Characterization of Signals from Multi scale Edges

Gowtham Bellala

Department of Electrical Engineering, University of Michigan,
Ann Arbor

1. INTRODUCTION

Points of sharp variations are often amongst the most important features for analyzing the properties of transient signals or images. In images, they are generally located at the boundaries of important image structures and hence are used to discriminate objects from their background in computer vision and pattern recognition applications. Edge detection is also used in detecting flaws in metal surface and in many more feature based image processing applications, hence has been the primary focus of the researchers over the last few decades.

The ability to reconstruct images from multi scale edges has many applications in signal processing. It allows us to process the image information with edge based algorithms. Characterizing images with edges will also lead to a good compression of the image, thereby reducing the memory requirements. The reconstruction of the algorithm from multi scale edges has mainly been studied in the zero crossing framework. But this has opened up the floor to issues regarding the uniqueness of the representation. There are known counter examples that prove that the positions of the zero crossings do not characterize uniquely the function.

Mallat hypothesized that for a complete and stable zero crossing representation, it is sufficient to record the positions where the wavelet transform of the signal has local extrema and the wavelet transform's values at these locations. This conjecture has been motivated by a reconstruction algorithm that is able to recover a close approximation of the original signal from the wavelet transform's local extrema positions and values.

In the next few sections, this algorithm has been explained and an application of the algorithm is demonstrated. Finally, the drawbacks of the algorithm and the cases where the algorithm fails have been discussed.

2. MULTISCALE EDGE DETECTION

Most multi scale edge detectors smooth the signal at various scales and detect sharp variations points from their first or second order derivative. The extrema of the first derivative correspond to the zero crossings of the second derivative and to the inflection points of the smoothed signal. The relationship of the wavelet transform to multi scale edge detected algorithms is explored here.

Any function $\theta(x)$ whose integral is equal to 1 and which converges to 0 at infinity is called a *smoothing function*. For example, a Gaussian function can be chosen as $\theta(x)$. Supposing that $\theta(x)$ is twice differentiable, define $\psi^a(x)$ and $\psi^b(x)$ as the first and second order derivative of $\theta(x)$

$$\psi^a(x) = \frac{d\theta(x)}{dx} \quad \text{and} \quad \psi^b(x) = \frac{d^2\theta(x)}{dx^2} \tag{1}$$

The functions $\psi^a(x)$ and $\psi^b(x)$ can be considered to be wavelets as their integral is equal to zero.

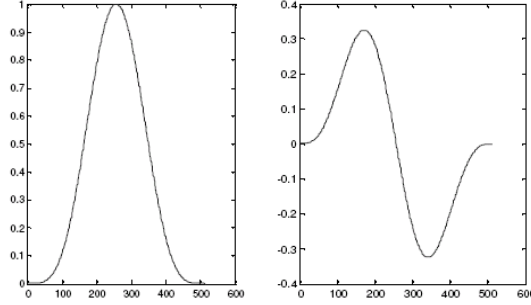$$\int_{-\infty}^{+\infty}\psi^a(x)dx = 0 \quad \text{and} \quad \int_{-\infty}^{+\infty}\psi^b(x)dx = 0$$

Figure 1: *A smoothing function and its corresponding wavelet*

A wavelet transform is computed by convolving the signal with a dilated wavelet. The wavelet transform of $f(x)$ at the scale s and position x, computed with respect to the wavelet $\psi^a(x)$, is defined by

$$W_s^a f(x) = f * \psi_s^a(x) \tag{2}$$

Similarly, the wavelet transform of $f(x)$ with respect to $\psi^b(x)$ is defined. Now, it can be shown that

$$W_s^a f(x) = f * (s\frac{d\theta_s}{dx})(x) = s\frac{d}{dx}(f * \theta_s)(x) \tag{3}$$

and a similar expression can be derived for the wavelet transform of $f(x)$ with respect to $\psi^b(x)$. Hence the wavelet transforms $W_s^a f(x)$ and $W_s^b f(x)$ are, respectively, the first and second derivative of the signal smoothed at scale s. Thus, the local extrema of $W_s^a f(x)$ correspond to the zero crossings of $W_s^b f(x)$ and to the inflection points of $f * \theta_s(x)$. Detecting zero crossings or local extrema are similar procedures, but the local extrema approach has some important advantages.

2.1   Dyadic Wavelet Transform in one dimension

A wavelet is a function $\psi(x)$ whose average is zero. The dilation of $\psi(x)$ by a factor 2^j is denoted by $\psi_{2^j}(x)$ where

$$\psi_{2^j}(x) = \frac{1}{2^j}\psi(\frac{x}{2^j})$$

The wavelet transform of $f(x)$ at the scale 2^j and at the position x is defined by the convolution product

$$W_{2^j}f(x) = f * \psi_{2^j}(x) \tag{4}$$

The sequence of functions $(W_{2^j}f(x))_{j\in Z}$ defined as in (4) is referred to as the *dyadic wavelet transform*. The Fourier transform of $W_{2^j}f(x)$ is given by the expression

$$\hat{W}_{2^j}f(\omega) = \hat{f}(\omega)\hat{\psi}(2^j\omega) \tag{5}$$

By imposing that there exists two strictly positive constants A_1 and B_1 such that

$$\forall \omega \in R, A_1 \le \sum_{j=-\infty}^{+\infty}\left|\hat{\psi}(2^j\omega)\right|^2 \le B_1 \tag{6}$$

it is ensured that the whole frequency axis is covered by dilations of $\hat{\psi}(\omega)$ by $(2^j)_{j \in Z}$ so that $\hat{f}(\omega)$, and thus, $f(x)$ can be recovered from its dyadic wavelet transform. The reconstructing wavelet $\chi(x)$ is any function whose Fourier transform satisfies

$$\sum_{j=-\infty}^{+\infty} \hat{\psi}(2^j \omega)\hat{\chi}(2^j \omega) = 1 \tag{7}$$

and the function $f(x)$ can be recovered from its dyadic wavelet transform with the summation

$$f(x) = \sum_{j=-\infty}^{+\infty} W_{2^j} f * \chi_{2^j}(x) \tag{8}$$

Using the Parseval Theorem, a norm equivalence relation can be derived from (5) and (6)

$$A_1 \|f\|^2 \leq \sum_{j=-\infty}^{+\infty} \left\| W_{2^j} f(x) \right\|^2 \leq B_1 \|f\|^2 \tag{9}$$

This proves that the dyadic wavelet transform is not only complete but also stable. If B_1/A_1 is closer to 1, it will be more stable. This dyadic wavelet transform is more than complete, it is redundant.

3.   ANALYSIS OF MULTI SCALE INFORMATION

The wavelet transform computes the derivative of the signal smoothed at different scales, but it is not clear as to how to combine these different values to characterize the signal variation. The wavelet theory gives an answer to this question by showing that the evolution across scales of the wavelet transform depends on the local Lipschitz regularity of the signal. A Lipschitz exponent is defined as follows

Definition 1: Let $0 \leq \alpha \leq 1$. A function $f(x)$ is uniformly Lipschitz α over an interval (a,b) if and only if there exists a constant K such that for any $(x_0, x_1) \in (a,b)$

$$|f(x_0) - f(x_1)| \leq K |x_0 - x_1|^\alpha \tag{10}$$

The relationship between the local Lipschitz regularity of the signal and the evolution of the wavelet transform across scales is given by the following theorem

Theorem 1: Let $0 < \alpha < 1$. A function $f(x)$ is uniformly Lipschitz α over (a,b) if and only if there exists a constant $K > 0$ such that for all $x \in (a,b)$, the wavelet transform satisfies

$$\left| W_{2^j} f(x) \right| \leq K (2^j)^\alpha \tag{11}$$

If the uniform Lipschitz regularity is positive, the condition implies that the amplitude of the wavelet transform modulus maxima should decrease when the scale decreases. On the contrary, the singularity at the abscissa 3 of Fig. 2 produces transform maxima that increases when the scale decreases. Such singularities can be described with negative a Lipschitz exponent, which means they are more singular than discontinuities.
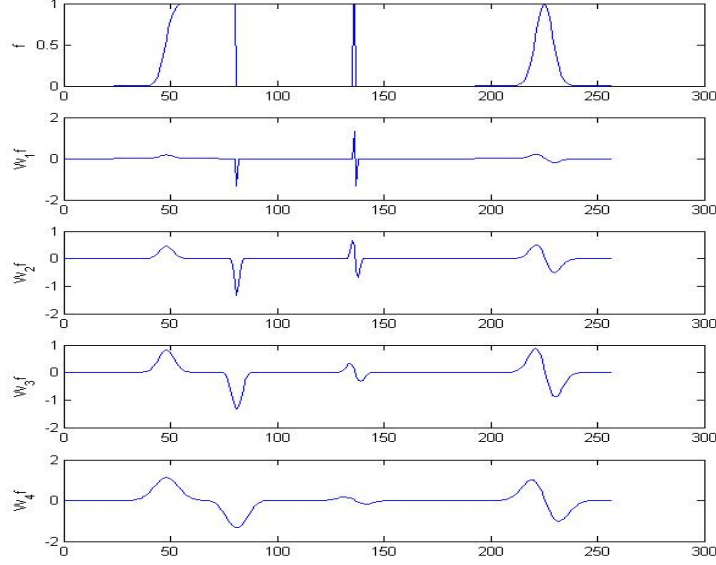
Figure 2: *The figure shows a signal with four sharp variation points having different Lipschitz regularity α and smoothing variance σ². The four plots below shows the wavelet transform of the signal at different scales.*

It can be noted from abscissa 2 of Fig. 2, that the maxima values of the wavelet transform remain constant over a large range of scales. It implies that the Lipschitz regularity α is equal to 0 at this point, which means that this singularity is a discontinuity. But, a signal is often not singular in the neighborhood of local sharp variations. An example is the smooth edge at the abscissa 1 of Fig. 2. It is generally important to estimate the smoothness of the signal variation in such cases. This is done by modeling the smooth variation as a singularity convolved with a Gaussian of variance σ².

$$f(x) = h * g_\sigma(x) \quad \text{with} \quad g_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{x^2}{2\sigma^2}) \tag{12}$$

Let $h(x)$ have a uniform Lipschitz regularity equal to α_o in the neighborhood of x_o. If the wavelet $\psi(x)$ is the derivative of a smoothing function $\theta(x)$, (3) proves that the wavelet transform of $f(x)$ can be written

$$W_{2^j} f(x) = 2^j \frac{d}{dx}(f * \theta_{2^j})(x) = 2^j \frac{d}{dx}(h * g_\sigma * \theta_{2^j})(x) \tag{13}$$

Suppose that the function $\theta(x)$ is close to a Gaussian function in the sense that the above equation can be rewritten as

$$W_{2^j} f(x) = 2^j \frac{d}{dx}(h * g_\sigma * \theta_{2^j})(x) \approx 2^j \frac{d}{dx}(h * \theta_{s_o})(x) = \frac{2^j}{s_o} W_{s_o} h(x) \tag{14}$$

This equation proves that the wavelet transform at the scale 2^j of a singularity smoothed by a Gaussian of variance σ^2 is equal to the wavelet transform of the non smoothed singularity $h(x)$ at the scale $s_o = \sqrt{2^{2j} + \sigma^2}$.

Equation (11) in Theorem 1 is valid for any scale $s > 0$. Hence

$$|W_s h(x)| \leq K s^\alpha \tag{15}$$

Inserting this inequality in (14)

$$|W_{2^j} f(x)| \leq K 2^j (s_o)^{\alpha-1} \tag{16}$$

This equation states that both Lipschitz exponent and the smoothness of the signal discontinuity can be retrieved from the wavelet transform modulus maxima at different scales. The conclusions are summarized in the following table.

α constraint	Meaning	Impact on Wavelet transform
$0 < \alpha <= 1$	f(x) is differentiable at x0. The change is gradual.	Amplitude decreases with scale.
$\alpha = 0$	F(x) is not differentiable at x0. A sharp change or discontinuity exists.	Amplitude remains the same across scales.
$-1 <= \alpha < 0$	F(x) is an impulse at x0.	Amplitude decreases with scale.

Figure 3: *Table summarizing conclusions drawn from Theorem 1*

4. SIGNAL RECONSTRUCTION FROM MULTI SCALE EDGES

4.1 Previous Results

The reconstruction of signals from multi scale edges has mainly been studies in the zero crossing framework. There are issues regarding the completeness of such a representation. In general, there are known counter examples that prove that the positions of the zero crossings of $W^b{}_s f(x)$ do not characterize uniquely the function *f(x)*. For example, the wavelet transforms of *sin(x)* and *sin(x) + 0.2sin(2x)* have the same zero crossings at all scales *s > 0*. A large family of such counter examples has been found by Meyer.

Mallat conjectured that to obtain a complete and stable zero crossing representation, it is sufficient to record the positions where $W^a f(x)$ has local extrema and the value of $W^a f(x)$ at the corresponding locations. A reconstruction algorithm based on this conjecture has been proposed.

4.2 Reconstruction Algorithm

Let $f(x) \in L^2(R)$. The algorithm reconstructs an approximation of the dyadic wavelet transform of *f(x)*, given the positions of the local maxima of $\left| W_{2^j} f(x) \right|$ and the values of $W_{2^j} f(x)$ at these locations. It is assumed that the wavelet $\psi(x)$ is differentiable in the sense of Sobolev, so that the wavelet transform of *f(x)* has at most, a countable number of modulus maxima. Let x^j_n be the abscissa where $\left| W_{2^j} f(x) \right|$ is locally maximum. Let *h(x)* be the reconstructed signal, then the maxima constraints on $W_{2^j} h(x)$ can be decomposed in two conditions.

1. At each scale 2^j, for each local maximum located at x^j_n, $W_{2^j} h(x^j_n) = W_{2^j} f(x^j_n)$.

2. At each scale 2^j, the local maxima of $\left| W_{2^j} h(x) \right|$ is located at the abscissa (x^j_n).

The value of $W_{2^j} f(x)$ at any x_o can be written as an inner product in $L^2(R)$. Thus, condition 1 is equivalent to

$$\left\langle f(u), \psi_{2^j}(x^j_n - u) \right\rangle = \left\langle h(u), \psi_{2^j}(x^j_n - u) \right\rangle \tag{17}$$

Let *U* be the closure in $L^2(R)$ of the space of functions that are linear combinations of functions in the family

$$\left\{ \psi_{2^j}(x^j_n - x) \right\}_{(j,n) \in Z} \tag{18}$$

It is easy to show that the functions *h(x)* that satisfy (17) for all abscissa x^j_n are the functions whose orthogonal projection on *U* is equal to the orthogonal projection of *f(x)* on *U*. Let *O* be the orthogonal complement of *U* in $L^2(R)$, which means that the space *O* is orthogonal to *U* and that

$$O \oplus U = L^2(R) \tag{19}$$

The functions that satisfy (17) for all abscissa $(x^j{}_n)$ can therefore be written

$$h(x) = f(x) + g(x) \text{ with } g(x) \in O \tag{20}$$

Condition 2 is more difficult to analyze as it is not convex. In order to solve the problem numerically, it is approximated to a convex constraint. Instead of imposing that the local maxima of the wavelet transform of $h(x)$ are located at the abscissa $(x^j{}_n)$, the condition that $\left|W_{2^j}h(x)\right|^2$ is as small as possible on an average, is imposed. This creates local modulus maxima at the positions $(x^j{}_n)$. To have as few modulus maxima as possible outside the abscissa $(x^j{}_n)$, the energy in the derivative of $W_{2^j}h(x)$ is also minimized. Hence the problem reduces to the following optimization problem

$$Min \sum_{j=-\infty}^{+\infty}\left(\left\|W_{2^j}h\right\|^2 + 2^{2j}\left\|\frac{dW_{2^j}h}{dx}\right\|^2\right)$$

$$h(x) = f(x) + g(x)$$

Instead of computing the solution itself, its wavelet transform is reconstructed with an algorithm based on alternate projections. Let K be the space of all sequences of functions $(g^{(x)}{}_j)$ such that

$$\left|(g_j(x))_{j \in Z}\right|^2 = \sum_{j=-\infty}^{+\infty}\left(\left\|g_j\right\|^2 + 2^{2j}\left\|\frac{dg_j}{dx}\right\|^2\right) < +\infty \tag{21}$$
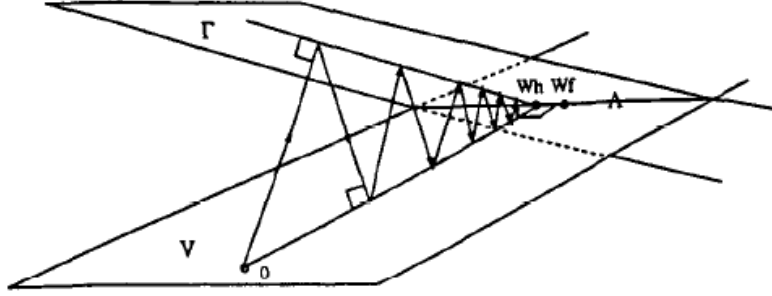

Figure 4: *Approximation of the wavelet transform of f(x) is reconstructed by alternating projections on an affine space Γ and on the space V of all dyadic wavelet transforms. The projections begin from zero element and converge to its orthogonal projection on Γ∩V.*

The norm defines Hilbert structure over **K**. Let **V** be the space of all dyadic wavelet transforms of functions in $L^2(R)$. Let **Γ** be the affine space of sequences of functions $(g^{(x)}{}_j) \in K$ such that for any index j and for all maxima positions $x^j{}_n$

$$g_j(x^j_n) = W_{2^j}f(x^j_n)$$

The dyadic wavelet transforms that satisfy condition 1 are the sequences of functions that belong to $\Lambda = V \cap \Gamma$. The sequence which satisfies both the conditions is the element of Λ whose norm is minimum. This is found by alternating projections on **V** and **Γ**. The projection operator on V is defined by

$$P_V = W o W^{-1} \tag{22}$$

4.3 Projection Operator on Γ

The operator P_Γ transforms any sequence $(g_j(x)) \in K$ into the closest sequence $(h_j(x)) \in \Gamma$ with respect to the norm defined. Let $\epsilon_j(x) = h_j(x) - g_j(x)$. Each function $h_j(x)$ is chosen such that

$$\sum_{j=-\infty}^{+\infty} \left\| \varepsilon_j \right\|^2 + 2^{2j} \left\| \frac{d\varepsilon_j}{dx} \right\|^2 \tag{23}$$

is minimum. To minimize this sum, each component is minimized separately. Let x_0 and x_1 be the abscissa of two consecutive modulus maxima of $W_{2^j} f(x)$. Since $(h_j(x)) \in \Gamma$,

$$\varepsilon_j(x_o) = W_{2^j} f(x_o) - g_j(x_o)$$
$$\varepsilon_j(x_1) = W_{2^j} f(x_1) - g_j(x_1) \tag{24}$$

Between the abscissa x_0 and x_1, the minimization of (23) is equivalent to the minimization of

$$\int_{x_o}^{x_1} \left(\left| \varepsilon_j(x) \right|^2 + 2^{2j} \left| \frac{d\varepsilon_j(x)}{dx} \right|^2 \right) dx \tag{25}$$

The Euler equation associated with this minimization is

$$\varepsilon_j(x) - 2^{2j} \frac{d^2 \varepsilon_j(x)}{dx^2} = 0 \tag{26}$$

for $x \in (x_0, x_1)$. The constraints (24) are the border conditions of this membrane equation. The solution is

$$\varepsilon_j(x) = \alpha e^{2^{-j}x} + \beta e^{-2^{-j}x} \tag{27}$$

where the constants α and β are adjusted to satisfy equations (24).
Any spurious oscillations that may result during the alternate projections can be suppressed by imposing sign constraints

$$\begin{cases} sign(g_j(x) = sign(x_n^j) & \text{if } sign(x_n^j) = sign(x_{n+1}^j) \\ sign(\dfrac{dg_j(x)}{dx}) = sign(x_{n+1}^j - x_n^j) & \text{if } sign(x_n^j) \neq sign(x_{n+1}^j) \end{cases}$$

The algorithm can be easily extended to the 2D case, as a simple product of 2 independent 1D operations along x and y directions, by considering a seperable 2D wavelet filter. The reconstruction algorithm is performed on the Lena Image and the results are shown in Fig. 5.
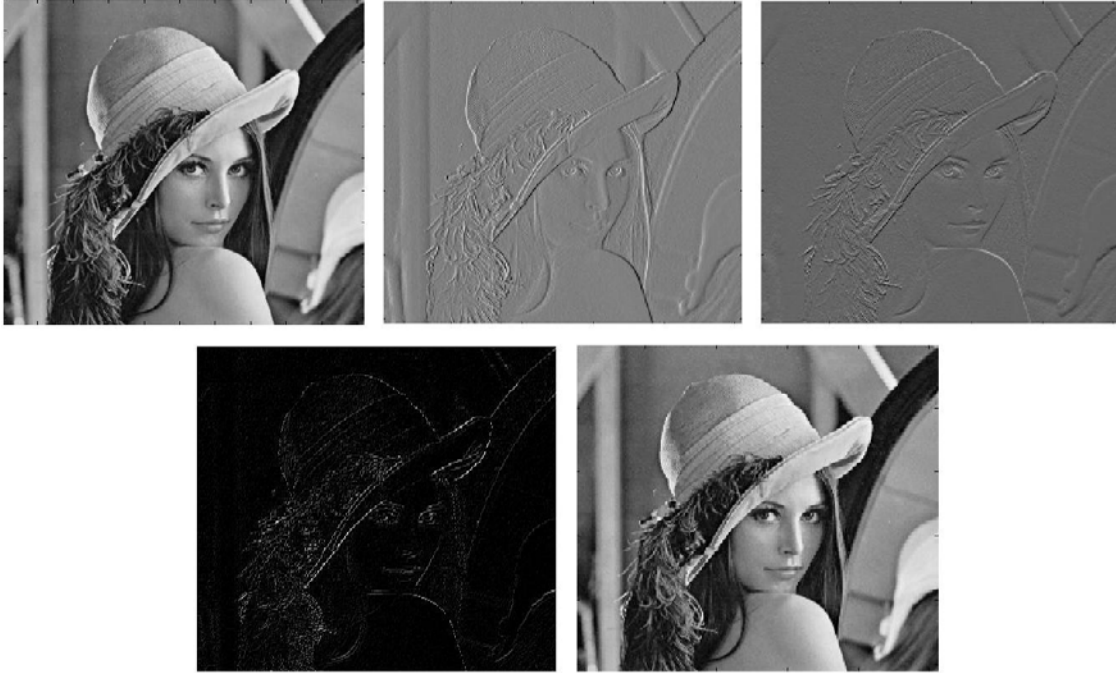

Figure 5: *a) True Image b) Wavelet transform of the image in x direction c) Wavelet transform of the image in y direction d)Edge map e) Reconstructed Image*

5. APPLICATION

An application of the reconstruction algorithm in Image restoration has been demonstrated. It is desired to reconstruct the original image from the noisy one. Two different images have been considered, the image in the first case is corrupted by additive white Gaussian noise where as the one in the second case is corrupted by impulse (salt and pepper) noise.

5.1 Gaussian Noise

It is desired to restore the image corrupted by white Gaussian noise. The restoration is performed from the estimated true edge map via the reconstruction algorithm described in Section 4. The advantage of using the wavelet transform is that images/signals have a sparse representation in this domain. Hence the effect of noise can be removed by setting the nonzero coefficients of the wavelet transform to zero. The threshold value for setting the wavelet coefficients to zero can be estimated using the standard Likelihood ratio test (LRT).

Restoration Algorithm:

1 The wavelet transform of the corrupted image is computed.
2 The variance (power) of the Gaussian noise is estimated from the wavelet transform via MAD estimator.
3 The wavelet transform is then thresholded to remove the effect of noise. The threshold value is set by the standard LRT test.
4 The image is reconstructed from the thresholded wavelet transform via the reconstruction algorithm described in Section 4.

Figure 6: *a) True Image b) Corrupted Image c) Reconstructed Image*

5.2 Impulse (Salt and Pepper) Noise

Here, the image has been corrupted by salt and pepper noise. This noise adds discontinuities in the image resulting in a lot of false edges on computing the wavelet transform. Hence the true edge map cannot be recovered here by thresholding. The effect of the reconstruction algorithm in the presence of impulse noise has been demonstrated in Fig. 7. The algorithm fails to restore the original image in this case.
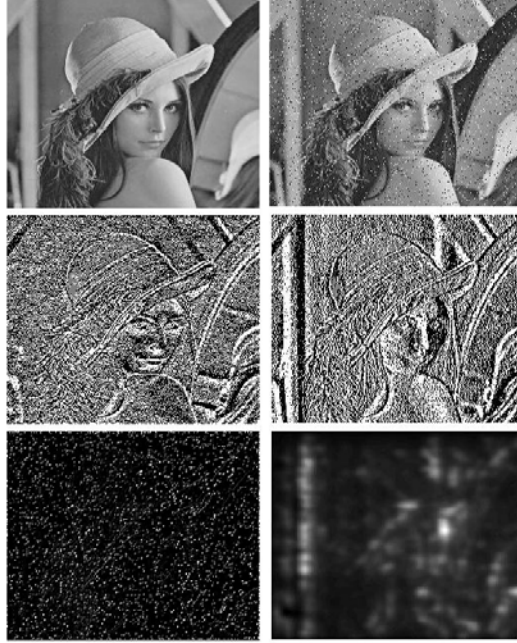
Figure 7: *a) Original Lena Image b) Image corrupted by Salt and Pepper noise c) Wavelet transform of the noisy image in y direction d) Wavelet transform of the noisy image in x direction e) Edge map of Lena after thresholding f) Reconstructed Image*

6. CONCLUSIONS

6.1 Completeness Issues

It has been proved by Meyer that the completeness of the representation conjectured by Mallat, depends on the choice of the smoothing function $\theta(x)$ and it is not valid in general. He also found a non countable family of functions $f_e(x) = f_o(x) + \chi_e(x)$, such that at all scales 2^j, the wavelet transform of $f_e(x)$ and $f_o(x)$ have the same extrema positions and values, where

$$f_o(x) = \begin{cases} \dfrac{1}{2\pi}(1+\cos(x)) & \text{if } |x| < \pi \\ 0 \end{cases}$$

A discrete analysis of the completeness conjecture was done independently by Berman, who found numerical examples to contradict the completeness conjecture.

6.2 Convergence Issues

Besides completeness issues, there are also convergence issues with the algorithm, especially when the solution is not stable, in which case, the alternate projections converge slowly. Fig. 7 shows an example where the reconstruction algorithm does not converge to the original image even after 100 iterations. The reason for the failure of the algorithm could be the incompetence of the edge detection algorithm in detecting certain edges which are visually perceptible to human eye.
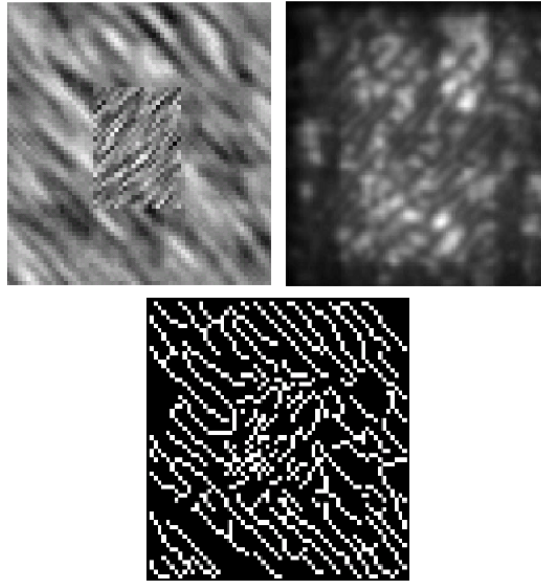
Figure 8: *a) True Image b) Reconstructed Image after 100 iterations c) Edge map via wavelet transform*

7. REFERENCES

[1] Characterization of Signals from Multi scale edges*, Mallat and Zhong, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992.*

[2] The Uniqueness question of discrete wavelet maxima representation, Berman*, Technical Report, Univ of Maryland, 1991.*

[3] Un contre exemple a la conjecture de Marr et a celle de S Mallat, Meyer, *Preprint, 1991.*

[4] Zero crossings of a wavelet transform, Mallat, *IEEE Transactions on Information Theory, 1991.*

[5] A Wavelet Tour of Signal Processing, Mallat

[6] Estimation, Filtering and Detection Theory, Alfred Hero, *Course packet UMich, 2001.*