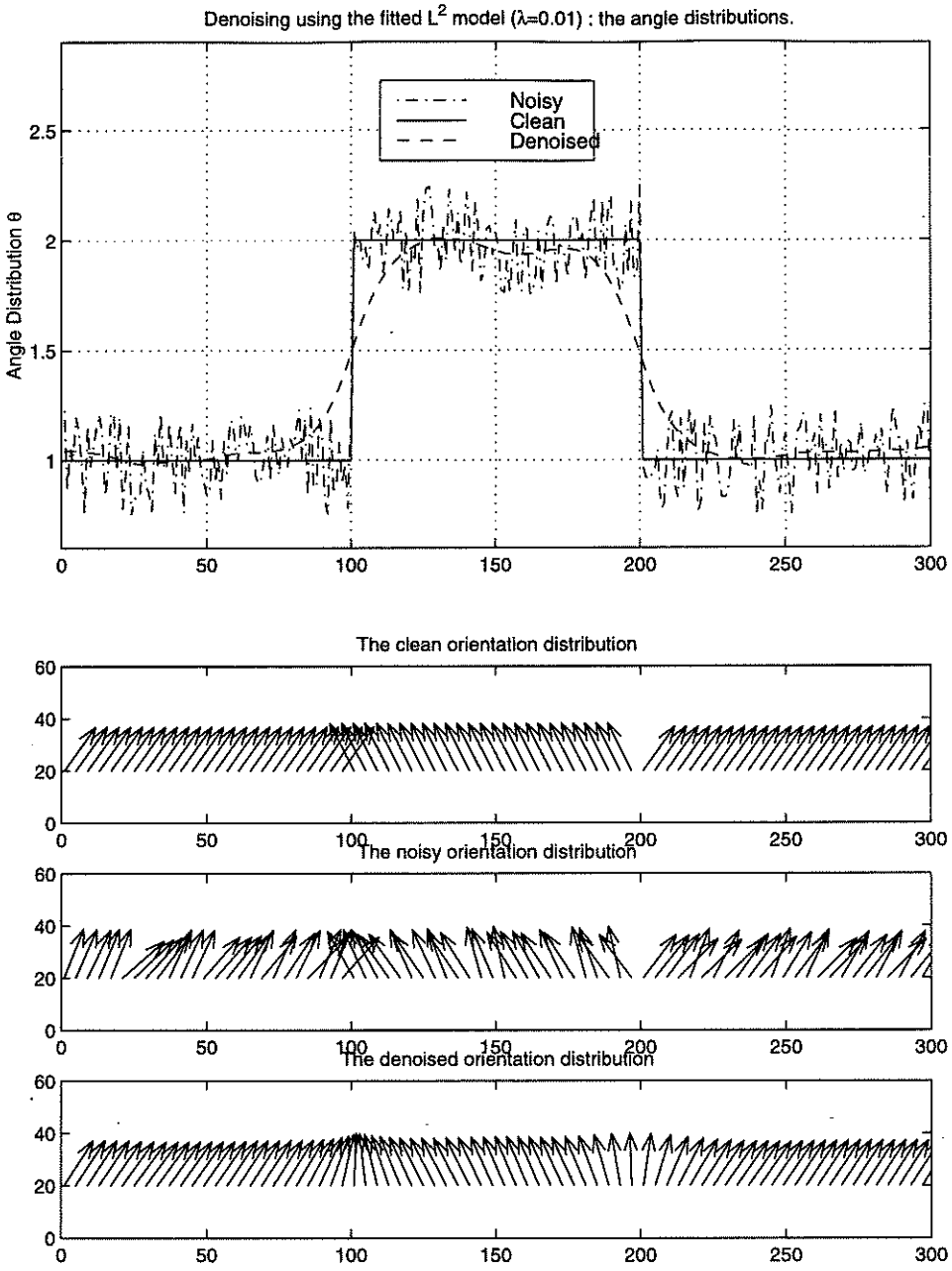

FIGURE 7. Denoising using the l^2 model Eq. (13). The orientation edges are smoothed out. (See Section 5.1.)

5.3. **Restoring wave lines of a fingerprint (Figure 9).** For fingerprints, the alignment pattern of wave lines is the most important feature. Therefore, instead of Eq. (13), we use Eq. (15), which yields better results.

We first apply the enhancement and diffusion techniques to the original raw fingerprint image to obtain the alignment distribution. This was done by computing the gradient flow of the processed image. We then have the

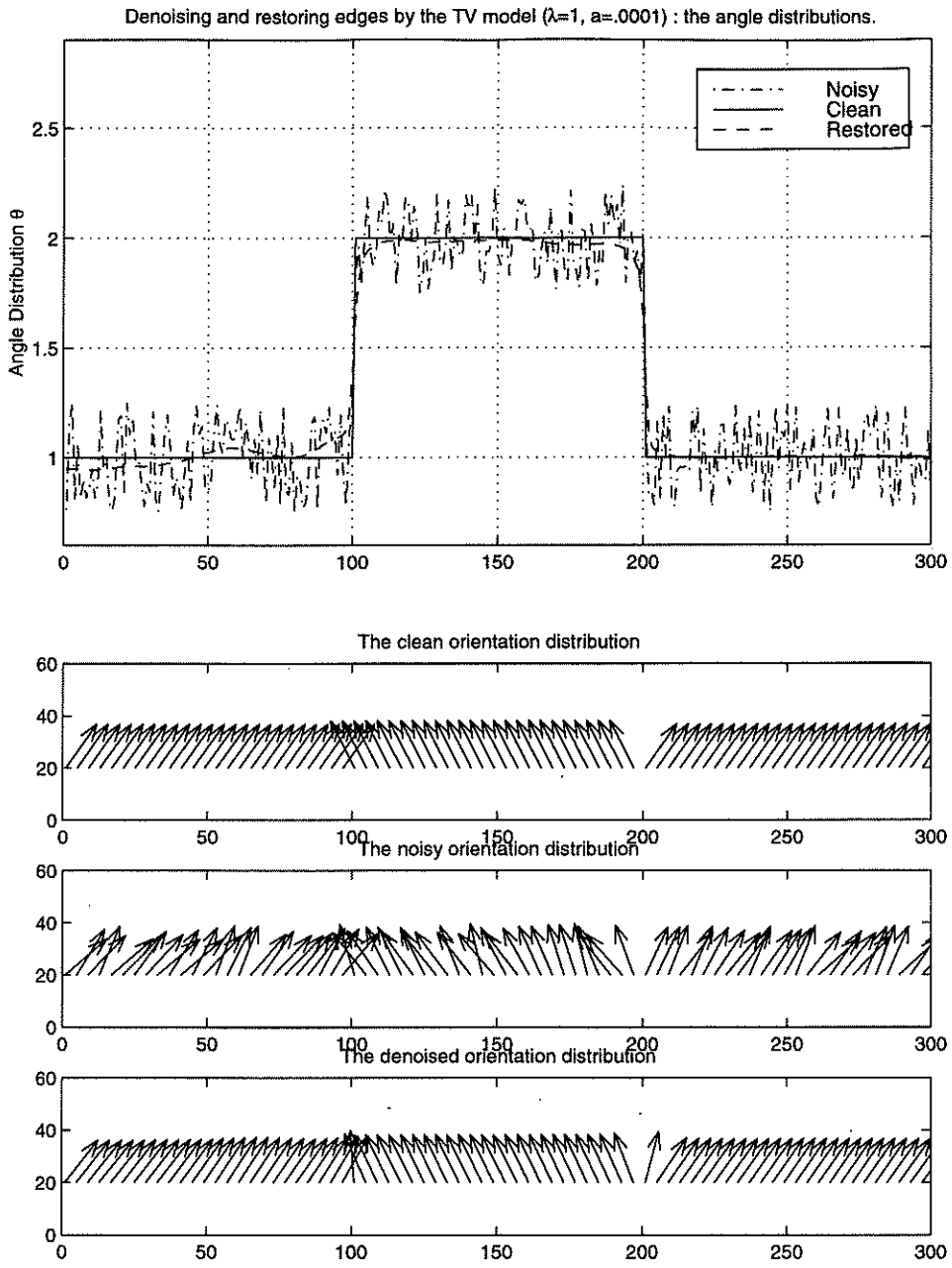

FIGURE 8. Denoising and restoration of edges through the TV model Eq. (14). The orientation edges are faithfully restored. (See Section 5.2.)

"heads of the arrows" removed. The alignment pattern has been superimposed on the fingerprint image in the top subplot Figure 9. Notice that it is indeed a noisy distribution regardless our smoothing effort.

In the bottom subplot, we apply the line model Eq. (15) with $\lambda = 0.1$ to denoise the above distribution. The result is very successful. The alignment

pattern coincides very well with the fingerprint except on a couple of local regions where the original image is not clear or damaged.

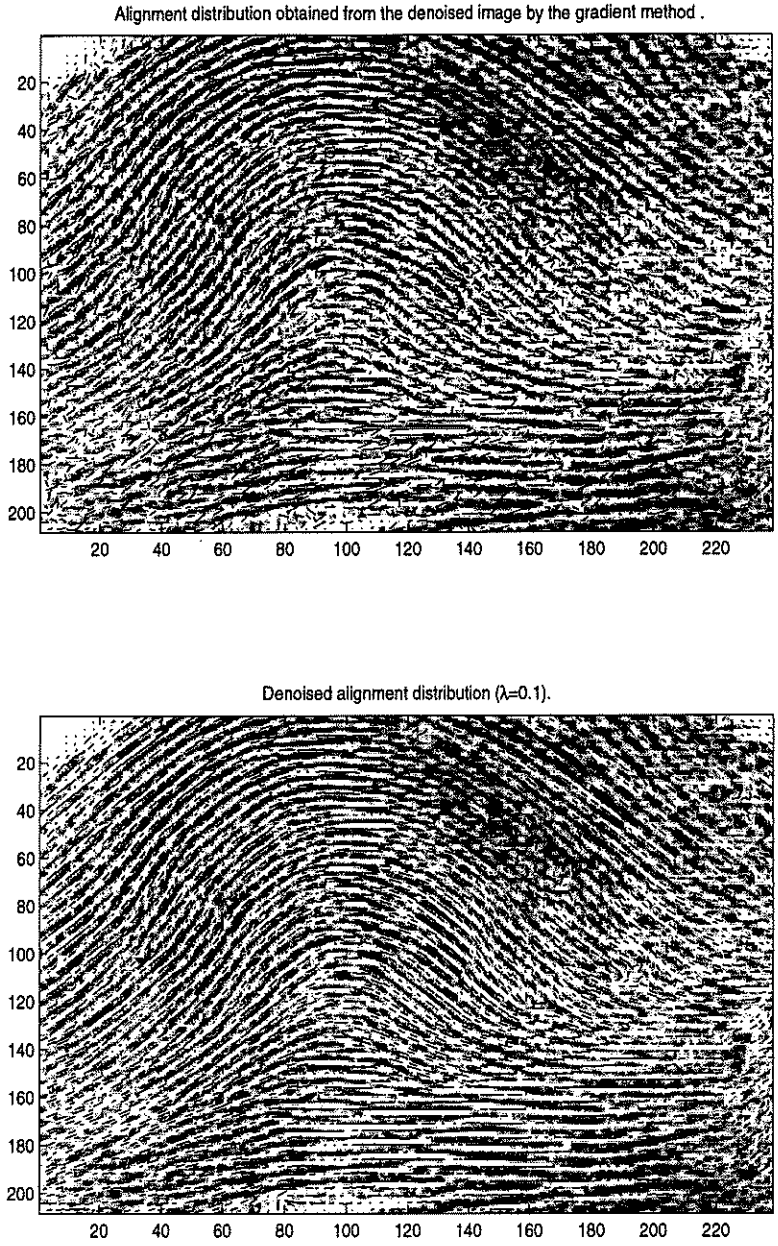

FIGURE 9. Restoring the alignment feature of fingerprints by the l^2 model Eq. (15). (See Section 5.3.)

5.4. **Restoration of optical flows by the l^2 model (Figure 10).** In Figure 10, we show an application of the l^2 model for denoising optical flows.

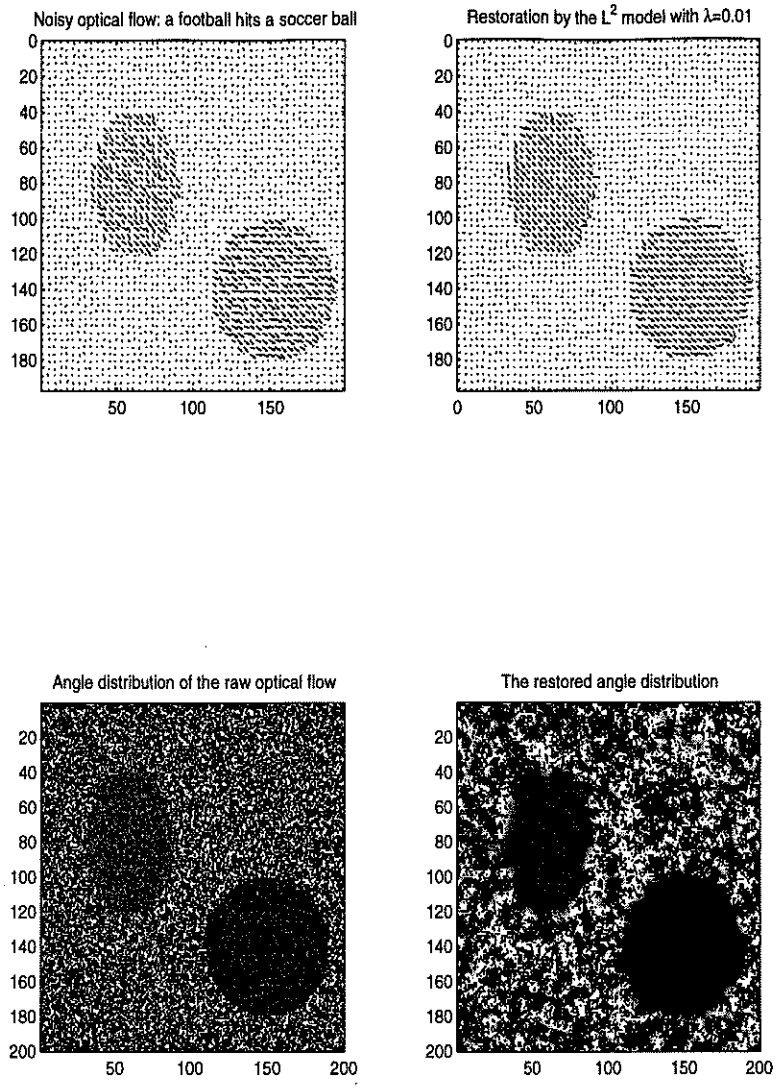

FIGURE 10. Restoration of optical flows by the l^2 model Eq. (13). (See Section 5.4.)

The upper left subplot shows the noisy optical flow of an imagined video recording the spinless rigid motion of a football and a soccer ball. The background is noisy.

The upper right subplot shows the restored optical flow obtained by the l^2 model. We have chosen $\lambda = 0.01$. The length of each arrow is not changed. The two subplots at bottom plot the corresponding angle distributions.

Observe that the motion near the boundaries is smeared due to the smoothing effect of the l^2 energy.

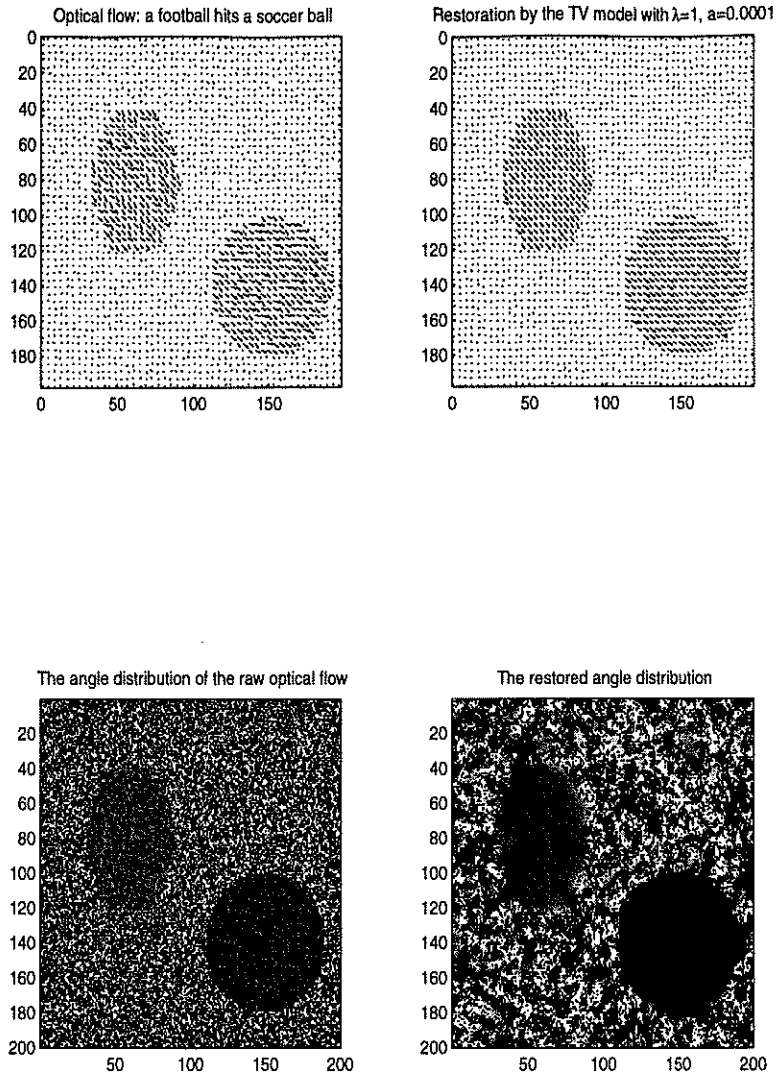

FIGURE 11. Restoration of optical flows by the TV model Eq. (14). (See Section 5.5.)

5.5. **Restoration of optical flows by the TV model (Figure 11).** Similar to Figure 10, in Figure 11, we have tested the TV model with $\lambda = 1$ and $a = 0.0001$. TV model restores the boundaries better than the l^2 model.

5.6. **Restoration of chromaticity (Figure 12 and 13).** In Figure 12, we apply the discrete chromaticity restoration model for a clown image with Gaussian noise (the top image). We first separate the vector RGB image $\mathbf{I}$ into the brightness component $B = \|\mathbf{I}\|$, and the chromaticity component $\mathfrak{f}^{(0)} = \|\mathbf{I}\|/B$. Then we apply the fixed-point TV filtering algorithm (31) to $\mathfrak{f}^{(0)}$ and get the optimal restoration $\mathfrak{f}$. Finally, we use the original brightness

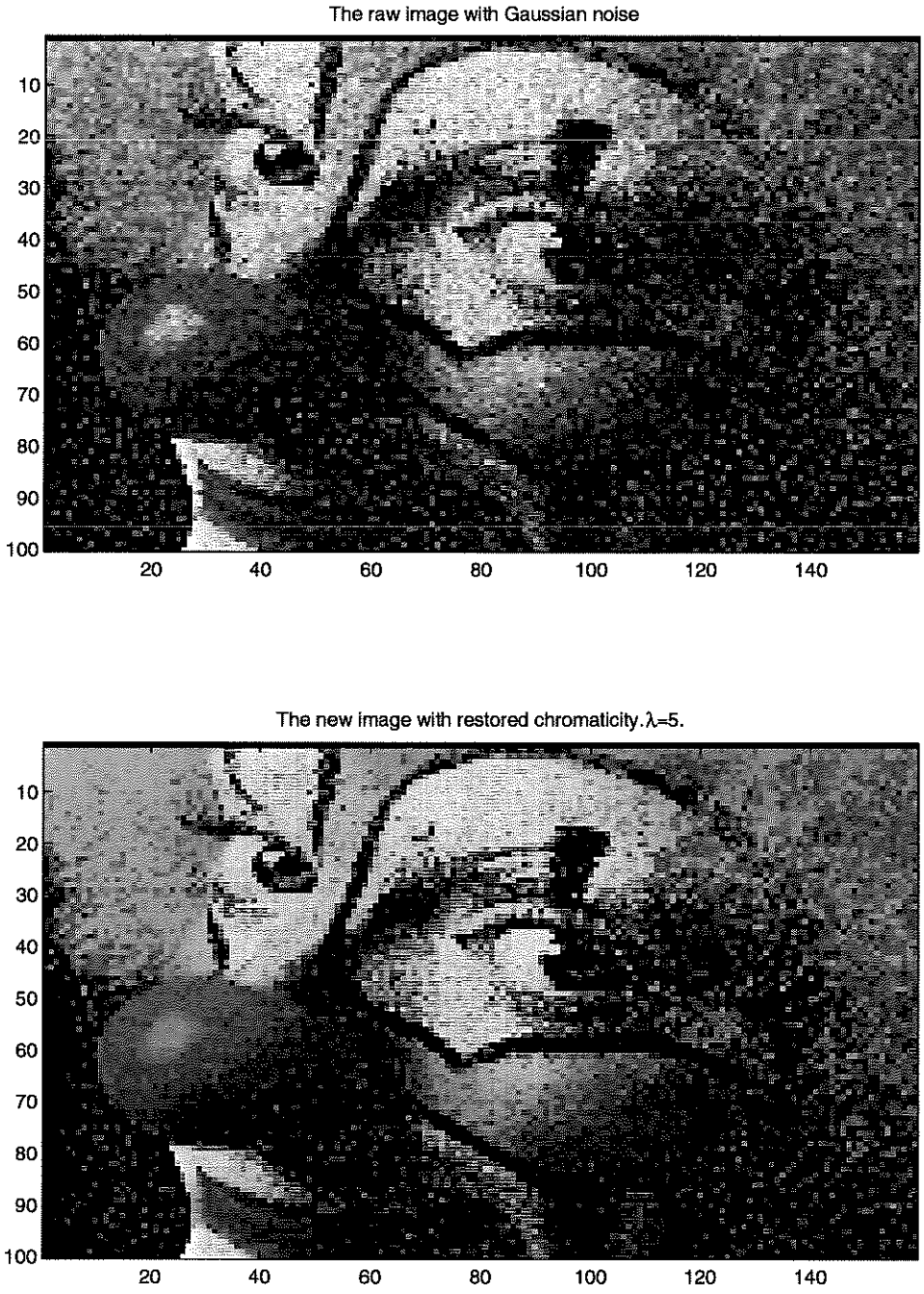

FIGURE 12. Restoration of chromaticity from Gaussian noise by the discrete TV model and spherical filtering Eq. (31). The brightness component remains unchanged. (See Section 5.6. The raw image is from the courtesey of G. Sapiro.)

value B to assemble the new image $\mathbf{I}_{\mathbf{new}} = \mathfrak{f} \times B$, which is shown in the bottom subplot. (One can also apply the classical flat restoration model to

the scalar field B along with the chromaticity restoration. See Tang, Sapiro and Caselles [18].)

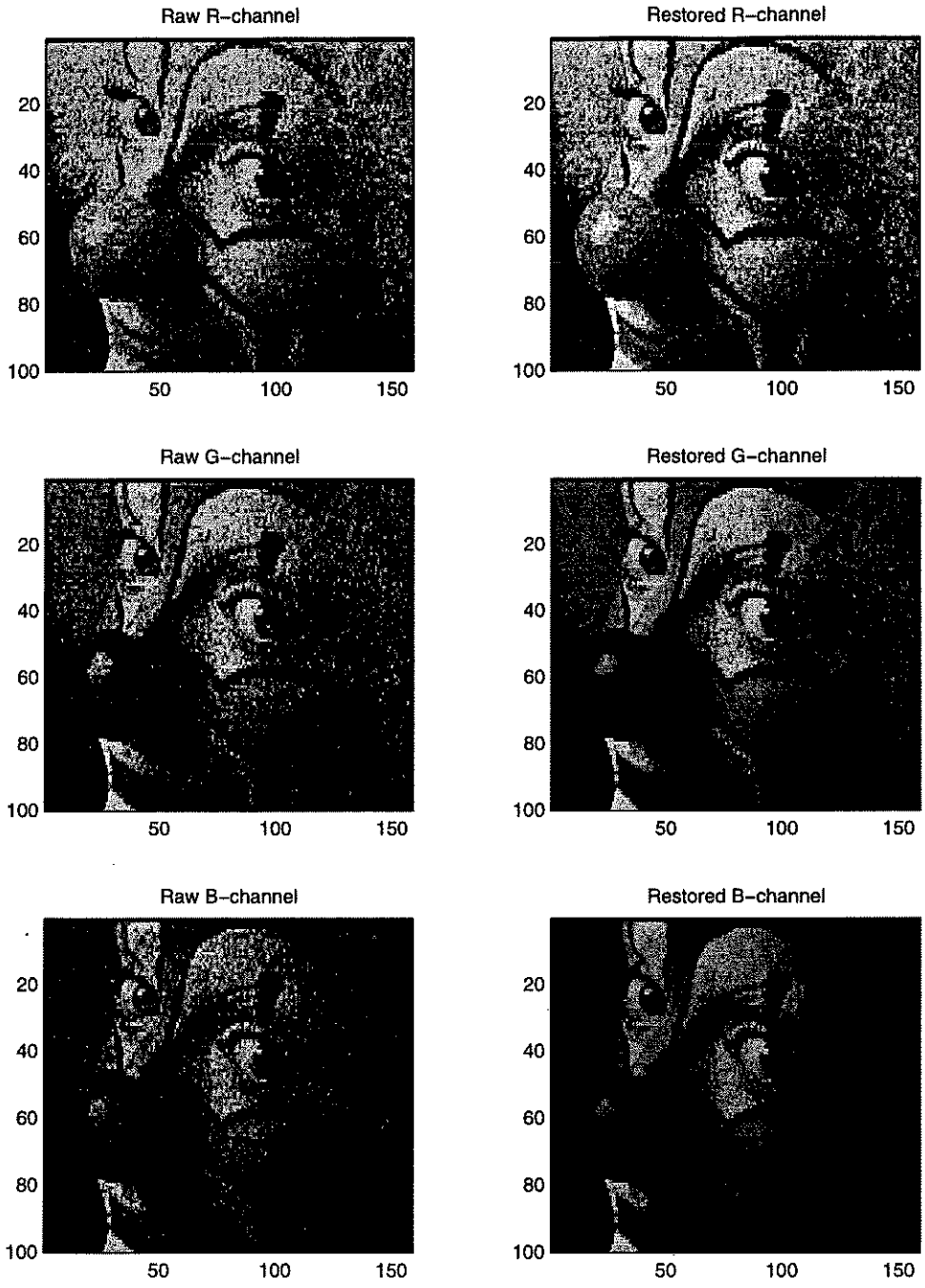

FIGURE 13. TV Restoration of chromaticity from Gaussian noise: the effect on each channel. (See Section 5.6.)

In Figure 13, we plot the raw and restored images (from Figure 12) channel by channel.

The restoration is quite successful. The visible noisy red and green dots have been almost completely removed. The eyes and dark lines resume their original black color. The nose and lips are now smoothly red. The restoration effect can be improved further if one applies an extra step of flat (or linear) restoration to the brightness component B (see Tang, Sapiro and Caselles [18]).

5.7. **Concluding remarks.** We have developed the general mathematical models for denoising and restoration of non-flat image features. The restoration models for three types of important non-flat features are studied in details. They are the orientation feature, the alignment feature, and the chromaticity feature. Algorithms resolving the nonlinearity of the models are constructed. Numerical evidences show that generally, the TV type of models perform better than the L^2 models.

Acknowledgments

It is a great pleasure to thank Professor Perona for bringing his work to our attention and for helpful discussions on the subject. We are also enormously grateful to Professors Sapiro and Osher for their numerous discussions, suggestions, and generous help.

References

[1] P. V. Blomgren and T. F. Chan. Modular solvers for constrained image restoration. Technical report, CAM 97-52, UCLA Department of Mathematics, 1997.

[2] A. Chambolle and P. L. Lions. Image recovery via total variational minimization and related problems. *Numer. Math.*, 76:167–188, 1997.

[3] T. Chan and L. Vese. Variational image restoration and segmentation models and approximations. submitted to *IEEE Tran. Image Process.*, 1999.

[4] F. R.K. Chung. *Spectral Graph Theory*, volume 92 of *Regional Conference Series in Mathematics*. Amer. Math. Soc., 1997.

[5] G. H. Golub and J. M. Ortega. *Scientific Computing and Differential Equations*. Academic Press, 1992.

[6] G. H. Granlund and H. Knuttson. *Signal Processing for Computer Vision*. Kluwer, Boston, 1995.

[7] S. Helgason. *Differential Geometry and Symmetric Spaces*. Academic Press, 1962.

[8] A. Marquina and S. Osher. Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal. Technical report, CAM 99-5, UCLA Department of Mathematics, 1999.

[9] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*, volume 14 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, Boston, 1995.

[10] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Applied. Math.*, XLII:577–685, 1989.

[11] S. Osher. Private communication. 1999.

[12] P. Perona. Orientation diffusion. *IEEE Trans. Image Proc.*, 7(3):457–467, 1998.

[13] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell.*, 12:629–639, 1990.

[14] L. Rudin and S. Osher. Total variation based image restoration with free local constraints. *Proc. 1st IEEE ICIP*, 1:31–35, 1994.

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

[16] G. Sapiro. Color snakes. Technical report, HPL-95-113, Hewlett Packard Computer Peripherals Laboratory, 1995.

[17] G. Sapiro and D. Ringach. Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Tran. Image Process.*, 5:1582–1586, 1996.

[18] B. Tang, G. Sapiro, and V. Caselles. Color image enhancement via chromaticity diffusion. Technical report, ECE-University of Minnesota, 1999.

[19] B. Tang, G. Sapiro, and V. Caselles. Direction diffusion. Technical report, ECE-University of Minnesota. To appear in *International Conference in Computer Vision*, 1999.

[20] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM J. Sci. Statist. Compt.*, 17:227–238, 1996.

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