A Quick Tour of Wavelets and PDE Techniques in Image Processing *

Hao-Min Zhou † Tony F. Chan ‡ Jianhong Shen §

Article Outline

Glossary

Wavelets

Wavelets are selected functions that generate orthonormal bases of the square integrable function space L^2 by using dilation and translations. The base functions have certain locality, such as compact support or fast decay property. And they are usually organized according to different scales or resolutions, which are called Multi-Resolution Analysis (MRA). Fast wavelet transforms are filtering procedures that compute the inner products of any given function and wavelet. The inner products are wavelet coefficients.

Digital Images

Digital images are usually referred as n dimensional data arrays recorded by optical or other physical devices, such as digital cameras, Radar, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), or generated by software in computer graphics. Most commonly seen digital images are 2 or 3 dimensional. When $n = 1$, they are conventionally called signals. And when $n > 3$, they are often hyper-spectral images.

Image Restoration

Image restoration rebuilds clearer images from given images that are disturbed or polluted in acquisition and transmission. Most commonly seen restoration tasks are denoising and deblurring. Denoising is to remove random perturbations to individual pixel values. Deblurring is a deconvolution procedure to restore pixel values and clean their neighbor's influence.

*Research supported in part by grants ONR-N00014-06-1-0345, NSF CCF-0430077, CCF-0528583, DMS-0610079, DMS-0410062 and CAREER Award DMS-0645266, NIH U54 RR021813, and STTR Program from TechFinity Inc.

†School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332. email:hmzhou@math.gatech.edu.

‡Department of Mathematics, the University of California, Los Angeles, CA90095-1555. email:chan@math.ucla.edu.

§Barclays Capital, Wall St. NY 10166, email: jackieneoshen@gmail.com.

Image Compression

Compression converts images from n dimensional data arrays into "0" and "1" bit streams so that they can be stored more efficiently. There are two types of compression, lossy or lossless compression depending on whether information is permanently lost or recoverable respectively. Many of the commonly used compression algorithms, such as the ones used by international image compression standards JPEG and JPEG2000, are transform based compression which consists of three basic steps: transform pixel values into frequency coefficients, quantization of the frequency coefficients, and coding to convert them into bit streams.

Image Segmentation

Segmentation divides images into subregions (segments) in which images have homogeneous or similar features.

Image Inpainting

Inpainting is an artistic word used to fill in damaged regions such as scratches in pictures, films and paintings. Digital image inpainting automatically fills in the missing regions in digital images automatically according to their neighboring information.

1 Definition of the Subject and Its Importance

It is no doubt that today's world has been experiencing the explosion of information in nearly all aspects of modern society, sciences and technologies. Visualization is one of the most direct and preferable manners to observe information carried by data, which is often massive in size and contains unwanted disturbance. To better reveal the information, specially when it is hidden, data has to be processed. For this goal, image processing, which includes many different tasks such as compression, restoration, inpainting, segmentation, pattern recognition and registration, has played a critical role. Historically, it has been viewed as a branch of signal processing, thus many classical methods are adopted from traditional Fourier based signal processing algorithms. In the past few decades, numerous new competing methods have emerged. Among them, wavelets, variational and PDE techniques and stochastic methods have demonstrated outstanding performance due to their special properties. For instance, wavelets have become the dominant tool in image processing because of their multiresolution structure, energy concentration ability and fast transform algorithms. The popularity of variational PDE techniques is driven by their extraordinary properties in understanding and manipulating geometrical features. Those new methods have been providing new avenues, some revolutionary, to make observations, understandings and discoveries in sciences and technologies. Many of them have been enjoying great success in different applications in medical, physical sciences, engineering and people's everyday life.

2 Introduction

Digital image processing analyzes or extracts certain information from digital images, which are often viewed as 2 or multi-dimensional data sets in mathematics. Each element in the data sets is called a pixel. Typical image processing tasks include segmentation, restoration, pattern recognition, analysis, compression, registration and motion detection [40, 46]. Image processing has a wide range of applications including communication, computer vision, acoustic, satellite imaging, medical and industrial diagnosis and many more.

Image processing tasks often require large scale computations, mainly due to the large amount of data to be processed. A typical gray scale still image with moderate resolution, such as 1024×1024, has over a million pixels. The size of a color image is three times larger than a gray scale image with the same resolution. A video sequence usually consists of over 24 color frames per second with each frame being a still image. A multi-spectral image contains a collection of several (usually more than 3) monochrome images of the same scene, each of them

taken with a different wavelength by a different sensor. In addition, many applications, such as airport screening and unmanned vehicle navigation, require real time response. All these demand efficient and reliable algorithms.

Traditional image processing methods are mainly based on Fourier/wavelets or statistical approaches. The best example is the current international image compression standards, JPEG and JPEG2000, which are largely based on discrete cosine transform (DCT) and wavelet transforms. For this reason, more images are stored using their wavelet coefficients. The great success of wavelets in image processing is built on their good properties, including multiresolution data structures, fast transform algorithms and superb energy concentration ability, which allows to approximate functions (images) using only a relative small number of coefficients.

Thousands of researchers have devoted their efforts to the development of wavelet theory, analysis, and algorithms in different applications. Groundbreaking contributions include Meyer's wavelet theory [53], Daubechies' compact support orthogonal wavelets [32], Mallat's multiresolution analysis [49, 50], Shapiro's progressive zero tree image coding algorithm [63], and many other works cited in books such as [27, 33, 43, 51], and [64].

Roughly speaking, wavelet transforms can express any square integrable functions by superpositions of wavelet basis functions, which are generated by dilation and translations from a few, if not a single, wavelet functions. The summation coefficients are called wavelet coefficients, which are standard L^2 inner products between wavelets and the given functions. Wavelet transforms are realized by filtering procedures. Usually, wavelet coefficients are classified into two types: low or high frequencies. Low frequency coefficients correspond to certain kinds of weighted local averages of the data values. High frequency coefficients are related to certain order derivatives. Therefore, high frequency coefficients are small for smooth functions and large for functions containing discontinuities.

In applications, it is inevitable that some of the wavelet coefficients, especially the high frequencies, are not available for intentional or involuntary reasons. For instance, in wavelet based image compression, insignificant (smaller in magnitude) high frequency coefficients are discarded on purpose to save more storage space. In lossy channel communication, coefficients are lost or damaged during the transmission due to unwanted disturbances. Obviously, with incomplete wavelet coefficients, one cannot re-compose the exact original functions. Many problems arise. One that has drawn the most attention is that oscillations are generated near discontinuities. This is the famous Gibbs' phenomenon in mathematics and edge artifacts in image processing.

Several directions have been taken to improve the performance of wavelet based image processing methods by reducing the Gibbs' oscillations, and by better preserving geometrical information in images. One strategy is using nonlinear thresholding procedures to allocate more storage resource to significant coefficients. Well-known examples include translation invariant denoising methods [30], wavelet hard thresholding, and wavelet shrinkage (also called soft thresholding) [36].

Another direction is building new geometry friendly wavelet-like multiresolution representations, such as ridgelets [8], curvelets [9], beamlets [35], bandelets [57] and many more recent developments. By introducing geometry into the construction of multiresolution representations, it is expected the decompositions have better performance near discontinuities.

The third direction is to modify the existing wavelet transforms so that fewer large high frequency coefficients are generated near discontinuities. Thus, less information is truncated in the thresholding process. Many methods have been proposed, such as Harten's remarkable general multiresolution framework [41] and its recent developments [2], the adaptive lifting scheme [29], and the adaptive Essential Non-Oscillatory (ENO) wavelet transforms [24, 25]. Many recent contributions are collected in [65].

In a different direction, PDE techniques for image processing, pioneered by Mumford-Shah's segmentation functional [55], Rudin-Osher-Fatemi's Total Variation (TV) restoration [59], and Perona-Malik's anisotropic diffusion [58], have emerged more recently. Due to their outstanding properties on handling geometrical information, different variational PDE models and methods have been proposed and studied for a variety of image processing goals, such as affine scale space [62], fundamental equations for image processing [1], total variation image analysis [15], active contour for segmentation [12, 22], blind deconvolution [23], image interpolation and inpainting [4, 5, 17, 19, 52], restoration [16, 18], and compression [26, 37]. The field is significantly enriched and many books have been published in recent years, see [3, 20, 54, 56, 61, 66, 45] and references therein.

Given the developments in both wavelets and PDE techniques in image processing, it is natural to think of combining their advantages to gain more benefits in the applications, especially when geometrical features are important. Well designed wavelet PDE methods can retain the good properties of wavelets, such as multires-

olution and fast algorithms. Meanwhile, they are able to use PDE concepts, such as gradients, curvatures to capture, control and manipulate the geometrical information to achieve image processing goals in more systematic manners. There are quite a few examples that have demonstrated the combined advantages in different applications [10, 14, 26, 28, 38, 48].

In this paper, it is not our intention to give a complete survey on either wavelets or PDE techniques in image processing. Instead, we will focus on a recent trend that combines them together. To be self-contained, we start with a brief introduction to wavelets, and followed by PDE techniques in image processing. We hope to use selected topics based on our experience to help readers, specially the beginners, to know some basic models and a few commonly used methodologies on the subject. The rest of the paper is arranged as following. Section 3 is a brief introduction to wavelets and their applications in image processing. Section 4 presents some well known PDE models in image processing. And we give some new developments of combining wavelets and PDE techniques in Section 5. A concise list of future directions is stated in the end.

3 Wavelets in Image Processing

Historically, Fourier decompositions, which express any given square integrable function by superpositions of sinusoidal functions, have been the major tool for image processing due to their efficient representations and fast Fourier transforms (FFT). This is particularly true for 1-D signals, such as audio sequences. However, all Fourier basis functions have global supports, which implies that any local change in the given function has to result in a global change in the representations. For this reason, Fourier basis are not efficient to represent local information, such as discontinuities. The well-known Gibbs' phenomena is an exhibition of this limitation. Unfortunately, most salient features such as edges and corners in images are local and discontinuous. Thus, all Fourier based methods for image processing suffer from the ringing artifacts.

Facing this shortcoming, it is highly desirable to have efficient representations which can better handle local information, specially discontinuities. Or more precisely, the basis functions should have local support or fast decay properties so that any local perturbation will only cause changes in a small neighborhood but not to far away places. To a certain extend, wavelets are designed to fill up this expectation, and have gained unsurpassed success in many applications of image processing.

After several decades of intensive studies, wavelets have been developed into a very rich mathematical theory. There are many different types of wavelets such as Meyer's wavelets, spline wavelets, and bi-orthogonal wavelets. Here, we present a very brief introduction based on Daubechies' compact supported wavelets, and their connections to compression and denoising.

3.1 Wavelets

Wavelets can be viewed as orthonormal bases of the square integrable function space $L^2(\mathbf{R})$. It starts with carefully selected scaling function $\phi(x)$ and corresponding wavelet $\psi(x)$ defined on finite support $[0, l]$, where l is a positive integer. We refer to [33] for the detailed selection procedure for $\phi(x)$ and $\psi(x)$. Many commonly used software such as MATLAB has built-in routines for the scaling and wavelet functions already.

The functions $\phi(x)$ and $\psi(x)$ satisfy the dilation equations (also called two-scale relations or refinement equations in some literature):

$$\phi(x) = \sqrt{2} \sum_{s=0}^{l} c_s \phi(2x - s), \tag{1}$$

and

$$\psi(x) = \sqrt{2} \sum_{s=0}^{l} h_s \phi(2x - s), \tag{2}$$

where the c_s's and h_s's are constants called low and high pass filters respectively. To give examples, the famous Haar wavelet selects

$$\phi(x) = \begin{cases} 1 & x \in [0, 1) \\ 0 & \text{otherwise} \end{cases},$$

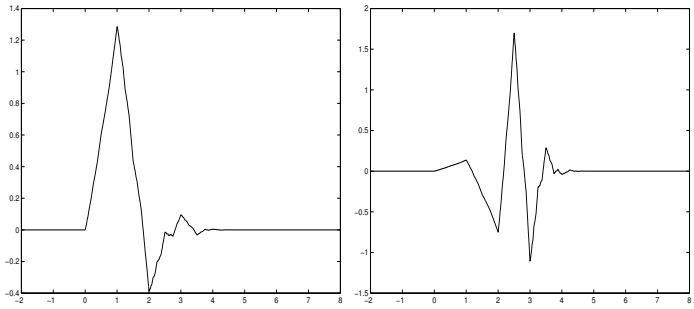

Figure 1: Left: The scaling function for Daubechies-6 wavelet. Right: The corresponding wavelet

and

$$\phi(x) = \begin{cases} 1 & x \in [0, \frac{1}{2}) \\ -1 & x \in [\frac{1}{2}, 1) \\ 0 & \text{otherwise} \end{cases},$$

which are step functions. We also plot the scaling and wavelet functions of Daubechies-6 in Figure 1.

Using dilation and translation, one can form families of functions from $\phi(x)$ and $\psi(x)$ as following,

$$\phi_{j,k}(x) = 2^{\frac{j}{2}}\phi(2^j x - k), \tag{3}$$

and

$$\psi_{j,k}(x) = 2^{\frac{j}{2}}\psi(2^j x - k), \tag{4}$$

where (j, k) are integers. Then the collection of $\psi_{j,k}(x)$ form an orthonormal basis of $L^2(\mathbf{R})$. This means that for any given function $f(x) \in L^2(\mathbf{R})$, one has

$$f(x) = \sum_{j,k} < f(x), \psi_{j,k}(x) > \psi_{j,k}(x), \tag{5}$$

where $< \cdot, \cdot >$ denotes the standard $L^2(\mathbf{R})$ inner product defined by

$$< f(x), g(x) >= \int_{\mathbf{R}} f(x)g(x)dx.$$

There are many desirable properties for the scaling functions and wavelets. Among them, locality and oscillations are most cited common features in all wavelets. Literally speaking, they make wavelets behave like localized small waves, which also explains the origination of the name.

The locality, which often refers to compact support or fast decay properties, enables wavelets to decompose or approximate functions locally. This meets the desire of many applications, particularly in image processing.

A good mathematical way to describe the oscillatory nature of wavelets is to use their vanishing moment property, which means

$$\int \psi(x)x^j dx = 0, \quad j = 0, 1, \cdots, p-1, \tag{6}$$

where p is a positive integer. In this case, the wavelet $\psi(x)$ is said to have p vanishing moments. The more vanishing moments, the more oscillations in wavelets in general.

Locality and oscillation together have been the main driving engineers for the success of wavelets in many applications.

3.2 Multi-Resolution Analysis

The success of wavelets also relies on their connection to Multi-Resolution Analysis introduced by Mallat [49, 50].

Consider the subspace of $L^2(\mathbf{R})$ defined by the scaling function $\phi_{j,k}(x)$,

$$V_j = \text{span}\{\phi_{j,k}(x), k \in \mathbf{Z}\},$$

for every fixed j. The dilation equation (1) implies that the subspaces form an ordered chain,

$$\cdots \subseteq V_{j-1} \subseteq V_j \subseteq V_{j+1} \subseteq V_{j+2} \cdots, j \in \mathbf{Z},$$

which also satisfies

$$\overline{\lim_{j\to\infty} V_j} = L^2(\mathbf{R}), \lim_{j\to-\infty} V_j = 0.$$

Here, larger indexes j correspond to finer resolutions or scales.

Similarly, one can define the subspaces generated by wavelets $\psi_{j,k}(x)$,

$$W_j = \text{span}\{\psi_{j,k}(x), k \in \mathbf{Z}\}.$$

The dilation equation (2) implies the following connection between W_j and V_j,

$$V_j = V_{j-1} \oplus W_{j-1}, j \in \mathbf{Z}. \tag{7}$$

Therefore $L^2(\mathbf{R})$ can be decomposed into,

$$L^2(\mathbf{R}) = V_J \oplus \sum_{j>J} W_j = \sum_{j=-\infty}^{\infty} W_j,$$

where J is an arbitrary reference resolution level. Consequently, $f(x) \in L^2(\mathbf{R})$ can be decomposed into a multi-resolution representation as,

$$f(x) = \sum_k \alpha_{J,k}\phi_{J,k}(x) + \sum_{j>J,k} \beta_{j,k}\psi_{j,k}(x), \tag{8}$$

where $\alpha_{j,k} =< f(x), \phi_{j,k}(x) >$ is called a low frequency (or scaling) coefficient, and $\beta_{j,k} =< f(x), \psi_{j,k}(x) >$ is a high frequency (or wavelet) coefficient. Without causing confusion, we call them wavelet coefficients for simplicity in this paper.

The decomposition (7) and the dilation equations (1), (2) lead to the following filtering and down-sampling procedures to compute the coarser scale wavelet coefficients from the finer scale coefficients,

$$\alpha_{j,k} = \sum_{s=0}^{l} c_s \alpha_{j+1,2k+s}, \tag{9}$$

and

$$\beta_{j,k} = \sum_{s=0}^{l} h_s \alpha_{j+1,2k+s}. \tag{10}$$

These are the famous fast wavelet transforms.

Apparently, fast wavelet transforms involve only the coefficients and can be started if one knows the low frequency coefficients $\{\alpha_{I,k}\}$ on a certain fine resolution I. Then, it is natural to ask how to obtain $\{\alpha_{I,k}\}$. Theoretically, $\{\alpha_{I,k}\}$ must be computed by $< f(x), \phi_{I,k}(x) >$ according to the definition. However, they are often replaced by the point-wise values $f(x_k)$ in practice, even though such an action is called a wavelet crime in [64]. The replacement makes sense when the function $f(x)$ is smooth and the resolution I is fine enough, because the low frequency coefficients $\alpha_{I,k}$, which are the weighted local averages of $f(x)$, are very close approximations to the point-wise values.

The above described wavelet transforms are for 1-D functions. Wavelet transforms for 2-D images are realized by simple tensor product in practice. More precisely, 2-D transforms are obtained by performing column-wise 1-D transforms followed by row-wise 1-D transforms.

3.3 Wavelet Thresholding and Image Processing

The wavelet representations (8) provide a mechanism to approximate functions in a multi-resolution fashion. For instance, the j-th scale (resolution) approximation is simply defined as

$$f_j(x) = \sum_k \alpha_{j,k}\phi_{j,k}(x) = \sum_k \alpha_{J,k}\phi_{J,k}(x) + \sum_{J<i<j,k} \beta_{i,k}\psi_{i,k}(x). \tag{11}$$

This multi-resolution approximation satisfies a standard error bound,

$$\|f(x) - f_j(x)\| \leq C2^{-jp}\|f^{(p)}(x)\|, \tag{12}$$

where C is a constant independent of j. It is obvious that the error is controlled by the vanishing moment p, the norm of the p-th derivative of $f(x)$, and the resolution j. Better approximations with more detailed information can be easily obtained by adding more terms for the finer resolutions. Many have argued that this convenient zoom-in and zoom-out multi-resolution approximation is by far one of the best mathematical models that mimic human perceptions.

In addition to the multi-resolution structure, the success of wavelet decomposition in image processing also depends on the sparsity of the wavelet coefficients. Simple integration by parts can show that high frequency wavelet coefficients satisfy $|\beta_{j,k}| = |f^{(p)}(x)|O(2^{-jp})$, which suggests that the wavelet coefficients are small if the function $f(x)$ is smooth enough. For images containing many smooth regions, such as most of the natural scenery images, it is easy to observe that a large number of the high frequency coefficients are insignificant, and therefore can be ignored in applications. Thresholdings are mathematical procedures to realize this observation.

Loosely speaking, thresholding is setting selected wavelet coefficients to be zero. There are many different types of thresholdings. In fact, the j-th scale approximation is one of them. It is constructed by ignoring all the scales higher than the given resolution j. This is often called linear thresholding, because the procedure is linear. Other nonlinear data dependent thresholdings, including commonly used hard and soft thresholdings, can achieve much better performance in image processing.

The hard thresholding simply sets any wavelet coefficients whose magnitudes are smaller than a given tolerance ϵ to be zero, i.e.

$$\bar{\beta}_{j,k} = \begin{cases} \beta_{j,k} & |\beta_{j,k}| > \epsilon \\ 0 & |\beta_{j,k}| \leq \epsilon \end{cases}.$$

Similar formula holds for the low frequency coefficients too.

The soft wavelet thresholding is slightly different from the hard thresholding. It is a shrinkage procedure. In addition to setting the coefficients whose magnitudes are smaller than the tolerance to zero, it reduces the magnitudes of other coefficients by ϵ as well,

$$\bar{\beta}_{j,k} = \begin{cases} sign(\beta_{j,k})(|\beta_{j,k}| - \epsilon) & |\beta_{j,k}| > \epsilon \\ 0 & |\beta_{j,k}| \leq \epsilon \end{cases},$$

where the $sign(\cdot)$ is the signum function.

The selection of the threshold ϵ has also been investigated by many groups. Among many proposed strategies, Donoho-Johnstone's $SQTWOLOG$ [36] and Stein's unbiased risk estimate have been widely used.

Thresholding procedures have accomplished remarkable success in image processing, specially in compression. It is easy to understand that wavelet thresholdings are useful in this application because one does not have to store the coefficients that are zero. However, it is more subtle in practical compression schemes. The problem is that not only one needs to remember the non-zero wavelet coefficients, but also their locations. The location information may occupy more storage space than the coefficients if they are recorded in a naive way. Shapiro's zero tree scheme [63] introduces a tree structure for wavelet coefficients based on their multiresolution property. A branch of the tree can be represented by a single bit '0' if all coefficients in the branch are zero. This is used in conjunction with thresholdings to achieve very efficient compression. Many well known state-of-the-art compression methods, such as Set Partitioning in Hierarchical Trees (SPIHT) [60] and Group Test Wavelet (GTW) [44] compression algorithms are based on the zero tree idea.

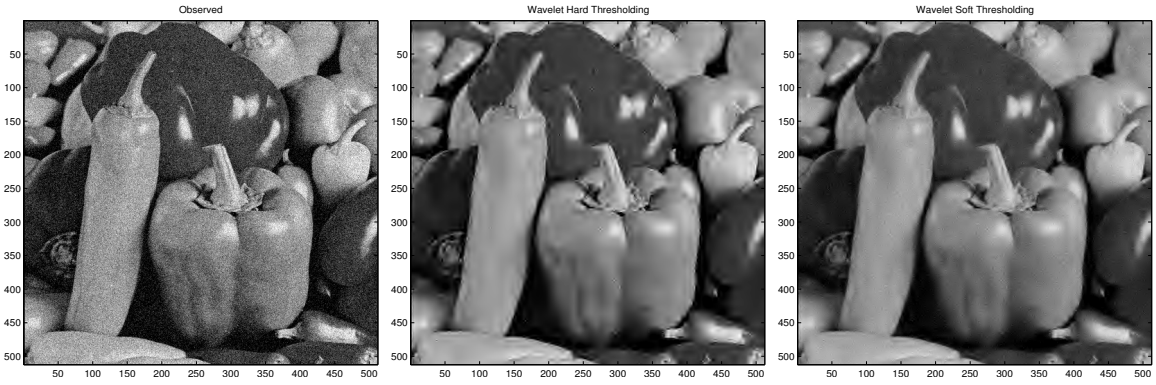

Figure 2: Left: Test image corrupted by white noise. Middle: Denoised image by hard thresholding. Right: Denoised image by soft thresholding

Simple thresholdings also provide fast and effective methods for noise removal. They have found many successful applications in communications, military and medical images. In Figure 2, we display the denoising effects of wavelet hard (middle) and soft (right) thresholdings of a test image with additive white noise (left).

From a mathematical point of view, the success of thresholdings can be explained by their connections to optimizations. It has been shown that many thresholding results are optimal in a certain sense. In layman's words, those thresholding results are best under certain criteria. For example, let us assume that the hard thresholding reconstruction

$$\bar{f}(x) = \sum_k \bar{\alpha}_{J,k}\phi_{J,k}(x) + \sum_{j,k} \bar{\beta}_{j,k}\psi_{j,k}(x)$$

has M nonzero wavelet coefficients. Then $\bar{f}(x)$ is the minimizer of the following optimization problem,

$$\min_g \|f - g\|_2, \quad \text{subject to} \quad g \text{ has at most } M \text{ nonzero wavelet coefficients.}$$

This leads to the conclusion that the hard thresholding gives the best M-term approximation in $L^2(\mathbf{R})$ among all possible combinations.

In a more general setting as discussed in [14] and [67], it is proved that the soft thresholding gives the minimizer of the following optimization problem

$$\min_g \{\|f - g\|_2 + 2\epsilon\|g\|_{B_1^1(L^1)}\},$$

where $B_1^1(L^1)$ is a Besov space. And the linear thresholding gives an approximate minimizer of the following optimization problem,

$$\min_g \{\|f - g\|_2 + 2\epsilon\|g\|_{W^m(L^2)}\},$$

where $W^m(L^2)$ is a Sobolev space. We refer readers to [14] for a detailed discussion.

4 PDE Techniques

Compared to wavelets, modern PDE techniques in image processing have appeared more recently, even though some traditional image processing methods can be interpreted from PDE perspective. For instance, the classical Gaussian filter for image denoising is accomplished by convolving the noisy image u_0 with the Gaussian kernel (also called heat kernel in literature) $G(x,t) = \frac{1}{2\pi t}\exp(-\frac{x^2}{2t})$,

$$u = G * u_0 = \int u_0(y)G(x-y,t)dy. \tag{13}$$

This denoised image u is actually the solution $u(x,t)$ of the following diffusion PDE,

$$u_t(x,t) = D\Delta u(x,t), \quad u(x,0) = u_0(x), \tag{14}$$

where Δ is Laplace operator, and $D = 1/2$ is diffusive coefficient.

Modern PDE techniques have drawn great attention and reached remarkable success in the past two decades. This is due to their extraordinary ability of handling geometrical features, which are lacking in traditional statistical or Fourier/wavelet based approaches. Two different strategies are commonly used to design PDE techniques for different image processing goals.

1 Construct PDE based evolution processes and incorporate geometry in the equations.

2 Pose image processing tasks in variational framework and derive corresponding Euler-Lagrange equations to compute the minimizers.

In both strategies, image processing goals are achieved by solving PDE's. Next, we use a few well-known examples to demonstrate these two strategies.

4.1 Anisotropic Diffusion for Denoising

Image denoising removes unwanted disturbances in images. Very often, those disturbances, such as white noise and pepper-and-salt noise, are highly localized and oscillatory. This makes it harder to separate noise from edges which are also local and discontinuous. As an anti-oscillation procedure, diffusion is a natural selection for denoising. As mentioned earlier, the classical Gaussian filter for denoising is equivalent to the linear isotropic diffusion (14). However, it has been observed in both experimental and theoretical studies that isotropic diffusion unavoidably smears sharp edges, corners and other geometrical features embedded in u_0 while filtering out noise. This is because it treats all orientations identically and never recognizes the presence of spatially coherent discontinuities – edges. In addition, the larger the diffusive coefficient D, the quicker the smoothing out.

To remedy this drawback, Perona-Malik [58] proposed using anisotropic diffusion instead,

$$u_t = \nabla \cdot (D(x,u,\nabla u)\nabla u). \tag{15}$$

The diffusivity coefficient D is data dependent and must sense the existence of edges, so that the PDE stops diffusion across the discontinuities. For this purpose, it is desirable to have D satisfying the following requirements,

$$D = \begin{cases} \text{large}, & \text{when } |\nabla u| \text{ is small on intra-regions,} \\ \text{small}, & \text{when } |\nabla u| \text{ is large near edges.} \end{cases} \tag{16}$$

Therefore, the evolution only smooths out the oscillations away from edges but not across them . In [58], D is selected as

$$D = g(|\nabla u|^2),$$

where g is a smooth positive concave function satisfying $g(+\infty) = 0$. For examples, g can be taken as

$$g(|\nabla u|^2) = e^{-\frac{|\nabla u|^2}{2\sigma^2}},$$

or

$$g(|\nabla u|^2) = \frac{1}{1 + b|\nabla u|^2},$$

where σ and $b > 0$ are constants.

In practice, the anisotropic diffusion (15) must face a challenge on how to compute the coefficient D robustly. This may be troublesome especially in the beginning of the diffusion process when u_0 contains highly oscillatory noise, because $|\nabla u|$ is large almost everywhere so D is small everywhere. Thus, the diffusion is not effective in removing noise. To overcome this difficulty, using a mollified image in g has been proposed in [13], which takes the form as

$$u_t = \nabla \cdot (g(|\nabla(G_\sigma * u)|^2)\nabla u), \quad u(x,0) = u_0(x),$$

where G_σ is again the Gaussian kernel.

Along the lines of anisotropic diffusion, much more research has been done including the well known general axiomatic scale-space theory in [1]. We refer readers to [20, 66] for more discussion.

4.2 Total Variation Image Denoising

A different viewpoint for denoising is to reduce the uncorrelated local oscillations in images. Mathematically speaking, total variation (TV) is a quantity that measures oscillations in functions. It is intuitive that oscillatory noise greatly increases the TV norm. Naturally, one can think denoising as reducing the total variations of images. In fact, this observation leads to the famous TV model proposed by Rudin-Osher-Fatemi [59],

$$\min_u \int |\nabla u| dx \quad \text{subject to} \quad \|u - u_0\|_2 \leq \sigma, \tag{17}$$

where σ is related to the noise level. The objective functional is to reduce oscillations in the reconstruction, and the constraint term is a fitting requirement. This optimization problem can be read as to find the least oscillatory image within a small ball of radius σ centered at the noisy image u_0.

The model is often re-formulated as a non-constraint minimization problem as

$$\min_u \int |\nabla u| dx + \frac{\lambda}{2}\|u - u_0\|_2, \tag{18}$$

where $\lambda \geq 0$ is a Lagrange multiplier, which is the factor that balances the competition between oscillations and fidelity. The smaller the λ, the fewer details in the denoised images. In extreme situations, the solution for (18) is a flat constant when λ is zero, or is the noisy image u_0 when λ is infinite.

The most outstanding advantage of TV denoising model (18) is that it allows sharp edges being preserved in the reconstruction. This implies that TV model has the ability of reducing small oscillations (noise) but not penalizing the edges. This feature has been well understood in the context of computational fluid dynamics (CFD), specially in shock capturing, where TV semi norm is intensively used. In fact, the authors of [59] are also experts in CFD and it is no doubt that (17) is inspired by numerical shock capturing.

Another attraction of TV denoising is its geometrical properties. For functions with finite TV semi norms, this can be seen clearly through an equivalent coarea formula,

$$\int |\nabla u\| dx = \int_{-\infty}^{+\infty} \int_{\{u=\gamma\}} ds d\gamma.$$

Here the term $\int_{\{u=\gamma\}}$ is the length of the level set $\{u = \gamma\}$. The TV semi norm is obtained by integrating along all level contours of $\{u = \gamma\}$ for all values of γ. This suggests that TV semi norm controls both the size of the jumps, and the geometry of the level sets.

The geometric connection of TV minimization is more visible if we analyze the optimization (18) by calculus of variation. The standard theory shows that the minimizer must satisfy the following Euler-Lagrange equation,

$$-\nabla \cdot (\frac{\nabla u}{|\nabla u|}) + \lambda(u - u_0) = 0. \tag{19}$$

The first term, the functional derivative of TV semi norm, is precisely the curvature of the image, which makes the method more geometric friendly. For noisy pixels, the jumps are isolated and their curvature is large. They will be wiped out much quicker than the edges that are coherent jumps with relatively smaller curvature.

The best known, not necessary the most efficient, algorithm to solve (18) is the gradient descent method, which introduces an artificial time to form an evolution PDE,

$$u_t = \nabla \cdot (\frac{\nabla u}{|\nabla u|}) - \lambda(u - u_0). \tag{20}$$

Compared to (15), the gradient descent of TV minimization (20) is also an anisotropic diffusion with a degenerate diffusive coefficient $D = 1/|\nabla u|$. And it satisfies the anisotropic diffusion requirement (16). In particular, if a

edge is sharp, D will be zero and no diffusion is performed across the edge. In practice, to prevent numerical blow-up caused by $|\nabla u| = 0$ in the denominator, it is often replaced by $\sqrt{|\nabla u|^2 + \epsilon}$, where ϵ is a small positive number. Actually, this replacement can be derived from variational framework too.

An interesting and natural question is why one wants to use TV semi norm in (18) instead of $\int |\nabla u|^2 dx$, which is the famous Sobolev H_1 semi norm in PDE's. In fact, a very similar calculation can show that the H_1 minimization leads to exactly the isotropic diffusion (14), which loses the geometrical properties.

4.3 Variational Models for Image Segmentation

The purpose of image segmentation is to divide an image into regions within which the image has similar features, such as intensity values, texture pattern, or belonging to same objects. Segmentation is a crucial building block for many high level image processing and vision tasks such as object detection, recognition, and tracking. Obviously, one image may produce different partitions because of different segmentation criteria. This non-uniqueness nature, which is also true for many other image processing tasks, makes the segmentation problem very challenging. There is extensive literature on the subject and many methods have been proposed using different strategies. For example, the celebrate intensity-edge mixture model is statistically based [39], the widely used active contour (also called snake) model [47] uses variational framework. We take the well known Mumford-Shah segmentation model [55] and Chan-Vese region based active contour model [22] as examples to demonstrate how mathematical formulations and computational strategies can contribute to segmentation.

The original Mumford-Shah segmentation model is stated in a variational format,

$$\min \lambda_1 \int_\Gamma ds + \frac{\lambda_2}{2} \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \frac{\lambda_3}{2} \int_\Gamma (u - u_0)^2 dx, \tag{21}$$

where λ_1, λ_2 and λ_3 are three constants, u is the partitions with different segments. Ω is the region where the image is defined and Γ is the interior boundary separating different segments. The first term is the length of the interior boundary curves. The second term is isotropic diffusion within each homogeneous region. Similar to the TV minimization model (18), the third term is the fitting term. From formulation (21), the segmentation is achieved by balancing the competitive three terms. Different ratios among λ_1, λ_2, and λ_3 give different partitions.

Mumford-Shah model has many desirable properties and is very general. Many other known models can be viewed as special cases of it. However, it also faces serious computational challenges because the partition boundary Γ is unknown. And thus the first term involving line integral along Γ has no easy way to compute. To ease the challenges, many other models are proposed for better computation properties. Among them, Chan-Vese's *active contour without edge* model [22] has gained remarkable success due to its simplicity and robustness.

Assume that C is a closed curve partitioning the segments. The model is designed to move C so that the following energy is minimized,

$$\min \lambda_1 \cdot \text{Length}(\Gamma) + \lambda_2 \cdot \text{Area}(inside(\Gamma)) + \lambda_3 \int_{inside(\Gamma)} (u_0 - c_1)^2 dx + \lambda_4 \int_{outside(\Gamma)} (u_0 - c_2)^2 dx, \tag{22}$$

where $\lambda_1, \lambda_2, \lambda_3$ and λ_4 are positive fixed parameters. This model uses piecewise constant approximations inside and outside the partition curves. If one picks $\lambda_2 = 0$, (22) becomes the minimal partition model which is a special case for (21).

Chan-Vese model (22) can also be formulated in a level set framework and it leads to a fast and robust computation method, which sparkles a large amount of follow-up researches in using level set based active contour methods for segmentations in different applications.

4.4 PDE Method for Image Inpainting

Image inpainting, or its mathematical synonym image interpolation, fills in missing or damaged image regions based on known surrounding information. It is a very fundamental problem having numerous prior work in existence. It also shares common ground with many other image processing tasks such as image replacement,

error concealment, edge completion and image editing. Here we only use 1) a third order nonlinear inpainting PDE by Bertalmio et al. [6], 2) a variational inpainting model by Chan-Shen [19] as two examples to illustrate how modern mathematics are used for this traditional labor intensive task, because image inpainting used to be done by hands.

Similar to segmentation, image inpainting is an inverse problem having possible multiple solutions. It is obvious that when information is missing, different people may have different ways to patch different information to the regions. And all of them may look reasonable. However, it is commonly agreed that the inpainted regions must have consistent geometrical features and texture patterns with their surroundings. For this reason, many of the inpainting methods are based on geometrical interpolations or extrapolations. One example is the remarkable third order inpainting PDE introduced by Bertalmio et al. [6]. In fact, the term *image inpainting* was used first by them and the work has stimulated a wave of interest in inpainting related problems.

The PDE is given as

$$u_t = \nabla(\Delta u) \cdot \nabla^{\perp} u, \tag{23}$$

where $\nabla^{\perp}$ is the orthogonal gradient direction (isophote direction as called in the original paper). The idea behind (23) is a brilliant intuition of information transport along broken level lines (isophotes). The PDE is solved only inside the inpainting regions with proper boundary conditions based on the gray values and isophote directions. It is discovered later that (23) actually connects to the famous Navier-Stokes equations in CFD [5].

The Chan-Shen's inpainting model tackles the problem from a different angle. It starts with a variational principle inspired by TV restoration model [59],

$$\min_u \int_\Omega |\nabla u| dx + \frac{\lambda}{2} \int_{\Omega \setminus D} (u - u_0)^2 dx, \tag{24}$$

where D is the inpainting region. Similar to (17), a straightforward interpretation of this model is that the minimizer u is the least oscillatory image that is close enough to the given image u_0 outside the inpainting region. For the regions inside of D, the restored u has no restriction except matching its surroundings in a least oscillatory fashion. This model can lead to a nonlinear data dependent PDE similar to (19) and can be solved numerically. The results are impressive and much follow-up work has been performed to analyze the model and extend it to include more sophisticated measurements such as Euler's Elastica into consideration for better curve treatments [17].

5 Wavelet Based Variational PDE Methods

As discussed in the previous sections, both wavelets and PDE techniques have been used extensively in image processing and achieved tremendous success in numerous applications. Their success is based on different properties of both approaches. Wavelets have multi-resolution data structure, energy concentration (sparsity) and fast algorithms. PDE techniques are geometrically friendly and often tied to variational principles. A closer look at both approaches can easily reveal that those properties do not overlap, and one cannot be used to replace the other. In a certain sense, they are complementary to each other. It seems natural to combine the advantages of both to gain benefits. In fact, many research efforts have been put forward along this direction.

There are two different strategies that have been explored to merge PDE techniques with wavelets,

(a) Use computational PDE skills to modify the standard wavelet transforms to form new transforms having better geometric properties.

(b) Design new wavelet based variational models for different image processing tasks.

In this section, we will select a few examples to demonstrate both strategies.

5.1 ENO-Wavelet Transforms

As mentioned earlier in Section 3, it is well-known that Fourier based algorithms suffer Gibbs' oscillations. Wavelets can remarkably reduce the severity of oscillations due to their locality. But they still exist unless

one retains all discontinuity related coefficients, which is not practical in many applications. To improve the image quality, one needs to reduce the oscillations by lowering the threshold ϵ in thresholding procedures. As a consequence, more coefficients, specially edge related ones, must be retained. This is why a majority of the storage is allocated to edge related coefficients in JPEG2000.

To further improve the performance and reduce the ringing artifacts, it is desirable to design wavelet-like transforms so that fewer significant high frequency coefficients are generated. One way to achieve this goal is to incorporate geometrical features in the design of wavelet-like basis (or redundant frame) functions so that discontinuous functions can be more effectively represented. Many efforts have been proposed such as curvelets [9], beamlets [35], bandelets [57] to name a few. A different way is to reduce the wavelet filter length. This is based on the fact that the larger the supports of wavelets, the longer the filters, and the worse the Gibbs' oscillations. Then, one can adaptively use shorter wavelet filters near the vicinities of edges. The adaptive lifting scheme proposed in [29] uses this idea.

ENO-wavelet transforms approach the problem from a different angle. It is inspired by the original Harten's multi-resolution framework [41], which has a profound impact on many new methods in the field. ENO-wavelet transforms borrow a key idea, the one side interpolation strategy, from ENO schemes for shock capturing [42]. Different from the fore-mentioned methods which adapt the filters or basis functions to better fit the data, ENO-wavelet transforms change the data near edge areas and feed them into the same standard wavelet filters. The data is changed in a special way so that the filters do not see the discontinuities.

Let us imagine that we filter around a jump discontinuity. A high frequency coefficient is large if the high pass filter is convolved with data across the jump. However, one can extend the data from both left and right sides in smooth ways and feed the extended data to the filters. Then the high frequency coefficients are small as the high pass filter only sees the smooth data on both sides. Of course, one may immediately question that near the jump region, we have actually two different pieces of data overlapping in the area. In fact, this is a serious issue, it causes a double storage problem, which means we have doubled the number of wavelet coefficients in the jump region. And this is directly against the goals of many image processing tasks, specially image compression. Fortunately, the problem can be avoided by a strategy called coarse level extrapolation, which extends the data in such ways that some of the jump related wavelet coefficients are predictable and do not need to be memorized. And the storage can be reduced to the same as that of standard wavelet transforms. We refer to [24] for detailed ENO-wavelet transform algorithms. Here we just point out the main ideas and some important results.

ENO-wavelet transforms can be used as functional replacements of standard wavelet transforms. Indeed, ENO-wavelet transforms perform standard wavelet transforms if no discontinuity is detected. ENO-wavelet transforms retain the essential properties and advantages of standard wavelet transforms such as energy concentration, multiresolution framework and fast transform algorithms, all without any edge artifacts. They also achieve uniform approximation accuracy up to the discontinuities. If $\hat{f}_j(x)$ is the j-th resolution approximation to $f(x)$ by using ENO-wavelet transforms, then

$$\|\hat{f}_j(x) - f(x)\| \leq C2^{-jp}\|f^{(p)}(x)\|_{\Omega\setminus\Gamma}, \tag{25}$$

where Γ is the set of discontinuous points. It is worth noting that the error (12) for standard wavelet transforms depends on the p-th derivative of $f(x)$ on the entire region Ω, which is unbounded if the discontinuous set Γ is not empty. In contrast, the error for ENO-wavelet transforms (25) depends on $f^{(p)}(x)$ only on the domain Ω excluding Γ. This ensures that ENO-wavelet transforms perform uniformly accurate regardless of the presence of discontinuities. That is probably the best result one may expect. In Figure 3, we show a comparison between ENO-wavelets and standard wavelets.

5.2 Wavelet Based Minimal Energy Methods for Denoising

As discussed in Section 4, anisotropic diffusion and total variation minimization for image denoising have great capability to extract image features, specially edges for better image quality preservation. However, it is also commonly recognized that such PDE techniques often post higher computational demands, because numerical solutions for nonlinear PDE's need to be computed iteratively. To achieve reasonable solutions, many iterations must be performed. This has been a major criticism for PDE techniques, specially when one compares them with wavelets, which have ultra fast filtering algorithms.

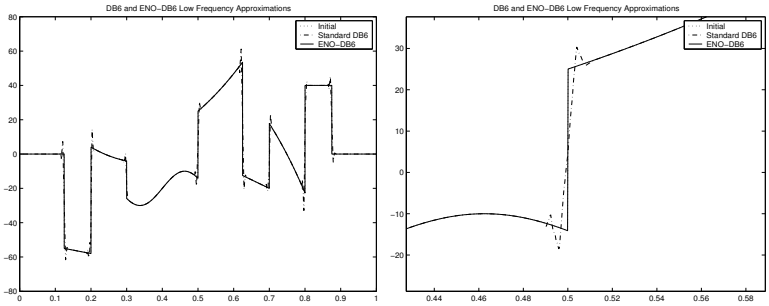

Figure 3: Left: The comparison between 4-level ENO-Daubechies-6 (solid line) and standard Daubechies-6 (dash-dotted line) approximations. Right: A zoom-in of picture on the left near a discontinuity. Standard Daubechies-6 generates oscillations near discontinuities, but the ENO-Daubechies-6 does not.

To retain the capability in feature extraction while eliminating the need of iterations for anisotropic diffusion, there have been efforts to formulate geometric friendly energy minimizations in wavelet spaces so that the minimizers can be obtained directly from wavelet coefficients without iterations. In fact, as discussed in Section 3.3, classical wavelet thresholdings, including linear, hard and soft thresholdings, have corresponding energy optimizations in certain functional spaces. But those minimization problems are not built to handle geometrical features. In [28], Chui-Wang suggested a new geometrical energy minimization in wavelet space given as

$$\min E_\lambda(\rho,\beta) = \lambda E_i(\rho,\beta) + \frac{1}{2}\|\beta-\alpha\|^2, \tag{26}$$

where $E_i(\rho,\beta)$ is a selected internal energy which can be expressed by the wavelet coefficients β. The second term is the standard L^2 fitting requirement. In their original paper, a blended internal energy is chosen as

$$E_i(\rho,\beta) = \sum_{j,k}(\rho(m_{j,k}(p)) + \rho(\beta^d_{j,k})), \tag{27}$$

where $m_{j,k}(p) = (|\beta^h_{j,k}|^p + |\beta^v_{j,k}|^p)^{1/p}$, and $\rho(s) = |s|$. In (27), the notations β^d, β^h and β^v are 2-D tensor product wavelet coefficients along diagonal, horizontal and vertical directions respectively. It is clear that the energy functional is resolution, orientation and spatial dependent. In this way, the energy functional can "see" the corners, edges in wavelet spaces because those geometrical structures create correlated wavelet coefficients along diagonal, horizontal and vertical directions respectively. When $p = 2$, the minimizers of (26) and (27) can be attained explicitly from the wavelet coefficients as

$$(\beta^*_\lambda)^{h,v}_{j,k} = (\beta^0)^{h,v}_{j,k}(1 - \frac{\lambda}{m^0_{j,k}})_+, \tag{28}$$

and

$$(\beta^*_\lambda)^{d}_{j,k} = \mathrm{sign}(\beta)^{0,d}_{j,k}(|\beta^{0,d}_{j,k} - \lambda|)_+, \tag{29}$$

where $(\cdot)_+$ denotes the nonnegative value function.

Along this line, there has been a recent trend in the computational harmonic analysis community to design data dependent nonlinear filters based on PDE techniques, for example, the adaptive digital TV filter presented in [18]. More recently, Chui and collaborators have proposed a new anisotropic filtering strategy based on ideas of finding approximate solutions of anisotropic diffusion equations discussed in Section 4.1. Their method realizes image denoising by one sweep of nonlinear filtering.

5.3 Diffusion Wavelets

Diffusion wavelets has been proposed by Coifman and collaborators in [31]. It is a different way to generalize classical wavelets using PDE and geometry concepts. The goal is to construct a multiresolution analysis

framework on general geometric structures, such as manifolds, graphs or even discrete point sets, so that image processing tasks can be performed for functions defined on these structures.

As discussed in Section 3.1, standard wavelet multi-resolutions are based on dilation and translations. However, this is often impossible for general data structures, specially when little geometric information is known. To overcome this difficulty, diffusion wavelets use dyadic powers of a diffusion operator T (with $\|T\| < 1$), such as the heat operator defined on the general data structure to create scales. The following two properties are crucial for constructing diffusion wavelets. One is that the spectral of high powers of T decay faster as the power gets higher. Consequently, one can use a few leading eigenfunctions of T^j (j large) to approximate the range spaces of T^j accurately.

The other property is that applying higher powers of T to local functions, such as Dirac delta functions defined on a point in a discrete data set produces smoother functions with wider supports, because T is a diffusion process. After a non-trivial process involving orthonormalization, which we refer to [31] for details, one can construct a multi-resolution analysis based on T^{2^j} ($j \in \mathbf{Z}^+$). Specially, once the multi-resolution analysis is formed, T^{j^2} can be expressed in a highly compressed format.

There are many potential applications, such as in data mining and learning theory. Here, we pick the following simple example to illustrate their usage. Let us consider computing the inverse of Laplacian $(I-T)^{-1}$ applied to an arbitrary vector f defined on a general data structure. The operator $(I-T)^{-1}$ is a deblurring process commonly seen in image restorations. It is well known that

$$(I-T)^{-1} f = \sum_{k=1}^{+\infty} T^k f.$$

Define

$$S_K = \sum_{k=1}^{2^K} T^k,$$

then an approximation to $(I-T)^{-1}$ can be achieved by

$$S_{K+1} f = (S_K + T^{2^K} S_K) f = \prod_{k=0}^{K} (I + T^{2^k}) f.$$

Since the powers T^{2^k} have been compressed in the multi-resolution analysis and can be efficiently applied to f, $(I-T)^{-1} f$ is computed efficiently.

5.4 TV Wavelet Inpainting

Wavelet inpainting, or more generally wavelet interpolation, refers to the problem of filling in missing or damaged wavelet coefficients due to lossy image transmission or communication. Obviously, the task is closely related to classical inpainting problems as discussed in Section 4.4, but also differs remarkably in that the inpainting regions are in the wavelet domain.

Working in the wavelet domain, instead of the pixel domain, changes the nature of the inpainting problem, since damages to wavelet coefficients can create correlated damage patterns in the pixel domain. For instance, there usually exists no corresponding regular geometric inpainting regions, which are however necessary for many PDE-based inpainting models in pixel domains. Such lack of spatial geometric regularity of inpainting regions also prohibits many other existent inpainting techniques applied to pixel domains. On the other hand, direct interpolation in the wavelet domain is also problematic, because wavelet coefficients are constructed to be uncorrelated in the L^2 sense and neighboring coefficients provide minimum information to the missing ones. In addition, degradation in wavelet inpainting problems is often spatially inhomogeneous, which demands different treatments in different regions.

A closer exam may find that all these new challenges are actually caused by a simple fact: Damage happens in the wavelet domain while human perception prefers to see images with certain regularity in pixel domain.

Therefore, it seems natural to create wavelet inpainting methods by filling in the coefficients in wavelet domain while controlling the regularity in the pixel domain. TV wavelet inpainting models presented in [21] exactly follow this strategy.

Two different models have been proposed based on the noise level in images. The first one is for noiseless images, in which the retained coefficients are considered to be correct and will not be alerted.
Model I:

$$\min_{\beta_{j,k}:(j,k)\in I} F(u,z) = \int_{\mathbf{R}^2} |\nabla_x u(\beta,x)|dx = \mathrm{TV}(u(\beta,x)), \tag{30}$$

where I is the inpainting index region in wavelet domain, and $u(\beta,x)$ has the wavelet transform:

$$u(\beta,x) = \sum_{j,k} \beta_{j,k}\psi_{j,k}(x).$$

For noisy images, since every coefficient is also corrupted by noise. Then one has to modify (denoise) them too.
Model II:

$$\min_{\beta_{j,k}} F(u,z) = \int_{\mathbf{R}^2} |\nabla_x u(\beta,x)|dx + \sum_{(j,k)} \lambda_{j,k}(\beta_{j,k}-\alpha_{j,k})^2, \tag{31}$$

and the parameter $\lambda_{(j,k)}$ is zero if $(j,k)\in I$; otherwise, it equals a positive constant λ.

Clearly, these two models are inspired by TV denoising model for their exceptional ability of handling geometries. Both models recover the wavelet coefficients so that the restored images are least oscillatory while matching the known information. The key difference is that the arguments are restricted to the inpainting regions I only for Model I, so the dimension of unknowns is the number of coefficients in I. While in Model II, the parameter λ is taken to be zero in the inpainting regions I in the wavelet domain, in contrast to the standard denoising and compression models, where λ is usually taken to be a constant everywhere. This difference essentially puts no constraint on the missing wavelet coefficients so that they can change freely. In Figure 4, we show an example of wavelet inpainting by the two models.

5.5 Compressive Sampling

Compressive sampling [7], also goes by the name compressed sensing [34], is an emerging theory addressing the sampling problem in image and signal processing. In information theory, the classical Shannon-Nyquist sampling theorem states that "Exact reconstruction of a continuous-time baseband signal from its samples is possible if the signal is bandlimited and the sampling frequency is greater than twice the signal bandwidth". More precisely, bandlimited signals refer to functions whose Fourier frequencies are in a bounded interval. And the sampling theorem says that a bandlimited signal can be fully reconstructed from its evenly spaced samples, provided that the sampling rate must exceed twice the maximum frequency in the bandlimited signal. This rate is often called Nyquist rate.

Compressive sampling considers a different scenario. In the simplest case, let us assume that a signal $f(t)$ is sparse in the frequency space or any other convenient spaces. For instance, the signal consists of only a few Fourier terms, i.e.

$$f(t) = \sum_{j=1}^{n} \beta_j e^{-ik_j t},$$

where n is an integer much smaller than the signal resolution N, $\{k_j\}$'s are frequencies, and i the imaginary unit. But the actual values of frequency k_j are not known. Obviously, $f(t)$ is a bandlimited signal. Without loss of generality, we assume k_n is the highest frequency. Then Shannon-Nyquist sampling theory requires at least $2k_n$ evenly spaced observations to exactly reconstruct $f(t)$. However, since the value k_n is not known, the actual number of samples (resolution N) needed may be much higher than $2k_n$.

Compressive sampling asks a different question: Can one exactly recover a sparse signal $f(t)$ using a small number of samples $\{f(t_j)\}_1^m$ observed at randomly selected time t_j $(j=1,\cdots,m)$? Here m may be much smaller

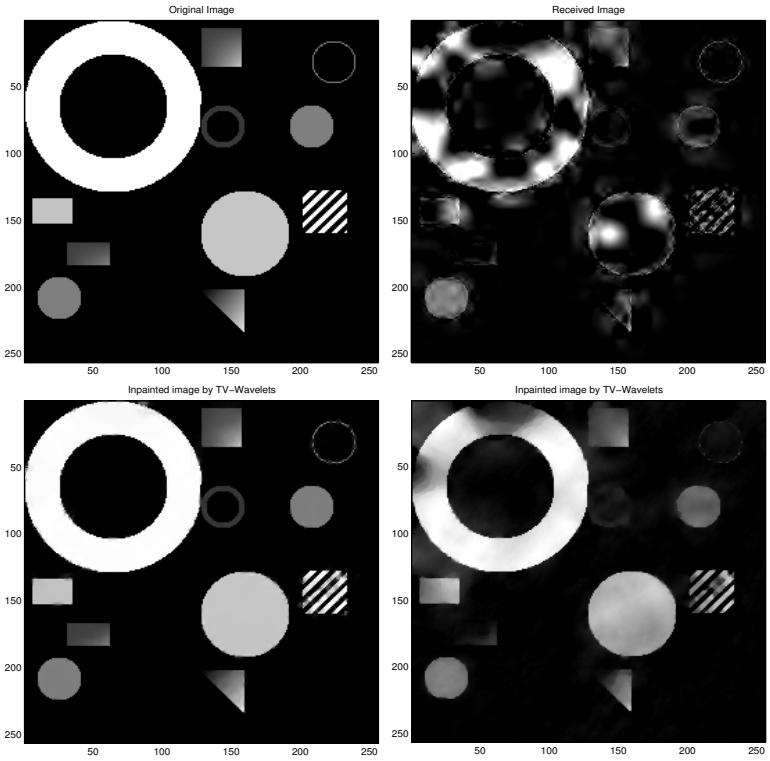

Figure 4: Upper-Left: Original synthetic image. Upper-Right: 50% of the wavelet coefficients are randomly lost, including some low frequency coefficients, which results in large damaged regions in the pixel domain. Notice that there are no well defined inpainting regions in the pixel domain. Lower-Left: Restored image by Model I, Lower-Right: Restored image by Model II. They not only fill in missing regions properly, but also restore the sharp edges and geometrical shapes.

than N. Given the knowledge of sparsity, ideally one can convert this problem into the following minimization problem,

$$\min \|\beta\|_{l_0}, \quad \text{subject to} \quad (F^*\beta)(t_j) = f(t_j), \tag{32}$$

where β is a vector containing the Fourier coefficients of the reconstructed signal, F is the Fourier matrix, and $\|\cdot\|_{l_0}$ is the l_0 norm of a discrete sequence, which is defined as the number of nonzero elements. Then F^* gives the inverse Fourier transform, and $(F^*\beta)$ is the reconstructed signal. l_0 minimization (32) finds the sparsest reconstruction $(F^*\beta)$ among all possible functions that agree with the observations $f(t_j)$.

However, l_0 minimization (32) is essentially a large non-convex integer optimization problem, which is computational prohibitive. Then compressive sampling suggests that it is still possible to exactly recover $f(t)$ from the samples $\{f(t_j)\}_1^m$. The exact reconstruction is realized by the following l_1 optimization,

$$\min \|\beta\|_{l_1}, \quad \text{subject to} \quad (F^*\beta)(t_j) = f(t_j). \tag{33}$$

In other words, the compressive sampling achieves exact recovery by finding the signal having the smallest l_1 norm in frequency space among all signals matching the sample values $f(t_j)$ at t_j.

There are many reasons to select l_1 norm in the optimization, including remarkable mathematical insights given in [11]. We will not intend to present their results here. Instead, we list the following two reasons that are more intuitive and may explain the essence of l_1 optimization in sparse recovery.

1 l_1 norm exhibits interesting sparsity in many applications. In other words, l_1 minimization in frequency space intends to drive more frequencies to zero.

2 l_1 norm has the least index p among all l_p norms that are convex.

The convex property ensures that (33) may be solved efficiently by the standard convex optimization algorithms.

It is believed that compressive sampling may have many implications. One of the most attractive potentials is that it suggests the possibility of new data acquisition protocols that translate analog information into digital form with fewer sensors than what was considered necessary. There are many interesting studies on how to advance the theory and applications, and even design new hardware to realize the implications. We refer to [7] for more information on the subject.

It is worth noting that even though TV wavelet inpainting and compressive sampling are developed independently, there is an interesting connection between them. For instance, the derivative of a piecewise constant image, ∇u, can be viewed as sparse in the pixel space. If one makes measurements in the wavelet space, then Model I (30) is the l_1 minimization of the derivative in the pixel space with constraints in the wavelet space, which fits well in the framework of compressive sampling. In this sense, they are complementary to each other, and can be viewed as dual formulations.

6 Future Directions

Driven by rapidly developing imaging sciences and technologies, the last couple of decades have witnessed the tremendous success of wavelets and PDE techniques in mathematical image processing. Many researchers have been working in the field and exciting new developments are constantly reported. However, compared to the even faster growing demands, there still is a large distance to meet the ever-increasing expectations. The following is just a very short list of directions that are or shall be pursuing in the near future.

(1) Developing more sophisticated models, methods to better preserve features for images, or general data sets in higher dimensions, such as video or hyper-spectral images. Merging traditional wavelets and PDE techniques seems to be promising along this direction. For example, developing wavelets and PDE models for segmentation is interesting. To our knowledge, it has not been attempted yet.

(2) New applications in high level vision, such as pattern recognition, auto navigation and tracking, which demands better understanding of the problems and more accurate extraction of the connections among data sets.

(3) Robust and efficient implementation strategies to compute the solutions of mathematical image processing models, specially those involving solutions of nonlinear PDE's.

The list is based on the authors' experience and reflects our own perspective. Certainly it does not cover all aspects of this large field. Interested readers are encouraged to read up-to-date literature to follow the latest advancements on the subject.

References

[1] L. Alvarez, F. Guichard, P. L. Lions and J. M. Morel, *Axioms and Fundamental Equations of Image Processing*, Arch. Rational Mechanics and Anal., 16 (1993), pp. 200-257.

[2] F. Arandiga and R. Donat, *Nonlinear Multiscale Decompositions: The approach of A. Harten*, Numerical Algorithms 23 (2000) 175-216.

[3] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculas of Variations*, Appl. Math. Sci., vol 147, Springer-Verlag, 2001.

[4] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro and J. Verdera, *Filling-in by Joint Interpolation of Vector Fields and Grey Levels*, IEEE Trans. Image Processing, 10(8), 2001, 1200-1211.

[5] M. Bertalmio, A.L. Bertozzi, and G. Sapiro. *Navier-Stokes, fluid dynamics, and image and video inpainting*, 2001 IEEE Conference on Computer Vision and Pattern Recognition. Dec. Kauai, Hawaii.

[6] M. Bertalmio, G. Sapiro, V. Caselles and C. Ballester, *Image Inpainting,* Tech. Report, ECE-University of Minnesota, 1999.

[7] E. Candés, *Compressive Sampling*, Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006.

[8] E. Candés and D. Donoho, *Ridgelets: a Key to Higher-dimensional Intermittency?*, Phil. Trans. R. Soc. Lond. A(1999).

[9] E. Candès and D. Donoho, *Curvelets: A Surprisingly Effective Nonadaptive Representation of Objects with Edges*, Tech. Report, Dept. of Stat., Stanford Univ., 1999.

[10] E. Candès, J. Romberg and T. Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information*, IEEE Trans. Inform. Theory, 52, pp489-509.

[11] E. Candès, and T. Tao, *Near Optimal Signal Recovery From Random Projections and Universal Encoding Strategies*, IEEE Trans. Inform. Theory, 52, pp 5406-5425.

[12] V. Caselles, R. Kimmel and G. Sapiro, *On Geodesic Active Contours*, Int. Journal of Computer Vision, 22(1), 1997, 61-79.

[13] F. Catte, P.L. Lions, J. M. Morel and T. Coll, *Image Selective Smoothing and Edge Detection by Nonlinear Diffusion*, SIAM J. Numer. Anal. 29, 1992, pp182-193.

[14] A. Chambolle, R. DeVore, N. Lee and B. Lucier, *Nonlinear Wavelet Image Processing: Variational Problems, Compression, and Noise Removal Through Wavelet Shrinkage*, IEEE Tran. Image Proc., Vol. 7, No. 3, Mar. 1998, pp. 319-333.

[15] A. Chambolle, P. L. Lions, *Image Recovery via Total Variational Minimization and Related Problems,* Numer. Math., 76, 1997, pp. 167-188.

[16] T. F. Chan, G. H. Golub, and P. Mulet, *A Nonlinear Primal-Dual Method for Total Variation-Based Image Restoration*, in ICAOS'96, 12th International Conference on Analysis and Optimization of Systems: Images, Wavelets, and PDEs, Paris, June 26-28, 1996, number 219 in Lecture Notes in Control and Information Sciences, pp. 241-252.

[17] T. F. Chan, S. H. Kang and J. Shen, *Euler's Elastica and Curvature Based inpainting* SIAM J. Appl. Math., 63(2) (2002), 564-592.

[18] T. F. Chan, S. Osher, and J. Shen, *The Digital TV Filter and Nonlinear Denoising*, IEEE Trans. Image Process., 10(2), pp. 231-241, 2001.

[19] T. F. Chan and J. Shen, *Mathematical Models for Local Non-Texture Inpainting*, SIAM J. Appl. Math., 62(3) (2002), 1019-1043.

[20] T. F. Chan and J. Shen, *Image Processing and Analysis - Variational, PDE, Wavelet and Stochastic Methods*, SIAM Publisher, Philadelphia, 2005.

[21] T. F. Chan, J. Shen, and H. M. Zhou, *Total Variation Wavelet Inpainting*, J. of Math. Imaging and Vision, Vol. 25, No. 1, 2006, pp 107-125.

[22] T. F. Chan and L. Vese, *Active Contour Without Edges* IEEE Transactions on Imge Processing, 10(2), Feb 2001, pp. 266-277.

[23] T. F. Chan and C. K. Wong, *Total Variation Blind Deconvolution*, IEEE Trans. Image Processing, 7 (1998), pp. 370-375.

[24] T. F. Chan and H. M. Zhou, *ENO-wavelet Transforms for Piecewise Smooth Functions*, SIAM J. Numer. Anal., Vol 40, No. 4 (2002), 1369–1404;

[25] Tony F. Chan and H. M. Zhou, *ENO-wavelet Transforms and Some Applications*, in the book *Beyond Wavelets*, editored by J. Stoeckler and G. V. Welland, Academic Press (2003).

[26] T. F. Chan and H. M. Zhou, *Optimal Constructions of Wavelet Coefficients Using Total Variation Regularization in Image Compression*, CAM Report, No. 00-27, Dept. of Math., UCLA, July 2000.

[27] C. K. Chui, *Wavelet: A Mathematical Tool for Signal Analysis*, SIAM, 1997.

[28] C. K. Chui and J. Wang, *Wavelet-based Minimal-Energy Approach to Image Restoration*, Applied. Comp. Harmonic Anal. Vol 23, Issue 1, 2007, pp 114-130.

[29] P. Claypoole, G. Davis, W. Sweldens and R. Baraniuk, *Nonlinear Wavelet Transforms for Image Coding*, Correspond. Author: Baraniuk, Dept. of Elec. and Comp. Sci., also Submit to IEEE Tran. on Image Proc., Preprint, 1999.

[30] R. Coifman and D. Donoho, *Translation invariant de-noising*, in Wavelets and Statistics, A. Antoniadis and G. Oppenheim, eds., Springer-Verlag, New York, 1995, pp. 125–150.

[31] R. Coifman and M. Maggioni, *Diffusion Wavelets*, Appl. Comp. Harm. Anal. 11, June 2006.

[32] I. Daubechies, *Orthonormal Bases of Compactly Supported Wavelets*, Comm. Pure Appl. Math. 41(1988), pp909-996.

[33] I. Daubechies, *Ten Lectures on Wavelets*, SIAM 1992.

[34] D. Donoho, *Compressed Sensing*, Tech Report, Stanford University, 2004.

[35] D. Donoho and X. Huo, *Beamlets and multiscale image analysis*, Multiscale and Multiresolution Methods, Springer Lecture Notes in Computational Science and Engineering (Editors T.J. Barth, T. Chan, and R. Haimes), vol. 20, pp. 149-196, 2002.

[36] D. Donoho, I. Johnstone, *Adapting to Unknown Smoothness via Wavelet Shrinkage*, J. Amer. Stat. Assoc., Vol. 90, 1995, pp1200-1224.

[37] D. Dugatkin, H. M. Zhou, T. F. Chan and M. Effros, *Lagrangian Optimization of A Group Testing for ENO Wavelets Algorithm*, Proceedings to the 2002 Conference on Information Sciences and Systems, Princeton University, New Jersey, March 20 - 22, 2002.

[38] S. Durand and J. Froment, *Artifact Free Signal Denoising with Wavelets*, in Proceedings of ICASSP'01, volume 6, 2001, pp. 3685-3688.

[39] S. Geman and D. Geman, *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*, IEEE Trans. Patt. Anal. Mach. Intell. 6, 1984, pp721-741.

[40] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison Wesley, 1993.

[41] A. Harten, *Multiresolution Representation of Data, II. General Framework*, Dept. of Math., UCLA, CAM Report 94-10, April 1994.

[42] A. Harten, B. Engquist, S. Osher and S. Chakravarthy, *Uniformly High Order Essentially Non-Oscillatory Schemes, III*, Journal of Computational Physics, v71 (1987), pp.231-303.

[43] E. Hernandez and G. Weiss, *A First Course on Wavelets*, CRC Press, 1996.

[44] E. S. Hong and R. E. Ladner, *Group testing for image compression*, IEEE Proceedings of the Data Compression Conference, Snowbird, UT, March 2000, 2-12.

[45] *Special Issue on Partial Differential Equations and Geometry-Driven Diffusion in Image Processing and Analysis*, IEEE Tran. on Image Proc., Vol. 7, No. 3, Mar. 1998.

[46] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, New Jersey, 1989.

[47] M. Kass, A. Witkin and D. Terzopoulos, *Snakes: Active Contour Models*, Intl. J. Comput. Vision, Vol 1, 1987, pp 321-331.

[48] F. Malgouyres and F. Guichard, *Edge Direction Preserving Image Zooming: a Mathematical and Numerical Analysis*, SIAM, J. Num. Anal., 39:1, 2001, pp 1-37.

[49] S. Mallat, *Multiresolution Approximation and Wavelet Orthonormal Bases of* $L^2(R)$, Tran. Amer. Math. Soc. 315(1989), pp.69-87.

[50] S. Mallat, *A Theory of Multiresolution Signal Decomposition: The Wavelet Representation*, IEEE Trans. PAMI 11 (1989), pp. 674-693.

[51] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.

[52] M. Masnou and J. Morel, *Level-lines Based Disocclusion*, IEEE ICIP, October, 1998, 259-263.

[53] Y. Meyer, *Wavelets and operators*, Advanced mathematics, Cambridge University Press, 1992.

[54] J. M. Morel and S. Solimini, *Variational Methods in Image Segmentation*, Birkhauser, 1994.

[55] D. Mumford and J. Shah, *Optimal Approximation by Piecewise Smooth Functions and Associated Variational Problems*, Comm, Pure Appl. Math. 42, 1989, pp. 577-685.

[56] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Springer-Verlag, 2002.

[57] E. Le Pennec and S. Mallat, *Image Compression with Geometrical Wavelets*, in IEEE Conference on Image Processing(ICIP), Vancouver, September, 2000.

[58] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion,* IEEE T PATTERN ANAL. 12: (7), July, 1990, pp. 629-639.

[59] L. Rudin, S. Osher and E. Fatemi, *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D, Vol 60(1992), pp. 259-268.

[60] A. Said and W. Pearlman, *A new, fast, and efficient image codec based on set partitioning in hierarchical trees*, IEEE Trans. Circuits and Systems for Video Technology, Vol 6 (1996), No. 3, 243-250.

[61] G. Sapiro, *Geometric Partial Differential Equations and Image Processing*, Cambridge University Press, 2001.

[62] G. Sapiro and A. Tannenbaum, *Affine Invariant Scale-Space*, Internet. J. Comput. Vision, 11 (1993), pp. 25-44.

[63] J. Shapiro, *Embedded Image Coding Using Zerotrees of Wavelet Coefficients*, IEEE Trans. Signal Processing, Vol 41 (1993), No. 12, 3445-3462

[64] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1996.

[65] G. V. Welland (editor), *Beyond Wavelets*, Academic Press, 2003.

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

[67] J. Xu and S. Osher, *Iterative Regularization and Nonlinear Inverse Scale Space Applied to Wavelet Based Denoising*, UCLA CAM Report 06-11, March 2006.