Wavelets Double Thresholding

Warith HARCHAOUI

August/September 2008

Like many Computer Vision tasks, the use of denoising methods in segmentation can improve the results of Computer Vision tasks [6]. For the sake of simplicity, we suppose that the noise is white and gaussian but our work can be generalized to correlated noise and the gaussian approximation is often used in the literature [1].

1 Noise estimation

The wavelet transform θ of a pure image I is sparse but the wavelet transform b of the noise N is not sparse, that is why the wavelet transform $x = \theta + b$ of the real image $J = I + N$ is almost the wavelet transform of the noise $x \simeq b$ except for the few peak coefficients of the pure signal θ. To estimate the noise, the authors in [4] have the original idea that the pure signal is an outlier from the noise point of view. While keeping this idea in mind, we need to estimate the standard deviation of the noise with a robust estimator [3].

$$\widehat{\sigma} = \frac{median(|X|)}{0.674500}$$

This method is robust to outliers (ie the pure image as mentioned above !) because of the the median estimator. Considering that we want to estimate the fortune of the Microsoft employees from a dataset containing the fortune of Bill Gates, the mean value will not be representative whereas the median will implicitly exclude the outlier.

2 Noise removal

The traditional method consists in thresholding the wavelets coefficients by $T = 3 \cdot \widehat{\sigma}$ to try to remove the noise in the images. In [4] the authors proposes a method to remove correlated noise[1] by estimating the standard deviation and thresholding at each scale j of the wavelet transform :

$$T_j = 3 \cdot \widehat{\sigma}_j$$

Indeed, the thresholding is done as if the noise was white at each scale i.e. each frequency band (approximately because in fact the frequency bands overlap). For the sake of clarity, we will only consider white gaussian noise.

On a database of 732 natural images[2], we tried different values instead of 3. For each image we have :

- an image I considered as pure
- a noisy image $J = I + n$ which is created with a random generator for the noise added n.
- a denoised image $\widehat{I}$ which comes from denoising methods

[1] which is not white
[2] http://perso.telecom-paristech.fr/ rabin/ICPR08/

to evaluate our algorithms we calculate the L2 norm of the linearized vector $r = I - \widehat{I}$.

For each image and several noise levels ($\sigma = 5\%, 10\%, 20\%$) we have selected the best threshold values from an L2 distance point of view by trying many threshold values. It seems that the best threshold is $3 \cdot \sigma$ as it is shown in this Figure 1.

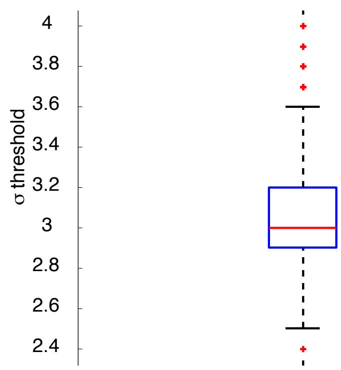

Figure 1: Best thresholds in our 732 images database

3 Double thresholding by 2σ and 4σ

As wavelets have non zero support, when we have a singularity, a wavelet needs some space to express itself as it is shown in Figure 2. By thresholding by $3 \cdot \sigma$ we suppress the low parts of the

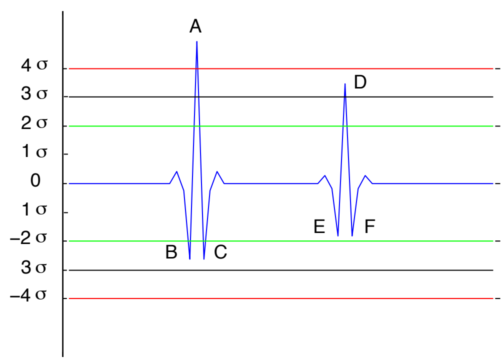

Figure 2: Wavelet transform of two Dirac distributions : one high (left) one low (right)

wavelet's tail and thus artefacts appear in the images : small horizontal and vertical bars as it is shown in Figure 3 and 4.

These artifacts are well-known to be one of the major drawback of wavelets thresholding. One way to avoid these artefacts is to analyse the wavelets coefficients with the same point of view as the authors in [2] for edge detection.

In a few words, the strategy is to select *seeds* i.e. points of which we are sure that they are edge points (with a high gradient above t_1) or in our case signal wavelet values (above $4 \cdot \sigma$ for example) and then *let the seeds grow* in the appropriate *field*. In [2], the field consists in points with a lower gradient value (but higher than $t_2 < t_1$) and in our case the field are the wavelets coefficients that are bigger than $2 \cdot \sigma$.

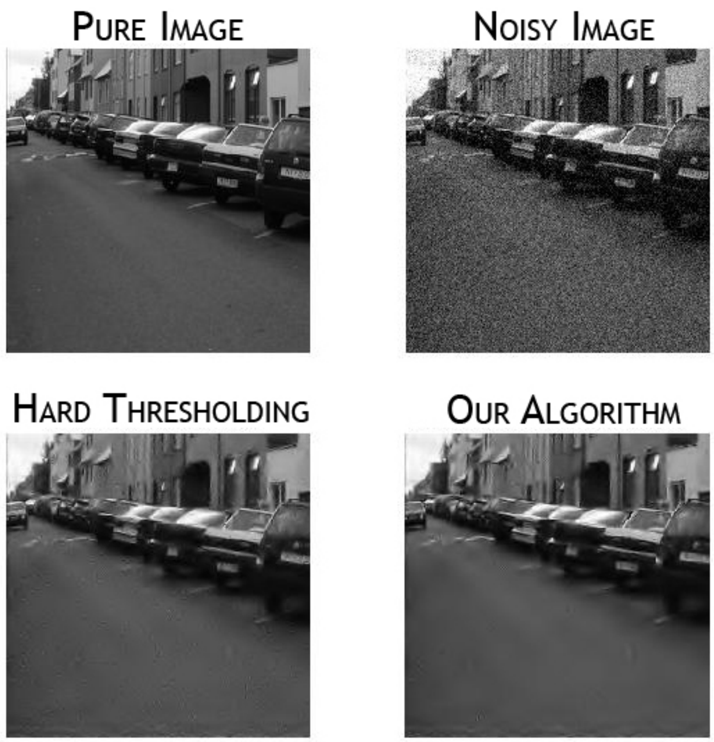

Figure 3: Artefacts in the seed/field thresholding image are fewer than in the simple thresholding image. You should zoom in to see the differences

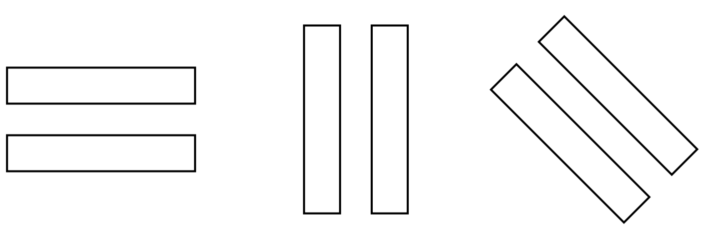

Figure 4: Artefacts that appear by cutting the tails of the wavelets coefficients

To accomplish that goal, we use the dilatation algorithm (that comes from Morphological Mathematics) to let the seeds grow and after that, a multiplication with the boolean matrix of the field is done. The dilatation algorithm is iterated n times where n has half the size of the wavelet support.

In summary, in the simple way of thresholding by $3 \cdot \sigma$ in Figure 2, we only keep A and D. In our thresholding technique, we keep :

- A because it is as seed i.e. higher than $4 \cdot \sigma$
- B and C because they are higher than $2 \cdot \sigma$ and near the seed A.

D is not kept although it is higher than $3 \cdot \sigma$ because it is not a seed and not near a seed, so the algorithm assume that it comes from noise. Moreover, the shape of the big wavelet is more kept in our technique than in a simple thresholding.

We compared our method in 732 images with different other methods including :

- Hard thresholding at $3 \cdot \sigma$
- Soft Wiener [5] thresholding at σ
- Soft Wiener thresholding at $3 \cdot \sigma$
- Our method with $2 \cdot \sigma$ et $4 \cdot \sigma$

with different noise level ($\sigma = 5\%, 10\%, 20\%$). In 94.95 % of the cases, our method have a better L2 norm.

4 Mathematical demonstration

N a big integer and the length of the wavelet transform of J. For the sake of clarity, we will only consider the first scale of the wavelets transform $x = \theta + b$ with θ the pure wavelets transform and b the noise.

$$(\forall i \in [1, N]) \quad x_i = \theta_i + b_i$$

$$b_i \sim N(0, \sigma^2) \quad iid$$

The wavelet support is $s + 1$ with s even. θ have p singularities in c_k ($k \in [1, p]$) therefore the supports of the wavelets are :

$$(\forall k \in [1, p]) \quad S_k = [c_k - \frac{s}{2}, c_k + \frac{s}{2}] \subset [1, N]$$

To be sure that the singularity will be shown at least its peak we suppose that :

$$(\forall k \in [1, p]) \quad |\theta_{c_k} + b_{c_k}| > 4\sigma \tag{1}$$

that means that the singularities are high enough.

$$S = \cup_{k=1}^{p} S_k \subset [1, N]$$

$$(\forall j \notin S) \quad \theta_j = 0 \tag{2}$$

because θ is the wavelets transform of a signal that is piecewise polynomial. We now define a threshold function :

$$(\forall i \in [1, N]) \quad A(x_i) = 1 \text{ or } 0$$

$$\text{Rem: } (A(x_i))^2 = A(x_i)$$

$$(\forall i \in [1, N]) \quad T(x_i) = x_i \cdot A(x_i)$$

So the expectation of the error is e:

$$e = E_b \left\{ \sum_{i=1}^{N} (T(x_i) - \theta_i)^2 \right\} = \sum_{i=1}^{N} E_b \left\{ (T(x_i) - \theta_i)^2 \right\} = \sum_{i=1}^{N} E_b \left\{ (T(\theta_i + b_i) - \theta_i)^2 \right\}$$

Because of (2), we have two cases whether we are in or out the wavelet support:

$$(\forall j \notin S) \quad (T(\theta_j + b_j) - \theta_j)^2 = (T(b_j))^2 = b_j^2 \cdot A(b_j)$$

$$(\forall j \in S) \quad (T(\theta_j + b_j) - \theta_j)^2 = \theta_j^2 + (b_j^2 - \theta_j^2) \cdot A(\theta_j + b_j)$$

$$e = \sum_{j \notin S} E_b \left\{ (T(b_j))^2 \right\} + \sum_{j \in S} E_b \left\{ (T(\theta_j + b_j) - \theta_j)^2 \right\}$$

$$e = \sum_{j \notin S} \int \left\{ b_j^2 \cdot A(b_j) \right\} \cdot p(b_j) \cdot db_j + \sum_{j \in S} \int \left\{ \theta_j^2 + (b_j^2 - \theta_j^2) \cdot A(\theta_j + b_j) \right\} \cdot p(b_j) \cdot db_j$$

In our case we are studying two kinds of thresholding : Hard and Double

$$A_{Hard}(x_i) = 1_{|x_i| > 3\sigma}(x_i)$$

$$A_{Double}(x_i) = 1_{|x_i| > 4\sigma}(x_i) + 1_{|x_i| \in [2\sigma, 4\sigma]}(x_i) \cdot max_{k \in [i - \frac{s}{2}, i + \frac{s}{2}] \setminus \{i\}} \left\{ 1_{|x_k| > 4\sigma}(x_k) \right\}$$

Our goal is to show that $e_{Hard} - e_{Double} > 0$

$$e_{Hard} - e_{Double} = K_{j \notin S} + K_{j \in S}$$

$$K_{j \notin S} = \sum_{j \notin S} \int \left\{ b_j^2 \cdot (A_{Hard}(b_j) - A_{Double}(b_j)) \right\} p(b_{j-\frac{s}{2}}) db_{j-\frac{s}{2}} \ \ldots \ p(b_{j+\frac{s}{2}}) db_{j+\frac{s}{2}}$$

$$K_{j \in S} = \sum_{j \in S} \int \left\{ \left(b_j^2 - \theta_j^2\right) \cdot (A_{Hard}(\theta_j + b_j) - A_{Double}(\theta_j + b_j)) \right\} p(b_{j-\frac{s}{2}}) db_{j-\frac{s}{2}} \ \ldots \ p(b_{j+\frac{s}{2}}) db_{j+\frac{s}{2}}$$

We evaluated $K_{j \notin S}$ numerically with a random generator and it is positive.

$$(\forall j \in S) \quad A_{Double}(x_j) = 1_{|x_j| \geq 2\sigma}(x_j)$$

$$(\forall j \in S) \quad A_{Hard}(x_j) - A_{Double}(x_j) = 1_{|x_j| > 3\sigma}(x_j) - 1_{|x_j| \geq 2\sigma}(x_j) = -1_{|x_j| \in [2\sigma, 3\sigma]}(x_j)$$

because of (1)

$$K_{j \in S} = \sum_{j \in S} \int \left\{ \left(\theta_j^2 - b_j^2\right) \cdot 1_{|\theta_j + b_j| \in [2\sigma, 3\sigma]}(\theta_j + b_j) \right\} p(b_j) db_j$$

Here we need to make a hypothesis on the θ_j distribution in the support S to compute this sum numerically by a random generator. We suppose that θ_j is a uniform random variable in $[-A, A]$, we get :

$$K_{j \in S} \propto E_{\theta, b} \left(\left(\theta^2 - b^2\right) \cdot 1_{|\theta + b| \in [2\sigma, 3\sigma]}(\theta + b) \right)$$

because of (2), $A \simeq 4\sigma$. We have, in Figure 5, the sign of $\propto E_b \left(\left(\theta^2 - b^2\right) \cdot 1_{|\theta + b| \in [2\sigma, 3\sigma]}(\theta + b) \right)$ and Figure 6 is the integral value $\propto K_{j \in S}$ which is positive if $A \gtrsim 1.8\sigma$.

Finally, we have shown that our algorithm has a better L2 norm than the classic hard thresholding.

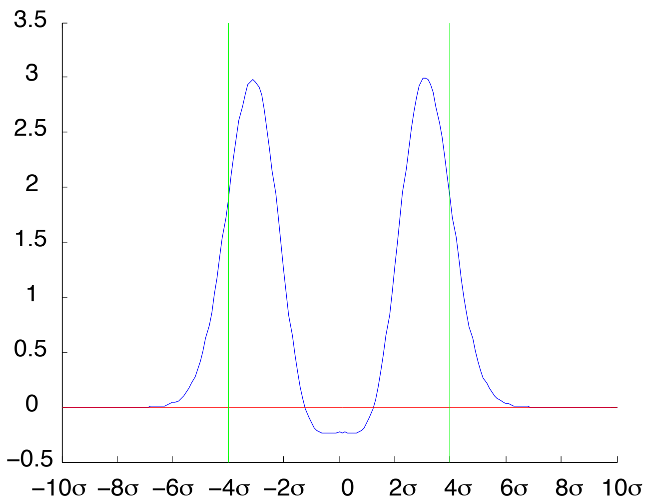

Figure 5: X: A ; Y: $\propto E_b\left(\left(\theta^2 - b^2\right) \cdot 1_{|\theta+b| \in [2\sigma,3\sigma]}(\theta + b)\right)$

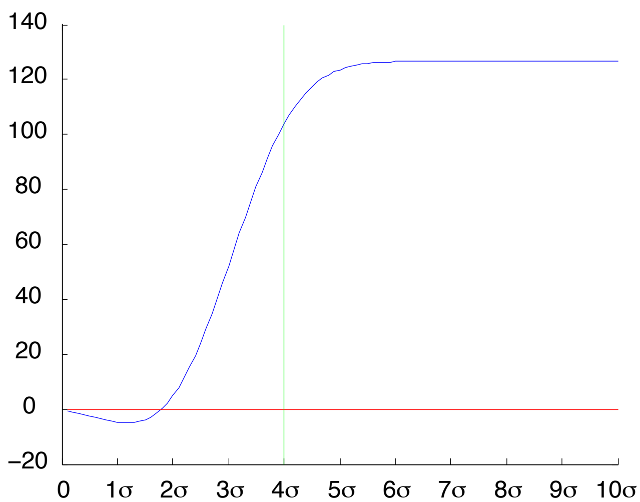

Figure 6: X: A ; Y: The sign of $K_{j\in S}$

5 Conclusion

In this report, we have shown a new thresholding algorithm which can be called hysteresis thresholding. A way to improve this algorithm is to use wavelets with different orientations instead of only horizontal, vertical and diagonal. We have used a symetric wavelet for that reason : try to be rotation-invariant. Our algorithm is translation-invariant because of the undecimated Wavelets transform.

One can point out that the L2 norm is not the criterion of quality for the human eye. The L2 norm has the advantage to be more easily computed but a great deal of work has to be done by using a more *biological* measure of quality.

References

[1] A. Buades, B. Coll, and J.M. Morel, *A Review of Image Denoising Algorithms, with a New One*, MULTISCALE MODELING AND SIMULATION **4** (2005), no. 2, 490.

[2] J. Canny, *A computational approach to edge detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence **8** (1986), no. 6, 679–698.

[3] P.J. Huber, *Robust statistics*, Wiley-Interscience Hoboken, NJ, 2004.

[4] I.M. Johnstone and B.W. Silverman, *Wavelet Threshold Estimators for Data with Correlated Noise*, Journal of the Royal Statistical Society: Series B (Methodological) **59** (1997), no. 2, 319–351.

[5] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.

[6] N. Paragios, Y. Chen, and O. Faugeras, *Handbook of Mathematical Models in Computer Vision*, Springer, 2006.