UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

The Digital TV Filter and Nonlinear Denoising

Tony F. Chan
Stanley Osher
Jianhong Shen

October 1999

CAM Report 99-34

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

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

The digital TV filter and nonlinear denoising

Tony F. Chan, Stanley Osher and Jianhong Shen

Abstract

Motivated by the classical TV (total variation) restoration model, we propose a new nonlinear filter – the digital TV filter for digital images, or more generally, data living on a graph. The digital TV filter is a lowpass filter, capable of denoising data without blurring jumps or edges. We show that such local filters in iterations solve a global total variational optimization problem. Applications are given in the denoising of 1-D irregularly sampled signals, 2-D (color) images and data with complicated topological structures, and non-flat image features such as chromaticity.

Keywords

Total variation (TV), denoising, restoration, nonlinear, adaptive, digital filters, graph, color images, chromaticity.

EDICS Number: **IP 1.2.**

I. Introduction

THE digital filter is a frequently used tool in signal and image processing. Denoising via filtering is perhaps the oldest and still the most common application. What makes denoising so challenging is that a successful filter also must preserve characteristic singular features of images such as edges. Preservation of important singularities is absolutely necessary in image analysis and computer vision since digital "objects" are very much defined (or detected) via edges (i.e. segmentation) and other singularities (e.g. the corners of eyes and lips). Linear filters inevitably smear edges or jumps and create ringing artifacts. Therefore, nonlinearity plays a crucial role in image restoration.

The work of T. F. Chan and J. Shen were supported in part by NSF Grant DMS-9626755 and by ONR under N00014-96-1-0277. S. Osher was supported by NSF Grant DMS-9706827. The authors are with the Department of Mathematics, UCLA, Los Angeles, CA 90095-1555, USA. ({chan, sjo, jhshen}@math.ucla.edu).

The digital TV filter we propose here is a nonlinear filter which can denoise images without blurring edges. It has a simple and fixed structure (i.e. a 5-point rectangular stencil at the target pixel for digital images). The filter coefficients change adaptively, based on the local information near the target pixel, and are given by a deterministic (as contrast to statistical) formula (see Section 4). In addition to its simplicity, the digital TV filter is also very flexible in the sense that it applies to general data living on a graph, vectorial signals such as color images, and even non-flat image features such as chromaticity (see Section 6).

The edge information enters the filter coefficients in a very special way (see Eq. (8)). The formula allows an intuitive interpretation as far as the edge adaptivity is concerned (see Remark 2). Unlike other conventional filters such as the median filter, the formula is not derived from a qualitative awareness of edges, but rather, from the precise digital translation of the classical TV restoration model invented by Rudin, Osher and Fatemi [21] and Rudin and Osher [20].

The digital TV filter can be considered as the digitized TV restoration model (see Osher and Shen [16]). The latter will be surveyed briefly in the next section. The TV restoration model is one of the most successful tools for image restoration (including both denoising and deblurring) and edge enhancement (see Rudin and Osher [20], Rudin, Osher and Fatemi [21], and also the monograph by Morel and Solimini [15]). It is a special yet perhaps the most frequently applied case of the general anisotropic diffusion model proposed by Perona and Malik [18], mainly due to its connections to mean curvature motion and interface evolution. (The latter subject is now attracting an enormous amount of attention throughout computational science.)

The TV restoration model (or general variational models) is defined for analog (i.e. continuous) signals. The Euler-Lagrange equation associated with the TV functional is a nonlinear partial differential equation (see next section). When applying it to a digital image, one has to carefully choose numerical schemes [20] to take care of the nonlinearity. Therefore, Osher and Shen [16] established a self-contained "digital" theory for the PDE method, in which knowledge of PDEs and numerical approximations is not required. (We should mention that a different approach taken by Coifman and Sowa [6] [7] also avoids PDEs by incorporating wavelets.) Similar digitizing work on the evolving Euler-Lagrange equation (i.e. the diffusion equation instead of the equilibrium equation) can also be found in Weickert [26]. However, the mathematical foundation of the latter is still the numerical discretization of PDEs on rectangular grids.

In the digitized formulation of the PDE method (and variational method) [16], the restoration equations are nonlinear algebraic ones instead of PDEs. Techniques such as linearization and the iterative method in numerical linear algebra typically lead to simple and local digital filters [16]. The digital TV filter is constructed in this way from the classical TV restoration model.

In what follows, we first survey in Section 2 the classical TV restoration model. In Section 3, we compare the digital TV filter with some closely related ideas and discuss the former's special attributes. Section 4 details the digital TV filter, and filtering process and their properties. Section 5 discusses specifically the digital TV filter for color (or vectorial) images and for one of the most important non-flat image features — chromaticity. The last section is devoted to issues of implemention and to applications.

II. The TV Restoration Model of Rudin, Osher and Fatemi

Let $u^0(x)$ be a signal which is assumed to be the noise contaminated version of a clean signal $u(x)$, i.e.

$$u^0(x) = u(x) + n(x).$$

Here $n(x)$ denotes random noise with

$$\mathrm{E}\, n(x) = 0, \qquad \mathrm{E}\, n^2(x) = \sigma^2. \tag{1}$$

Linear time-frequency method leads to the distortion of important singularities since both the noise and singularities share high frequency modes. Researchers are thus led to nonlinear operators. One recent approach is to use wavelets to decompose the signal into many resolutions and keep (or modify) more details where singularities are detected (see Donoho and Johnstone [10] and Donoho [9], for examples). Another important approach is based on the PDE point of view, and takes advantage of the locality and anisotropy of certain differential equations. (Besides these two relatively modern methods, the third important class consists of various nonlinear digital filters which we shall discuss in the next section.)

Among all differential operators, the diffusion class is the most widely applied in current image analysis. For the reason discussed above, although linear homogeneous diffusion may smooth out noise successfully, it could also blur edges and jumps. Therefore anisotropic diffusion attracted the attention of researchers (Perona and Malik [18], Rudin [19]).

The TV anisotropic diffusion model was invented by Rudin, Osher and Fatemi [21], Rudin and Osher [20], and is now one of the most successful tools for image restoration. Compared with *least squares* restoration models [12], the major difference of the TV model is the transition "from 2 to 1." That is, one minimizes the total variation (the L^1 norm of the gradient instead of its L^2 norm)

$$\mathrm{TV}[u] = \int_\Omega |\nabla u| dx. \tag{2}$$

The simplicity of implementation of L^2 models is lost and complexity due to nonlinearity emerges. But the gain of this transition is enormous, both in image processing and in mathematics. The latter is due to its deep connection to nonlinear functional analysis (e.g. function spaces of bounded variations) and geometry (i.e. mean curvature motions and interface evolutions), and the former because of its effectiveness in restoring images.

The assumptions on the noise (1) now lead to two constraints for the minimization of the TV norm:

$$\int_\Omega u dx = \int_\Omega u^0 dx, \qquad \frac{1}{|\Omega|}\int_\Omega (u-u^0)^2 dx = \sigma^2. \tag{3}$$

Here Ω denotes the continuous signal domain, $|\Omega|$ its area, and ∇u the gradient. Therefore, (2) and (3) define a constrained optimization problem.

Because of the translation invariance of the TV norm: $\mathrm{TV}[u+c] = \mathrm{TV}[u]$ for any constant c, the first constraint is in fact automatically encoded (see Chambolle and Lions [3]). Therefore, typically, we need only consider the second fitting constraint. By introducing a Lagrange relaxation parameter λ, one can define a new energy functional

$$J[u] = \int_\Omega |\nabla u| dx + \frac{\lambda}{2}\int_\Omega (u-u^0)^2 dx. \tag{4}$$

The Euler-Lagrange equation of J is

$$-\nabla \cdot \left(\frac{\nabla u}{|\nabla u|}\right) + \lambda(u-u^0) = 0, \tag{5}$$

and the infinitesimal steepest descent evolution gives

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

Setting $\lambda = 0$, or considering no fitting constraint, we get the famous Osher-Rudin TV diffusion or the *weighted* mean curvature motion (see Morel and Solomini [15], Rudin and Osher [20]):

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right).$$

To avoid singularities in flat regions or at extrema, $|\nabla u|$ in Eq. (5) is regularized to

$$|\nabla u|_a = \sqrt{|\nabla u|^2 + a^2},$$

for some small positive parameter a (10^{-4} for instance). Then the modified Euler-Lagrange equation (5) in fact minimizes the regularized energy functional

$$J_a[u] = \int_\Omega |\nabla u|_a dx + \frac{\lambda}{2} \int_\Omega (u - u^0)^2 dx.$$

The TV model has also led to successful schemes for deblurring images, which is usually beyond linear traditional methods since inverse problems like deblurring are highly ill-posed. Let j be the blurring kernel. For simplicity, assume it is spatially homogeneous. Then a noisy blurred image is modeled by

$$u^0 = j * u + n.$$

To restore u is to find u from u^0, and the TV model is one of the most powerful tools for this multi-task restoration (see Rudin and Osher [20], Rudin, Osher and Fatemi [21]).

III. Some Related Nonlinear Digital Filters

As mentioned above, the major problem for a linear constant filter (or *weighted averaging* with fixed weights) is that in iterations it blurs intrinsic singularities such as jumps or edges (just as the Gaussian does during diffusion processing.) But the good news about these filters is that they are simple — with fixed filter supports and uniform filter coefficients. So for a long time, researchers have been asking the following question:

"For fixed filter support, is there an explicit formula for the averaging weights so that the filter can avoid blurring intrinsic singularities?"

This long-standing question now gets an answer from our digital TV filters. We shall give a simple formula for the filter coefficients $h_{\alpha\beta}$ in the next section. The filter has the simplest structure, as far as we know, among all filters designed to preserve edges or jumps.

In this section, we briefly review and comment on some existing work closely related to the digital TV filter.

A. Digital TV filters v.s. the edge adaptive filters of Lev, et. al.

As early as two decades ago, Lev, Zucker and Rosen [13] proposed a possible way to choose the filter coefficients. The support of their filter is the 3×3 neighborhood centered at the target pixel. The filter has *twelve* different versions $D_1, D_2, \cdots, D_{12}$ corresponding to twelve different patterns of local variations of the image. For instance, for the seventh case (see [13]),

$$D_7 = \frac{1}{6} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

For their type of filters, essentially one needs an *edge detector* first to check which particular type of local variations one is encountering at the target pixel. One of the authors' important ideas — that one should never average across the edges (see [13]) — is also critical in the modern practice of *anisotropic diffusion* (see Perona and Malik [18], Rudin [19] and Weickert [26]). The digital TV filter we present here combines the two steps, i.e. edge detection along with an automatic modification of the filter coefficients. The simple formula of digital TV filters given in the next section achieves this two-fold goal.

B. Digital TV filters v.s. median filters.

Median filters are the well-known tool in the denoising and restoration of both grey level and color images (see Glasbey and Horgan [11], Gonzalez and Woods [12]). They are based on the local histogram or local order statistics of the image. To get the filtering output at a target pixel α, one needs all the image values (grey level values or RGB vectors) within a window around α, typically a 7×7 or 5×5 square window (see [11]). Then the new output at α is the median of these sample image values. This means that the output is one of the existing values inside the window. From the viewpoint of weighted averaging filters, the median filter is an *exclusive filter* in the sense that the filter coefficients $h_{\alpha\beta} = 0$ or 1. Because of the lowpass condition, one and only one filter coefficient can be non-zero. Therefore, for the median filter, all the filter coefficients are constants and the non-linearity resides in the way of determining the unique neighboring pixel β such that $h_{\alpha\beta} = 1$. The median filter and digital TV filter differ in three aspects:

(a) For the digital TV filter, the filter coefficients $h_{\alpha\beta}$'s are given by a fixed explicit formula (see Eq. (8)). Median filters are exclusive as mentioned above.

(b) The window (or, filter support) for the digital TV filter is small. For image processing, it can be simply the standard rectangular 5-node stencil, instead of 25 or 49 nodes for median filters. The large size of a median filter is due to to its statistical origin — no realistic statistics can be obtained from a small sample of data.

(c) The median filter is designed so that a single step of filtering minimizes an l^1 norm based cost function inside the window (see Trahanias et. al. [24] [25]). The digital TV filter, on the other hand, is constructed with a global l^1 minimization problem in mind and is applied in an iterative fashion.

C. Digital TV filters for chromaticity v.s. the VDF of Trahanisa et. al.

The digital TV filter is *universal* in the sense that it applies to grey level images, multi-channel color images (vectorial), and even one of the most important non-flat image features —chromaticity (see Tang, Sapiro and Caselles [23], [22], Chan and Shen [5]). In particular, the digital TV filter for chromaticity (see Section 5) will be a potential competitor for the VDF (vector directional filter) invented by Trahanisa et. al. [24] [25], which is closely connected to spherical median filters.

D. Digital TV filter: no need for stopping time

To the best of our knowledge, for the most existing digital filters, the raw noisy image u^0 is "abandoned" right after the first iteration. That is, the filtering process can be described as

$$u^0 = u^{(0)} \longrightarrow u^{(1)} \longrightarrow u^{(2)} \longrightarrow \cdots$$

At step k, $u^{(k)}$ depends solely on $u^{(k-1)}$. The ignorance of u^0 at later steps causes intrinsic singularities to be smeared step by step (the discrete scale-space filter is an extreme example [26]. Therefore, a stopping time is required for such filtering processes.

It is often a difficult task to determine an optimal stopping time. This is avoided by the digital TV filter since it recycles u^0 at each step (see next section):

$$u^{(0)} \xrightarrow{u^0} u^{(1)} \xrightarrow{u^0} u^{(2)} \xrightarrow{u^0} \cdots$$

The presence of u^0 at each step constantly reminds the filter not to forget the noisy image, which has information about the original singular features such as jumps and edges. We should point out, however, although the stopping time is unnecessary for the digital TV filter, the difficulty goes into the determination of the degree of influence that u^0 should impose at each step. Fortunately, simple ways do indeed exist for making a nearly optimal decision (see the last section on applications).

IV. Graphs and the Digital TV Filter

In this section, we introduce the digital TV filter and study its properties.

A. Graphs and edge derivatives

The digital domain is modeled by an undirected graph $[\Omega, E]$ with a finite set Ω of vertices (or nodes) and an edge dictionary E. (The necessity of the graph model in data analysis can be found in Osher and Shen [16], Alpert et. al. [1]). If α and β are linked by an edge, we write $\alpha \sim \beta$. A digital signal u is a function on Ω:

$$u : \quad \Omega \to R.$$

The value at node α is denoted by u_α. The *local variation* or *strength* $|\nabla_\alpha u|$ at any node α is defined by

$$|\nabla_\alpha u| := \sqrt{\sum_{\beta \sim \alpha} (u_\beta - u_\alpha)^2}. \tag{6}$$

[*A Note.* There are a couple of other definitions existing in the literature. For example, the *wirelength* definition inspired by the network cell-placement problem in Alpert et al. [1]. The one we propose here corresponds to the direct generalization of

$$|\nabla u| = \sqrt{u_{x_1}^2 + u_{x_2}^2 + \cdots + u_{x_n}^2}$$

in the n-D continuous case.]

For any positive number a, the regularized local variation is

$$|\nabla_\alpha u|_a = \sqrt{|\nabla_\alpha u|^2 + a^2}.$$

Next, we define the *edge derivative.* Let e denote the edge $\alpha \sim \beta$. Then the edge derivative of u along e *at* α is defined to be

$$\left.\frac{\partial u}{\partial e}\right|_{\alpha} := u_{\beta} - u_{\alpha}.$$

Apparently,

$$\left.\frac{\partial u}{\partial e}\right|_{\alpha} = -\left.\frac{\partial u}{\partial e}\right|_{\beta}, \qquad \text{and} \qquad |\nabla_{\alpha} u| = \sqrt{\sum_{e \vdash \alpha} \left[\left.\frac{\partial u}{\partial e}\right|_{\alpha}\right]^2},$$

where $e \vdash \alpha$ means that α is one node of e.

B. The digital TV filter

The digital TV filter contains two tunable parameters:

– a small positive parameter a called the *regularization parameter*;

– a positive parameter λ called the *fitting parameter.*

(In Section IV, we will address on how to choose these two parameters in applications.) For a given noisy signal u^0, the *digital TV filter* $\mathcal{F}^{\lambda,a}$ is a nonlinear data-dependent filter:

$$\mathcal{F}^{\lambda,a}: \quad u \to v.$$

Here u is any existing signal on Ω and v the output. For simplicity, we shall denote $\mathcal{F}^{\lambda,a}$ by $\mathcal{F}$. For any node $\alpha \in \Omega$,

$$v_{\alpha} = \mathcal{F}_{\alpha}(u) = \sum_{\beta \sim \alpha} h_{\alpha\beta}(u) u_{\beta} + h_{\alpha\alpha}(u) u^0_{\alpha}. \tag{7}$$

Here the filter coefficients $h_{\alpha\beta}$ are given by

$$h_{\alpha\beta}(u) = \frac{w_{\alpha\beta}(u)}{\lambda + \sum_{\gamma\sim\alpha} w_{\alpha\gamma}(u)}, \qquad h_{\alpha\alpha}(u) = \frac{\lambda}{\lambda + \sum_{\gamma\sim\alpha} w_{\alpha\gamma}(u)}. \tag{8}$$

$$w_{\alpha\beta}(u) = \frac{1}{|\nabla_{\alpha} u|_a} + \frac{1}{|\nabla_{\beta} u|_a}.$$

Apparently, h is a lowpass filter in the sense that at any node α,

$$h_{\alpha\alpha} + \sum_{\beta\sim\alpha} h_{\alpha\beta} = 1.$$

The complete algorithm at node α is therefore

TV Filtering at α:

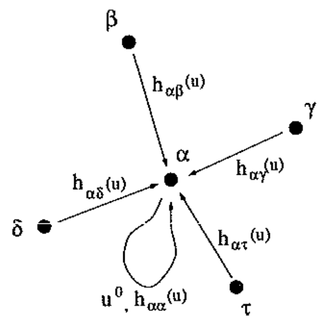


Fig. 1. The digital TV filter at node α

- Compute the local variation $|\nabla u|_a$ at α and all its neighbors β;
- Compute the weights $w_{\alpha\beta}$;
- Compute the filter coefficients $h_{\alpha\alpha}$ and $h_{\alpha\beta}$;
- Filtering: $\mathcal{F}_\alpha(u) = \sum_{\beta\sim\alpha} h_{\alpha\beta}u_\beta + h_{\alpha\alpha}u^0_\alpha$.

  Assign a linear order to all nodes:

$$\alpha_1 < \alpha_2 < \cdots < \alpha_D.$$

Then the process of *TV filtering* is defined as follows

TV Filtering Process:

- Initialize $u^{(0)}$ (by setting $u^{(0)} = u^0$, typically but not necessary).
- For $k = 1, 2, \cdots$

  For $\alpha_1 < \alpha_2 < \cdots < \alpha_D$

  $u^{(k)}_{\alpha_j} = \mathcal{F}_{\alpha_j}(u^{(k-1)})$;

  End

  End

For the TV filtering process, the computational cost can be reduced by noticing that the local variations at *all* nodes can be worked out in the beginning of each k-loop and utilized for all α-subloops.

For image processing, the underlying graph can simply be taken to be the rectangular net. Therefore, only five immediate neighbors are involved in one step of filtering. This is the *locality* of the digital TV filter, which makes parallelization easily.

Thoerem 1 *If the above TV filtering process converges to some signal u, then u sat-*

isfies

$$\sum_{e \vdash \alpha} \frac{\partial}{\partial e} \frac{-1}{|\nabla u|_a} \frac{\partial u}{\partial e}\bigg|_\alpha + \lambda(u^0_\alpha - u_\alpha) = 0, \quad \alpha \in \Omega. \tag{9}$$

Here $e \vdash \alpha$ *means that* α *is one node of* e.

Proof: The limit apparently satisfies

$$u_\alpha = \mathcal{F}_\alpha(u), \qquad \alpha \in \Omega,$$

or equivalently,

$$(\lambda + \sum_{\beta \sim \alpha} w_{\alpha\beta}) u_\alpha = \sum_{\beta \sim \alpha} w_{\alpha\beta} u_\beta + \lambda u^0_\alpha; \tag{10}$$

$$\sum_{\beta \sim \alpha} w_{\alpha\beta}(u_\beta - u_\alpha) + \lambda(u^0_\alpha - u_\alpha) = 0. \tag{11}$$

Therefore, it suffices to show that

$$\sum_{\beta \sim \alpha} w_{\alpha\beta}(u_\beta - u_\alpha) = \sum_{e \vdash \alpha} \frac{\partial}{\partial e} \frac{-1}{|\nabla u|_a} \frac{\partial u}{\partial e}\bigg|_\alpha . \tag{12}$$

We leave this verification to our readers. ■

Remark 1 *Eq. (9) is the digital version of*

$$\operatorname{div}\left[\frac{\operatorname{grad}(u)}{|\operatorname{grad}(u)|_a}\right] + \lambda(u^0 - u) = 0,$$

in the classical continuous case, where the first term geometrically means the curvature of the level curve of u *when* Ω *is a domain in* $\mathbb{R}^2$. *(Notice the sign difference between the graph curvature term and classical curvature term.)*

Corollary 1 *If the TV filtering process converges, then the limit signal* u *is the unique minimizer for the fitted TV energy*

$$FTV[v] = \sum_{\alpha \in \Omega} |\nabla_\alpha v|_a + \frac{\lambda}{2} \sum_{\alpha \in \Omega} (v_\alpha - u^0_\alpha)^2. \tag{13}$$

Proof: It suffices to notice that the left hand side of Eq.(9) is exactly the negative gradient of FTV at α, and FTV is strictly convex (see Chambolle and Lions [3], Osher and Shen [16]). ■

Proposition 1 *In a TV filtering process with the fitting parameter* $\lambda = 0$ *on a* con- nected *graph, the limit signal must be a constant.*

Proof: Since $\lambda = 0$, we have $h_{\alpha\alpha} = 0$ for any node α. Therefore, if u is the limit signal, then

$$u_\alpha = \sum_{\beta \sim \alpha} h_{\alpha\beta} u_\beta.$$

Since h is a lowpass filter with positive coefficients (regardless its nonlinear dependence on the data itself), we therefore have the *maximum principle*:

$$\min_{\beta \sim \alpha} u_\beta \le u_\alpha \le \max_{\beta \sim \alpha} u_\beta,$$

and any of the two equalities holds if and only if u is flat at α, i.e. $u_\beta \equiv u_\alpha$ for all $\beta \sim \alpha$. This implies that any peak or valley (local or global) of u must be flat. Since the graph is connected, it is now easy to see that u must be a constant signal. This completes the proof. ■

Remark 2 (Adaptivity of the TV filter) *The adaptive property of the digital TV filter can be easily understood qualitatively as follows. Consider those $u^{(k)}$'s that are near the limit. Suppose the local variation $|\nabla_\beta u^{(k)}|$ is very large (and large enough to be distinct from noise) near a node α, then we can interpret it as the evidence of an intrinsic jump (or edge) inherited from u^0 at this location. We surely do not want it to be distorted. The digital TV filter achieves this goal since*

$$\textit{large } |\nabla u^{(k)}| \quad \Rightarrow \quad \textit{small } w_{\alpha\beta} \textit{ (compared to } \lambda) \quad \Rightarrow \quad h_{aa} \textit{ is nearly 1.}$$

Now if on the other hand, the local variation $|\nabla u|$ is small, which implies that the $u^{(k)}$ is nearly flat at that location, then we think of it as the denoised version of $u^{(0)}$, and we certainly do not want it to go back to the original noisy data. Therefore, we expect a small $h_{\alpha\alpha}$ and a pure lowpass filtering on $u^{(k)}$ (i.e. $\sum_{\beta \sim \alpha} h_{\alpha\beta}$ is nearly 1), which makes $u^{(k+1)}$ even flatter. This is precisely the work done by the TV filter since

$$\textit{small } |\nabla u^{(k)}| \quad \Rightarrow \quad \textit{large } w_{\alpha\beta} \textit{ (compared to } \lambda) \quad \Rightarrow \quad h_{aa} \textit{ is nearly 0.}$$

The exact (quantitative) balance is encoded into the formulae of digital TV filters. Nevertheless, this qualitative interpretation does add more "readability" to the digital TV filter, like many other existing ones.

V. Digital TV Filters for Color Images and Chromaticity

In this section, we discuss the natural extension of the digital TV filter to color images and chromaticity.

A. Color images or vectorial signals

A color image takes values in the RGB $\mathbb{R}^3$ space:

$$\mathbf{I}: \quad \Omega \to \mathbb{R}^3.$$

$\mathbf{I}_\alpha$ is then the RGB value at pixel α. Let $|\cdot|$ denote the Euclidean norm. Then the local variation is again well defined:

$$|\nabla_\alpha \mathbf{I}| = \left[\sum_{\beta \sim \alpha} |\mathbf{I}_\beta - \mathbf{I}_\alpha|^2 \right]^{\frac{1}{2}}.$$

Since the digital TV filter essentially only depends on this scalar quantity, it naturally applies to color images (or more generally, vectorial signals). The formulae in the previous section for the scalar case are all valid.

Proposition 2 *If a convex set Σ contains all vectors $\mathbf{I}_\alpha^{(0)}$ (the initial signal) and $\mathbf{I}_\alpha^0$ (the noisy signal) for all $\alpha \in \Omega$, then for all k and $\alpha \in \Omega$, $\mathbf{I}_\alpha^{(k)} \in \Sigma$.*

Proof: It follows easily from that fact that the digital TV filter is a lowpass filter with positive coefficients. ■

In particular, for color images, the limit image $\mathbf{I}$ will also reside in the first octant in the RGB space if we start the filtering process with a color image. This makes perfect sense since negative values do not correspond to RGB values.

In the literature of color image processing, there exist many important non-linear filters, such as *Vector Median Filters* (VMF) and *Vector Directional Filters* (VDF) (see [24] [25]). These filters are designed based on statistical considerations. They solve a local l^1 optimization problem within a sampling window. The digital TV filter is also a local filter, yet the filtering process solves a global or functional l^1 optimization problem. This is the fundamental difference between the digital TV filter and most of the existing nonlinear digital filters.

B. Digital TV filter for chromaticity

Several authors have recently studied the denoising and restoration of non-flat image features (Perona [17], Tang, Sapiro and Caselles [22], [23]), Chan and Shen [5]). Non-flat image features are those that live on non-Euclidean manifolds. Orientation (i.e. the unit circle $\mathbf{S}^1$), alignment (i.e. the real projected line $\mathbb{RP}^1$), and chromaticity (i.e. the sphere $\mathbf{S}^2$) are typical examples of non-flat image features that are important in pattern analysis, study of optical flows, and restoration of color images. The TV (or L^1) model performs better than the L^2 model for non-flat image features as well, as expected from the classical literature (see Chan and Shen [5]). In this section, we discuss the digital TV filter for chromaticity.

Let

$$\mathbf{I} : \Omega \rightarrow \mathbb{R}^3_+ = \{(r, g, b) \ : \ r, g, b > 0\}$$

denote a color image, and $|\cdot|$ the Euclidean norm in $\mathbb{R}^3$. Then $B = |\mathbf{I}|$ encodes the brightness (or luminance) information, while $\mathbf{u} = \mathbf{I}/|\mathbf{I}|$ records the color saturation of the image. $\mathbf{u}$ is therefore called *chromaticity* in image processing. Apparently, the manifold for chromaticity is the sphere $\mathbf{S}^2$.

There are many distinct ways of measuring the distance between two chromaticity "points" $\mathbf{u}$ and $\mathbf{v}$, among which the familiar and convenient two are:

— Geodesic distance (or *intrinsic*) d_i:

$$d_i(\mathbf{u}, \mathbf{v}) = \cos^{-1}\langle \mathbf{u}, \mathbf{v} \rangle,$$

where $\langle \mathbf{u}, \mathbf{v} \rangle$ is the Euclidean product. d_i is the length of the shortest arc of the big circle that links the two points.

— Embedded distance d_e:

$$d_e(\mathbf{u}, \mathbf{v}) = |\mathbf{u} - \mathbf{v}|.$$

Infinitesimally, they are equivalent to the first order. Experiments also suggest that they make no distinguishable difference to human visual perception when applied to denoising and restoration. Therefore, in what follows, we shall apply the embedded distance d_e only, which is relatively easier to work with.

Define the local variation of a chromaticity signal $\mathbf{u}$ to be

$$|\nabla_\alpha \mathbf{u}| = \left[\sum_{\beta\sim\alpha} d_e^2(\mathbf{u}_\alpha, \mathbf{u}_\beta)\right]^{\frac{1}{2}}.$$

(Generally, one can replace d_e by any reasonable distance d, see Chan and Shen [5]). Similar to the scalar case, we can define the weights $w_{\alpha\beta}$ and filter coefficients $h_{\alpha\beta}$. Then the digital TV filter $\mathcal{F}$ for chromaticity is defined accordingly. Let $\mathbf{u}^0$ be the target noisy chromaticity signal $\mathbf{u}^0$. For any chromaticity $\mathbf{u}$,

— $\mathbf{w}_\alpha = \sum_{\beta\sim\alpha} h_{\alpha\beta}\mathbf{u}_\beta + h_{\alpha\alpha}\mathbf{u}^0_\alpha$,

— $\mathcal{F}_\alpha(\mathbf{u}) := \mathbf{w}_\alpha/|\mathbf{w}_\alpha|$.

Notice that $\mathbf{w}_\alpha$ cannot be zero for chromaticity since chromaticity in fact lives on the positive octant of the sphere and the TV filter is a lowpass filter with positive coefficients. The normalization step (i.e. the second step) is not accidental at all. It is a part of a whole *global* non-flat optimization process. This is stated by the following theorem. The asserted link is valid only when we choose the distance function to be the embedded distance. The proof is similar to Theorem 1, or see Chan and Shen [5].

Starting from an initial chromaticity distribution $\mathbf{u}^{(0)}$, we can iterate the TV filter as in the scalar case. Set

$$\mathbf{u}^k_\alpha = \mathcal{F}_\alpha(\mathbf{u}^{(k-1)}), \qquad k = 1, 2, \cdots, \ \alpha \in \Omega.$$

Thoerem 2 *Suppose that the TV filtering process converges to some chromaticity signal* $\mathbf{u} : \Omega \to \mathbf{S}^2$*, then at any* $\alpha \in \Omega$,

$$\mathbf{\Pi}_{\mathbf{u}_\alpha}\left[\frac{\partial}{\partial e}\frac{-1}{|\nabla \mathbf{u}|_a}\frac{\partial}{\partial e}\mathbf{u}\bigg|_\alpha + \lambda(\mathbf{u}^0_\alpha - \mathbf{u}_\alpha)\right] = \mathbf{0},$$

where $\mathbf{\Pi}_\mathbf{w}$ *denotes the orthogonal projection onto the tangent space of* $\mathbf{S}^2$ *at* $\mathbf{w}$*, i.e.*

$$\mathbf{\Pi}_\mathbf{w}(\mathbf{v}) = \mathbf{v} - \langle \mathbf{v}, \mathbf{w}\rangle\, \mathbf{w}.$$

Therefore, $\mathbf{u}$ *is an equilibrium signal with respect to the fitted total variation functional on all chromaticity distributions* $\mathbf{v} : \Omega \to \mathbf{S}^2$*:*

$$FTV[\mathbf{v}] = \sum_{\alpha\in\Omega} |\nabla_\alpha \mathbf{v}| + \frac{\lambda}{2}\sum_{\alpha\in\Omega} d_e^2(\mathbf{v}_\alpha, \mathbf{u}^0_\alpha).$$

VI. Applications

In this section, we present four applications of the digital TV filter and filtering process. First, we address on some issues regarding implementation.

(a) One attribute of the digital graph model is it does not require any artificial "boundary" condition. In the classical literature, the continuous anisotropic diffusion equation is usually accompanied by the adiabatic Neumann condition (see Rudin and Osher [20]). In the digital model, the boundary condition has been encoded into the *structure* of the graph and the definition of the local variation $|\nabla_\alpha u|$. For instance, each of the four corner pixels has only two neighbors and a typical boundary pixel has three neighbors. A simple checking on our definition of the *local variation* $|\nabla_\alpha u|$ verifies that the above boundary structure of the graph indeed corresponds to a flat outward extension of u, or the discrete outward Neumann condition.

(b) The regularization constant a is purely for the purpose of computational stability (to avoid a zero denominator). It has no significant influence as long as we set it very small. For applications in image processing, one can choose a typical order of 10^{-4}. Its influence has been further controlled in our digital model, since the digital TV filter is a lowpass filter and thus the computation is always stabilized.

(c) The regularization constant λ is important for the restoration effect. Practical concerns and estimations are discussed in Rudin and Osher [20], Blomgren and Chan [2]. In terms of the digital model, an estimation of the optimal λ from a current signal u is by (see [20])

$$\lambda \simeq \frac{1}{\sigma^2}\frac{1}{|\Omega|}\sum_{\alpha\in\Omega}\sum_{\beta\sim\alpha} w_{\alpha\beta}(u_\beta - u_\alpha)(u_\alpha - u^0_\alpha).$$

Here,

- σ^2 is the variance of the noise, which is known or can be estimated from homogeneous regions in the data.
- $w_{\alpha\beta}$ are the weights in the definition of the digital TV filter.
- $|\Omega|$ is the size of the graph, i.e. the total number of nodes.

The formula suggests that λ is comparable to $1/\sigma^2$. Therefore, a good economic strategy for the digital TV filter is to start the filtering process with $\lambda = 1/\sigma^2$, then after every K steps ($K = 10$, say), λ is updated according to the above formula. Typically, one needs only about 6 updatings in one application. Our experiments suggest that human vision seems to be insensitive to the accuracy of λ. Therefore, in what follows, we simply fix some value close to $1/\sigma^2$ for the entire TV filtering process.

(d) For those who wonder what will happen if one replaces the fitted TV energy (see Eq. (13)) by the fitted l^2 energy

$$\mathrm{FL}^2[u] = \frac{1}{2}\sum_{\alpha\in\Omega} |\nabla_\alpha u|^2 + \frac{\lambda}{2}\sum_{\alpha\in\Omega}(u_\alpha - u_\alpha^0)^2, \tag{14}$$

we introduce the *fitted linear filter* $\hat{h}_{\alpha\beta}$. For any $\alpha \in \Omega$, let d_α denotes its degree (i.e. the total number of edges from α). Define

$$\hat{h}_{\alpha\beta} = \frac{2}{\lambda + 2d_\alpha}, \qquad \hat{h}_{\alpha\beta} = \frac{\lambda}{\lambda + 2d_\alpha}. \tag{15}$$

Then it is not difficult to show that an h–filtering process minimizes the preceding fitted l^2 energy (or see Osher and Shen [16]). Such a linear filter is doomed to smear edges and/or create oscillations, as one can see from the first example.

(e) We do not address the convergence issue in this paper. Our experiments have shown that usually 60 or 80 rounds of TV filtering are satisfactory enough for human vision. For related convergence analysis, we refer to Dobson and Vogel [8], Chan and Mulet [4], Marquina and Osher [14].

A. Digital TV filter applied to a 1-dimensional signal

In Fig. 2, the top one plots the "shape" of a one-dimensional landscape with two "cliffs." Plotted right below is the "measured" height data at *irregularly* sampled spots. Due to the error of measuring tools and human factors, the set of measured data are noisy, which is modeled in computers by polluting the height data with a random noise. To compare with the digital TV filter, in the third figure is plotted the restoration result by the fitted linear filter (see Eq.(15)). One observes that the two cliffs have been artificially expanded. The last one shows the restoration and denoising by the digital TV filter with $\lambda = 3$ and $a = 0.0001$. The noise is successfully diminished without significantly degrading the intrinsic jumps in the clean data. The TV filter clearly performs better than the l^2 fitted linear filter.

B. Beyond image processing: data on a Sierpinski graph

In our second example, the digital TV filter is applied to data living on a level 3 Sierpinski graph (Fig. 3). (For possible applications of similar graph models in oceanography, see Osher and Shen [16]). In order to visualize the noisy data and the restoration effect by the TV filter, we have labeled the 42 nodes in a linear order.

The subplot on top in Fig. 4 shows the noisy data u^0 according to the labeling in Fig. 3. The digital TV filtering ($\lambda = 4, a = .0001$) is applied to u^0, and the restoration result is plotted in the bottom. Notice that we apply the digital TV filtering to the original graph topology, not the labeled linear chain. As a result, we may easily interpret the overshooting at node 25, which is the "boundary" node between the right and left wings of the Sierpinski graph.

C. Digital TV filter applied to a color image

In Fig. 5, the vectorial digital TV filtering (see Section V-A) is applied to a noisy color image of stones. Notice that the noise has been successfully removed without smearing the sharp edges of the stones.

D. Digital TV filter applied to chromaticity restoration

In Fig. 6, we apply the digital TV filter modified for chromaticity to a clown image with Gaussian noise (the top image). We first separate the vector RGB image $\mathbf{I}^0$ into the brightness component $B = |\mathbf{I}^0|$, and the chromaticity component $\mathbf{u}^0 = \mathbf{I}^0/B$. Then we apply the digital TV filtering process to $\mathbf{u}^0$ and get the optimal restoration $\mathbf{u}$. Finally, we keep using the original brightness value B to assemble the new image $\mathbf{I}_{\text{new}} = B \times \mathbf{u}$, which is shown in the bottom subplot. (One can also apply the scalar digital TV filtering to the scalar field B along with the chromaticity restoration. See Tang, Sapiro and Caselles [23].)

The restoration is quite successful. The visible noisy red and green dots have been swept out. The eyes and dark lines resume their original black color, and the nose and lips are now smoothly red.

References

[1] C. Alpert, T. Chan, A. Kahng, I. Markov, and P. Mulet. Faster minimization of linear wirelength for global placement. *IEEE Trans. Computer-Aided Design*, 17(1):3–13, 1998.

[2] P. Blomgren and T. Chan. Modular solvers for constrained image restoration problems. CAM Report 97-52, Department of Mathematics, UCLA, 1997.

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

[4] T. Chan and P. Mulet. On the convergence of the lagged diffusivity fixed point method in image restoration. *SIAM J. Num. Anal.*, 36(2):354–367, 1999.

[5] T. Chan and J. Shen. Variational restoration of non-flat image features: models and algorithms. Submitted to *SIAM J. Appl. Math.*, 1999.

[6] R. R. Coifman and A. Sowa. Combining variational calculus and wavelets for image enhancement. *Appl. Comp. Harmon. Anal.*, to appear, 1999.

[7] R. R. Coifman and A. Sowa. New types of Total Variation diminishing flows. Preprint, 1999.

[8] D. C. Dobson and C. R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM J. Numer. Anal.*, 34(5):1779–1791, 1997.

[9] D. L. Donoho. De-noising by soft thresholding. *IEEE Trans. Inf. Th.*, 41:613–627, 1995.

[10] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. Technical Report, Department of Statistics, Stanford University, 1992.

[11] C. A. Glasbey and G. W. Horgan. *Image Analysis for the Biological Sciences.* Statistics in Practice. John Wiley & Sons, New York, 1995.

[12] R. C. Gonzalez and R. E. Woods. *Digital Image Processing.* Addison-Wesley, New York, 1992.

[13] A. Lev, S. W. Zucker, and A. Rosenfeld. Iterative enhancement of noisy images. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-7(6):435–442, 1977.

[14] A. Marquina and S. Osher. *Lecture Notes in Computer Science*, volume 1682, chapter "A new time dependent model based on level set motion for nonlinear deblurring and noise removal", pages 429–434. 1999.

[15] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*, volume 14. Birkhäuser, Boston, 1995.

[16] S. Osher and J. Shen. Digitized PDE method for data restoration. *Analytic-Computational Methods in Applied Mathematics,* to appear, 2000.

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

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

[19] L. Rudin. Images, numerical analysis of singularities and shock filters. Caltech C.S. Dept. Report TR 5250:87, 1987.

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

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

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

[23] B. Tang, G. Sapiro, and V. Caselles. Direction diffusion. *International Conference in Computer Vision,* to appear, 1999.

[24] P. E. Trahanias, D. Karako, and A. N. Venetsanopoulos. Directional processing of color images: theory and experimental results. *IEEE Trans. Image Process.*, 5(6):868–880, 1996.

[25] P. E. Trahanias and A. N. Venetsanopoulos. Vector directional filters — a new class of multichannel image processing filters. *IEEE Trans. Image Process.*, 2(4):528–534, 1993.

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

Tony F. Chan(M'98) received the B.S. degree in engineering and the M.S. degree in aerospace engineering, both in 1973, from California Institute of Technology, Pasadena, and the Ph.D. degree in computer science from Stanford University, Stanford, CA, in 1978. He is currently the Department Chair of the Department of Mathematics, University of California, Los Angeles, where he has been a Professor since 1986. His research interests include PDE methods for image processing, multi grid, and domain decomposition algorithms, iterative methods, Krylov subspace methods, and parallel algorithms.

Stanley Osher received his B.S. from Brooklyn College (1962) and M.S. (1964) and Ph.D. (1966) from the Courant Institute, NYU. He has been a Professor of Mathematics at UCLA since 1977. He is coinventor of the level set method, ENO and WENO high resolution schemes, and TV based image restoration.From 1988-1995 he was co-CEO and cofounder of Cognitech, Inc., Santa Monica. He has been a Fulbright fellow, Alfred P. Sloan Fellow, received the NASA Public Service Group Award, and was an invited speaker at the International Congress of Mathematicians. His interests include PDE based image processing, moving front problems and numerical methods.

Jianhong Shen received his B.S. degree from the University of Science and Technology of China in 1994, and Ph.D degree from Massachusetts Institute of Technology in 1998. He is currently a CAM (Computational and Applied Mathematics) Assistant Professor at UCLA. His current research interests are in the area of wavelets analysis, analysis and computation of multiscale phenomena, and wavelets and PDE methods in image processing and computer vision.

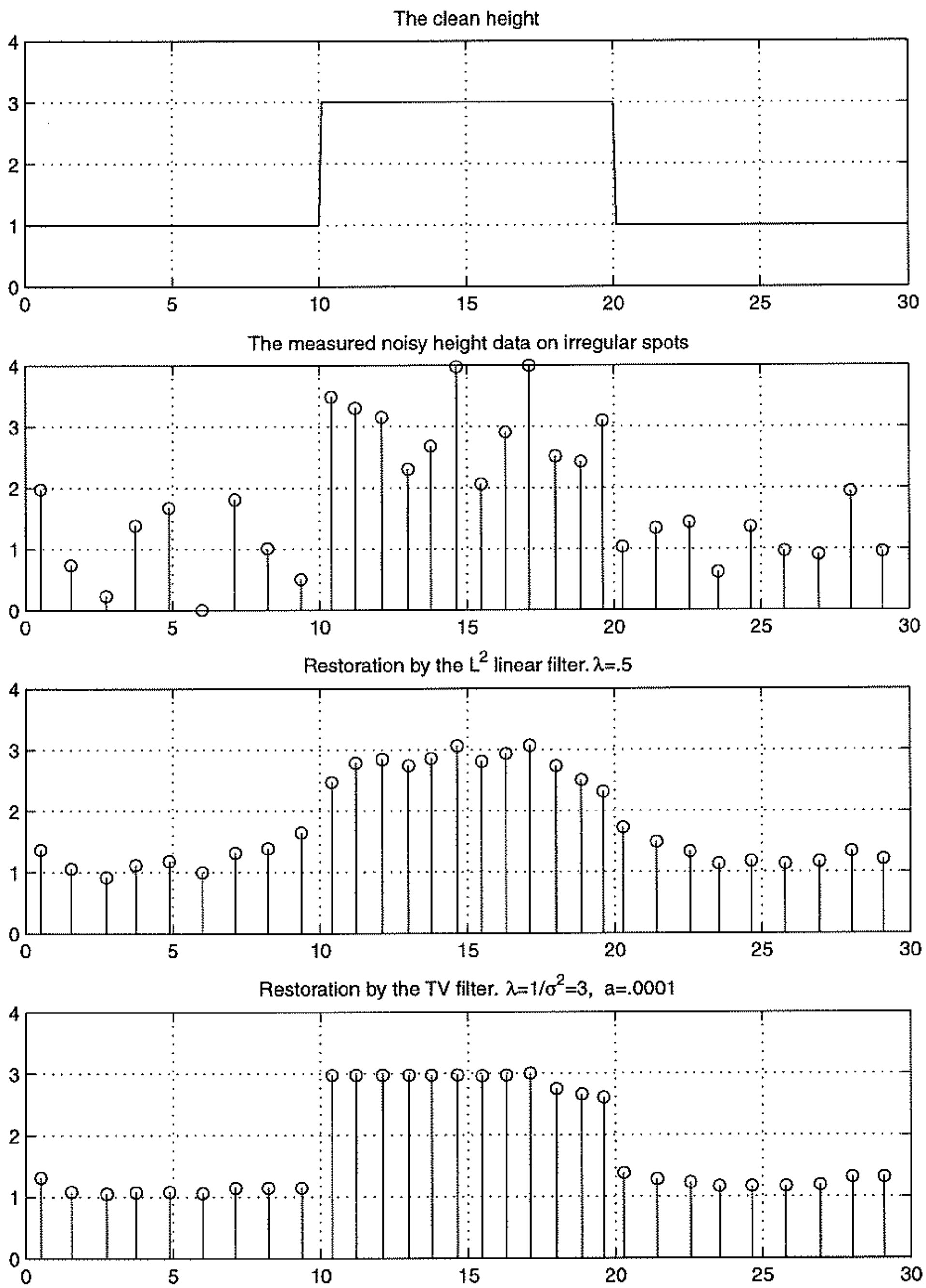


Fig. 2. The TV filtering on a 1-dimensional signal (see Section VI-A).

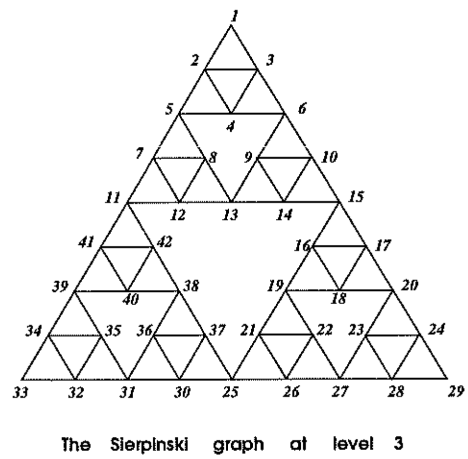
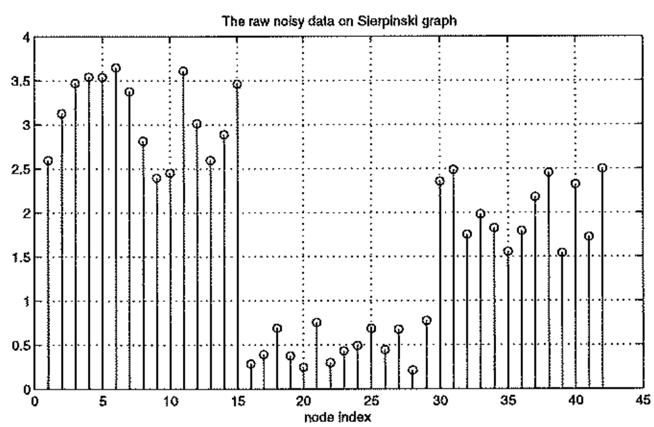


Fig. 3. The 42 nodes of the Sierpinski graph at level 3 (see Section VI-B).



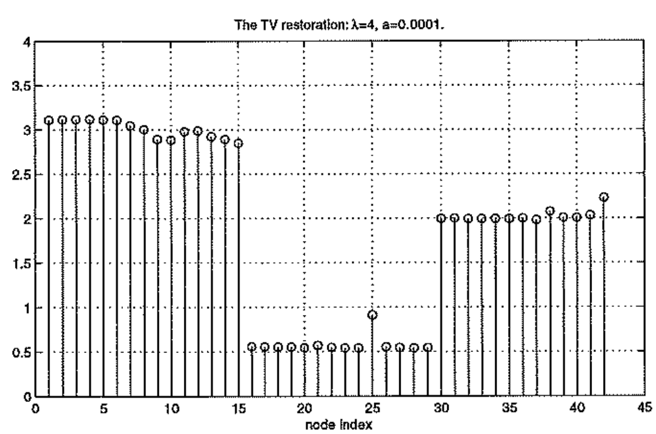


Fig. 4. The restoration result of the digital TV filter on the Sierpinski graph.

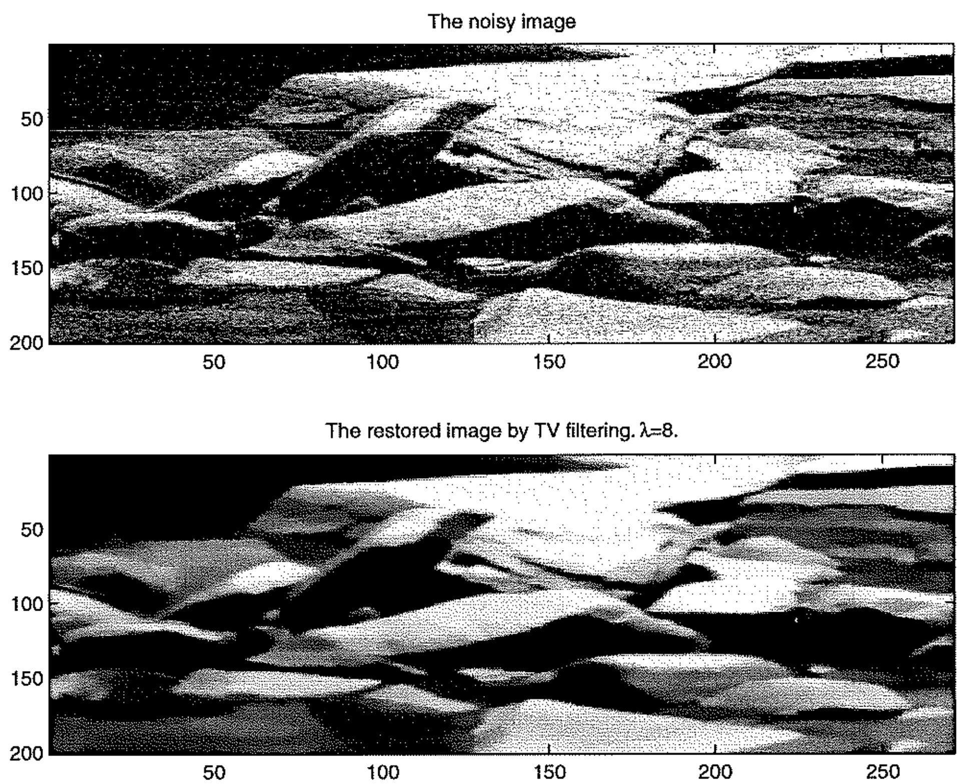


Fig. 5. The digital TV filter applied to a noisy color image (see Section VI-C).

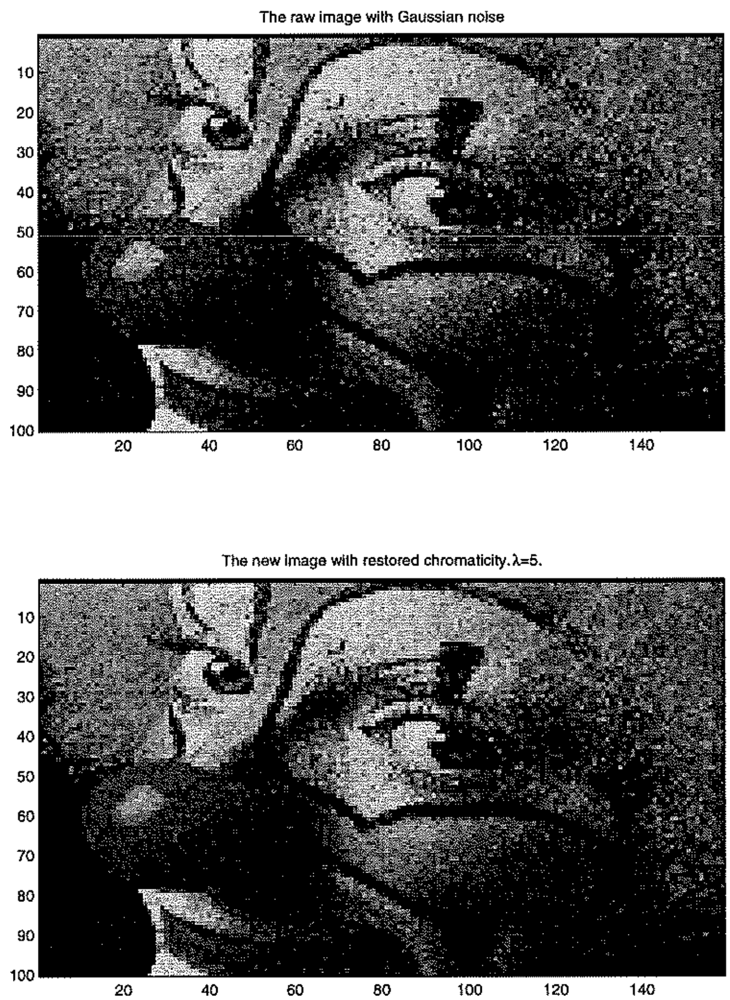


Fig. 6. The digital TV filter applied to chromaticity restoration (see Section VI-D).