ADAPTIVE WAVELET IMAGE DECOMPOSITION USING LAD CRITERION

Ana Sović and Damir Seršić

Faculty of Electrical Engineering and Computing, University of Zagreb
Unska 3, 10000 Zagreb, Croatia
phone: + (385) 1-6129-973, fax: + (385) 1-6129-652, email: ana.sovic@fer.hr, damir.sersic@fer.hr
web: http://maja.zesoi.fer.hr/~sdamir/eusipco

ABSTRACT

In this paper, an adaptive separable 2D wavelet transform is proposed. Wavelet transforms are widely used in signal and image processing due to its energy compaction property. Sparser representation corresponds to better performance in compression, denoising, compressive sensing, sparse component analysis and many other applications. The proposed scheme results in more compact representation then fixed wavelet. Instead of the commonly used least squares criterion, least absolute deviation (LAD) is introduced. It results in more accurate adaptation resistant to outliers. The advantages of the proposed method have been shown on synthetic and real-world images.

1. INTRODUCTION

Two-dimensional wavelet transform is widely used in image analysis, compression, denoising and many other applications [1][2][3]. The analytical strength of wavelets lies in its energy compaction property, or, in other words, in sparseness of image representation in the wavelet domain. The wavelet transforms can be efficiently realized using filter banks. Among the others, the lifting scheme is one of the most effective realizations of wavelet filter banks [4].

The key of sparseness is the polynomial annihilation property: well chosen wavelets provide for desired number of vanishing moments, which cancel out the polynomials of the corresponding order. In fact, many real-world images can be locally well approximated by polynomials, which results in sparse representation. It is essential for many hot research topics, such as compressive sensing, blind separation of underdetermined mixtures and many others [5][6].

Still, there are several drawbacks. More vanishing moments correspond to longer support of the wavelets, thus producing ringing of wavelet coefficients on image edges. It causes visible artifacts, which is especially annoying in image analysis. Moreover, periodic patterns in images (like sine waves) are not well represented by finite-order polynomials. Hence, it results in a sub-optimal wavelet representation: sine waves in input images result in sine waves in the wavelet domain.

It motivated us to realize an pixel-wise adaptive wavelet filter bank that overcomes the aforementioned drawbacks. The idea is to have fixed part(s) of the bank which provide for desired number of vanishing moments, and variable part(s) which provide for pixel-wise adaptation. The vanishing moments must be preserved regardless of the adaptation, and the adaptation must behave well on edges in images. It should result in more compact representation for a wide class of inputs, including periodicities and image components with long-tail Taylor series.

In this paper, we propose a separable implementation of an adaptive wavelet filter bank with desired number of vanishing moments and adjustable number of adaptive parameters. Due to its robustness (insensitivity to outliers), we propose use of the least absolute deviation (LAD) criterion [7] on the fixed size sliding window for the adaptation. We have shown advantages of the proposed scheme over the fixed wavelets, as well as over the least squares error (LS) adapted wavelets; when applied on synthetic and real-world images.

2. SEPARABLE 2D ADAPTIVE WAVELET FILTER BANK

The lifting scheme is a numerically efficient and a memory saving implementation of wavelet filter banks [4]. Furthermore, it enables pixel-wise adaptation that ensures perfect reconstruction by a simple change of sign, thus enabling a construction of the second-generation wavelets [8]. Although there are several ways of implementing wavelets in 2D, and some of them are non-separable, or even adaptive [3], in this work we use a separable implementation. It has shown several advantages, due to its: a) simplicity and efficiency, b) independent adaptation across the rows and columns.

At first, every row of the image is analyzed, as shown in Figure 1. Primal lifting step S and dual lifting step T consist of fixed and variable part that ensure zero moments and enable adaptation, respectively. The same analysis is conducted across every column of the intermediate results A and D.

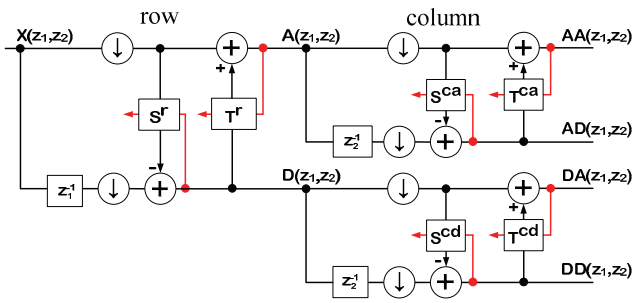

Figure 1 – Separable 2D adaptive lifting scheme.

Due to decimation, four resulting images *AA*, *AD*, *DA* and *DD* are quarter size of input image $X(z_1,z_2)$.
Essentially, the separable implementation consists of three one-dimensional adaptive filter banks applied in different directions, one of which is shown in Figure 2. It is realized using the adaptive lifting scheme (detailed description in [9]). Wavelet details *D* are outputs of high-pass filter defined by *S* in the primal lifting step, and approximations are outputs of low-pass filter defined by *T* in the dual lifting step. The primal lifting step shown in Figure 2 has fixed part that ensures at least two vanishing moments ("wired" $b_0 = 1$). We can provide more vanishing moments by successive setting parameters b_k to 1, and we can use the remaining subset of parameters b_k for adaptation. In that case, depending on the adaptation criterion, the analysis function could better match the local properties of input signal *X*. The analogy stands for the dual lifting step. Finally, the reconstruction is conducted by reversing the order and corresponding signs of the lifting steps.
At each point, the adaptation algorithm uses multiple inputs U_k to successfully predict single output Y_d and to minimize the error *D* after primal lifting step, as well as multiple inputs V_k to predict output Y_a and minimize *A* after dual lifting step. Wide class of input signals produce zero wavelet coefficients: polynomials due to the fixed part, sine waves due to the adaptive part (see paragraph 4), all in purpose to produce the sparsest possible representation.

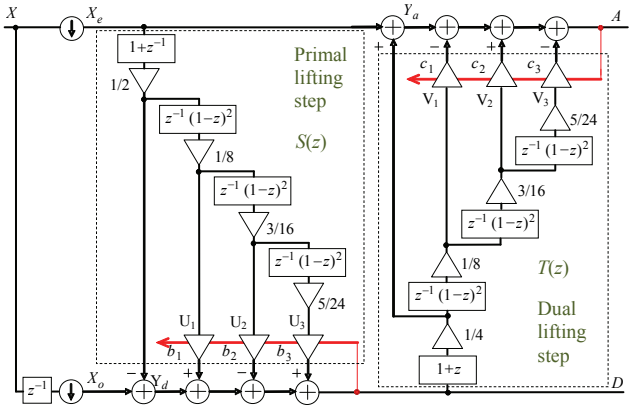

Figure 2 – Adaptive wavelet filter bank. The argument *z* is a placeholder for z_1 or z_2, depending on direction of application.

3. LAD VERSUS LSE MINIMIZATION

3.1 MISO system

In general, a system with multiple inputs and a single output (MISO) is given by difference equation:

$$\begin{aligned} y(n) + a_1 y(n-1) + \cdots + a_{n_a} y(n-n_a) \\ = b_{11}u_1(n) + \cdots + b_{1n_{b1}}u_1(n-n_{b1}+1) \\ + b_{21}u_2(n) + \cdots + b_{2n_{b2}}u_2(n-n_{b2}+1) + \cdots \\ + b_{n_u1}u_{n_u}(n) + \cdots + b_{n_un_{bn_u}}u_{n_u}(n-n_{bn_u}+1), \end{aligned}$$

where n_u is the number of inputs $u_1, u_2, \ldots u_{n_u}$; n_a, n_{b1}, n_{b2},… are the system orders and $a_1,\ldots,b_{11},\ldots$ are the system parameters. To calculate several shifts of $y(n)$, we use a matrix equation $\mathrm{Y} = \Psi \cdot \theta$, where

$$\mathrm{Y} = \begin{bmatrix} y(n) \\ y(n-1) \\ \ldots \\ y(n-n_w+1) \end{bmatrix}, \quad \theta = \begin{bmatrix} -a_1 \\ \ldots \\ -a_{n_a} \\ b_{11} \\ \ldots \\ b_{n_un_{bn_u}} \end{bmatrix},$$

$$\Psi = \begin{bmatrix} y(n-1) & y(n-2) & \ldots & y(n-n_w) \\ \vdots & \vdots & \ldots & \vdots \\ y(n-n_a) & y(n-n_a-1) & \ldots & y(n-n_a-n_w+1) \\ u_1(n) & u_1(n-1) & \ldots & u_1(n-n_w+1) \\ \vdots & \vdots & \ldots & \vdots \\ u_{n_u}(n-n_{bn_u}+1) & u_{n_u}(n-n_{bn_u}) & \ldots & u_{n_u}(n-n_{bn_u}-n_w+2) \end{bmatrix}^T.$$

and n_w is the window length of Y.
Assume that we know the inputs and the outputs, and we do not know parameters θ. The outputs are known up to some measurement uncertainty, e.g. they contain noise. So the parameters can be estimated using some error minimization technique.

3.2 Least squares (LS) and least absolute deviation (LAD)

If estimated parameters are $\hat{\theta}$, estimation error is:

$$\epsilon = \mathrm{Y} - \Psi\hat{\theta}.$$

Dimension of ϵ is n_w. L_2 norm based cost function is:

$$F(\theta) = ||\epsilon||_2^2 = \epsilon^T\epsilon,$$
$$F(\theta) = (\mathrm{Y} - \Psi\hat{\theta})^{\mathrm{T}}(\mathrm{Y} - \Psi\hat{\theta}) = \sum_{i=1}^{n_w}(y_i - \Psi_i\hat{\theta})^2,$$

where y_i is *i*-th element of vector Y, and Ψ_i is *i*-th row of matrix Ψ. The well-known LS solution for minimum of *F* is:

$$\hat{\theta} = (\Psi^{\mathrm{T}}\Psi)^{-1}\Psi\mathrm{Y}. \quad (1)$$

L_1 norm based cost function is:

$$F(\theta) = ||\epsilon||_1 = |\epsilon| = \sum_{i=1}^{n_w}|y_i - \Psi_i\hat{\theta}|. \quad (2)$$

One possible way for finding minimum L_1 parameters is iteratively reweighted least squares (IRWLS). Using this method we can express the L_1 cost function similarly as the L_2:

$$F(\theta) = \sum_{i=1}^{n_w}|y_i - \Psi_i\hat{\theta}|\,W_{ii}|y_i - \Psi_i\hat{\theta}|,$$

where W is a diagonal weight matrix with diagonal elements $W_{ii} = 1/|\mathrm{y_i} - \Psi_i\hat{\theta}|$. Since we do not know $\hat{\theta}$ in advance, we do not know W_{ii} as well. But, we can find them iteratively [10]:

1. take some initial $\hat{\theta}_0$, $k = 1$;
2. calculate elements of matrix W;
3. find estimated parameters $\hat{\theta}_k = (\Psi^{\mathrm{T}}\mathrm{W}\Psi)^{-1}\Psi^{\mathrm{T}}\mathrm{W}\mathrm{Y}$;
4. if $\hat{\theta}_k \neq \hat{\theta}_{k-1}$ take $k = k + 1$ and repeat steps 2 - 4; else stop with $\hat{\theta} = \hat{\theta}_k$ (up to desired precision).

The other way of finding LAD adapted parameters is based on linear programming (LP). Cost function can be stated as:

$$\min \sum_i t_i,$$

$$\text{subject to } -t_i \le y_i - \Psi_i \hat{\theta} \le t_i,$$

where t_i are non-negative slack variables [11], and the minimization problem can be solved using any known LP technique.

We use either one or another way to solve our L_1 adaptation problem. Efficient L_1 minimization is a hot research topic [12], which is beyond the scope of this paper.

4. ADJUSTABLE LIFTING STEPS

At first, we observe each row of input image $X(z_1, z_2)$ and analyze it using the adaptive wavelet filter bank (Figure 2). The primal lifting step predicts odd samples from neighboring evens. Fixed part ensures two vanishing moments: it annihilates linear component of the observed row producing intermediate signal Y_d. Now, we use input signals u_1, u_2 and u_3 and adjust parameters b_1^r, b_2^r and b_3^r in order to predict Y_d. The prediction error remains in intermediate details D. We observe these signals on a sliding window of the length n_w. Hence, our optimization matrices are:

$$\mathrm{Y} = \begin{bmatrix} y_d(n) \\ y_d(n-1) \\ \dots \\ y_d(n-n_w+1) \end{bmatrix}, \qquad \theta = \begin{bmatrix} b_1^r \\ b_2^r \\ b_3^r \end{bmatrix},$$

$$\Psi = \begin{bmatrix} u_1(n) & u_2(n) & u_3(n) \\ u_1(n-1) & u_2(n-1) & u_3(n-1) \\ \dots & \dots & \dots \\ u_1(n-n_w+1) & u_2(n-n_w+1) & u_3(n-n_w+1) \end{bmatrix}$$

Desired parameters θ can be found by minimizing L_2 norm applying (1) or by minimizing L_1 norm (2) using linear programming techniques or iteratively reweighted least squares method.

Similar procedure is conducted in the dual lifting step. Here, we predict Y_a from inputs v_1, v_2 and v_3 by adjusting parameters c_1^r, c_2^r and c_3^r. Prediction error is intermediate approximation A.

Finally, following the separable 2D generalization from Figure 1, we apply the same procedure to every column of the intermediate results D and A. Using the chosen criteria, we calculate parameters b_1^{ca}, b_2^{ca}, b_3^{ca}, c_1^{ca}, c_2^{ca} and c_3^{ca} for approximation A and parameters b_1^{cd}, b_2^{cd}, b_3^{cd}, c_1^{cd}, c_2^{cd} and c_3^{cd} for details D. The final results are four quarter-size images designated as *AA, AD, DA* and *DD*. The calculated parameters should be used on the reconstruction side, as well.

5. RESULTS

To illustrate the proposed method, we adapt only one parameter (b_1) across the rows and columns. For the simplicity, all other parameters in primal and dual steps are set to zero.

By adjusting parameter b_1, we actually change the high-pass filter, and its zero locus plots are shown in Figure 3. Fixed part of the primal lifting step ensures two zeros in $z = 1$.

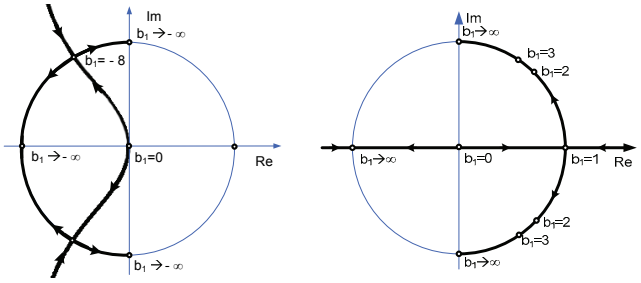

Figure 3 – Zero locus plot of the HP filter, parameter b_1. Left: negative values, right: positive values of b_1.

Considering that, we created an artificial image (left hand side of Figure 4) composed from four cosine wave segments with different frequencies and angles. Upper two segments have vertical cosine lines. Left lower segment contains sum of horizontal and vertical cosine lines, while the remaining segment contains tilted cosine wave. In the centre of the first segment there are two white pixels – outliers. Resolution of the image is 200x200 pixels. Filter parameters $b_1^{\mathrm{r}}(n)$ are being adapted for each position n in every row, on a sliding window. The adaptation is conducted on interval $[n-15, n+15]$ by the least squares algorithm and by the least absolute algorithm to produce intermediate results A and D. The same procedure is then used column-wise to find filter parameters $b_1^{\mathrm{ca}}(n)$ and $b_1^{\mathrm{cd}}(n)$.

Parameters $b_1^{[\cdot]}(n)$ adjust to the frequencies of the analyzed signals in purpose to cancel out the cosine waves. Both adaptive approaches, LS and LAD, result in significantly sparser representation, when compared to the fixed wavelets (biorthogonal, 2 vanishing moments). One-level decomposition is shown in Figure 4.

Adapted parameters $b_1^{[\cdot]}(n)$ are shown in Figure 5. The edge between the two patterns of the image is significantly narrower for the LAD criterion. Transition area for the LAD (minimum L_1) might be only one pixel, which is not the case for the LS (minimum L_2), where it depends on the adaptation window length. The influence of the outlier is clearly visible in the LS adapted parameters, while it is not present at all when using the LAD.

To additionally illustrate properties of the proposed adaptive wavelets, we extracted one row of image x and corresponding rows of the DA coefficients and the b_1^r parameters, as shown in Figure 6. The row contains an outlier, which is pointed by an arrow. Clearly, the DA coefficients produced by the adaptive wavelets are more concentrated then by the fixed wavelets. The sparsest coefficients are result of the LAD criterion, with shorter transition areas and less ringing near the edges. The difference between the two criteria is clearly visible in adaptive parameters b_1^r. The adaptation window length determines the transition area on the edges and around the outliers when the LS is applied. Moreover, the transition area is very short when the LAD is used and insensitive on the window length.

In real images, success of the adaptation depends on noise, spatial frequency content and spatial stationarity.

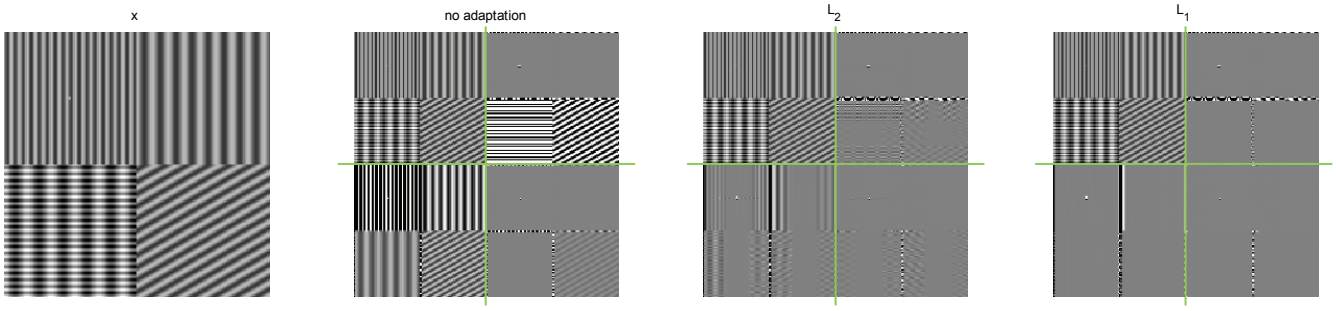

Figure 4 – Left: input image (x). Right: decimated one level wavelet decomposition using fixed wavelets (no adaptation), LS adaptive (L_2) and LAD adaptive wavelets (L_1). In each wavelet decomposition, upper left quarter is approximation AA; upper right quarter is detail AD, lower left quarter is detail DA, and lower right quarter is detail DD. Gray-levels of details are rescaled for visibility. Both adaptive approaches significantly reduce details. L_1 adaptation is almost optimal: very sparse representation.

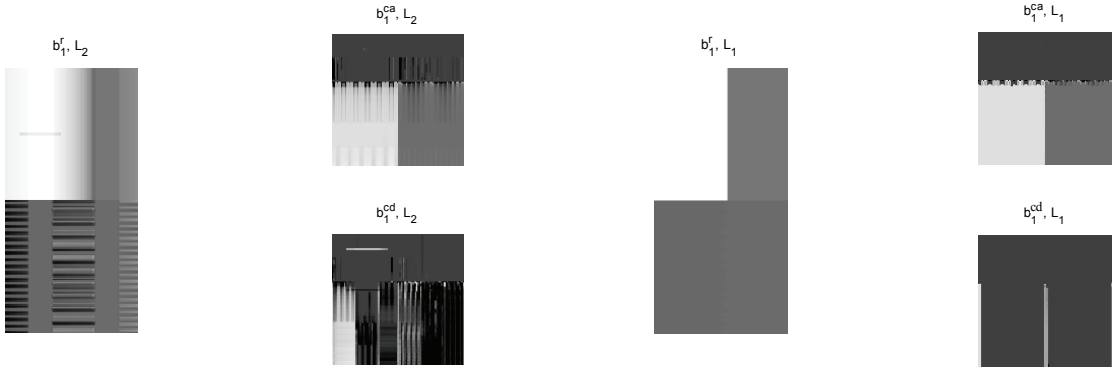

Figure 5 – Adapted parameters. Left to right: b_1^{r}, b_1^{ca} (up) and b_1^{cd} (down) adapted using least squares (minimum L_2); b_1^{r}, b_1^{ca} (up) and b_1^{cd} (down) adapted using least absolute deviation (minimum L_1). Proposed sliding-window L_1 adaptation is almost optimal: it adjusts to the frequency content of the input image, with sharp transitions between different areas.

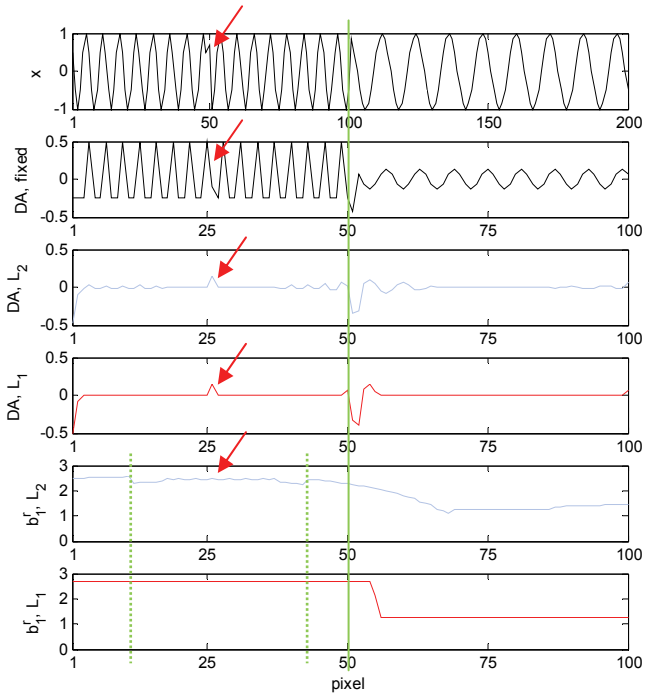

Figure 6 – One row of image x (50th). Top to bottom: original signal, details (fixed wavelets), details (L_2 adapted), details (L_1 adapted), parameters (L_2), parameters (L_1). Transition area and outlier influence is determined by the sliding window length for the L_2 norm and almost invisible for the L_1.

Figure 7 (left hand side) shows a magnified part of the real image (Barbara, part of clothes) contaminated by the salt and peppers noise (1% pixels, Gaussian amplitude distribution, $\sigma = 100$). The adaptation is conducted in the same way as for the previous synthetic image. The results of one-level wavelet decomposition for fixed and adaptive wavelets are presented in Figure 7.

In the case of fixed wavelets, periodic parts of the image result in periodic wavelet coefficients. The influence of outliers significantly decreases the performance of the LS adaptation (minimum of L_2): the periodicities survive in coefficients, as well as the outliers. On the contrary, the LAD adaptation works fine: the adaptation is not affected by the outliers. Hence, the periodicities are mostly canceled out.

In Figure 8 one row of the real image is shown, as well as the intermediate wavelet coefficients D and parameters b_1^{r} for both adaptive criteria. The consequences of the outlier (marked by arrows) on the LS adaptation are evident.

Computational times for described image decomposition methods, simulated in MATLAB using an average PC are presented in TABLE I.

In this work, the adaptation algorithms were not optimized for efficiency, which is left for the future research. Clearly, the adaptation results in almost optimally sparse representation, but it is paid with significant computational burden. The LAD adaptation is numerically demanding, due to lack of the minimum L_1 closed form solution.

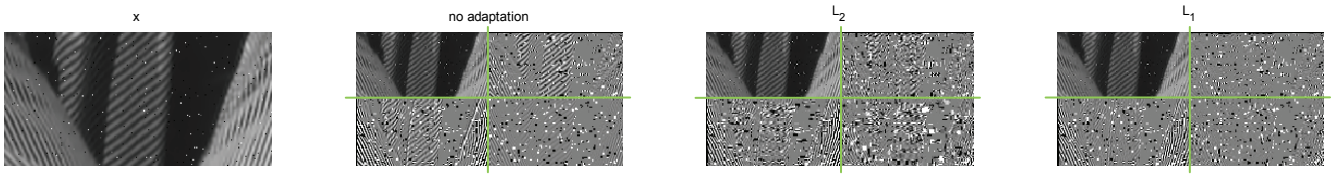

Figure 7 – A magnified part of Barbara. Left: input image (x). Right: decimated one level wavelet decomposition using fixed wavelets (no adaptation), LS adaptive (L_2) and LAD adaptive wavelets (L_1). In each wavelet decomposition, upper left quarter is approximation AA; upper right quarter is detail AD, lower left quarter is detail DA, and lower right quarter is detail DD. Gray levels of details are rescaled for visibility. L_2 adaptation is sensitive to outliers, L_1 adaptation cancels out most of periodicities.

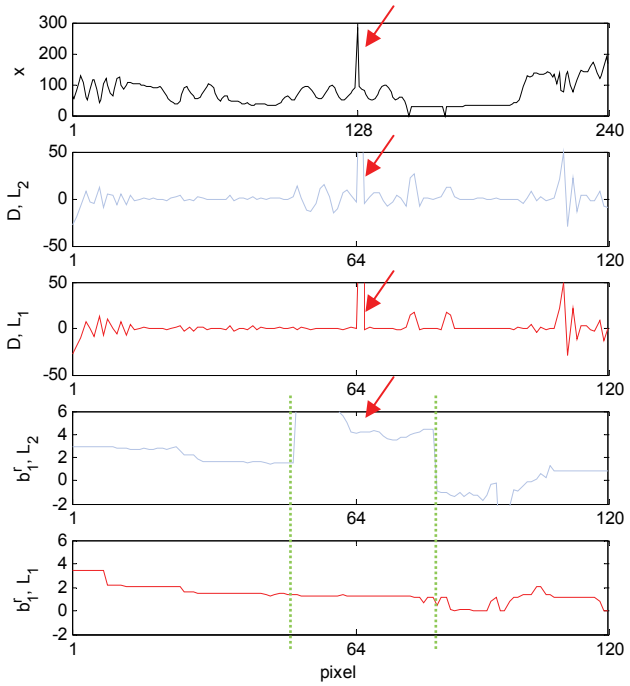

Figure 8 – One row of the real image (magnified part of Barbara). Top to bottom: input signal x, details (L_2 adapted), details (L_1 adapted), parameters (L_2), parameters (L_1). The LAD adaptation is robust to outliers (marked by arrows), which is not the case for the LS adaptation.

TABLE I. AVERAGE RUN – TIMES IN SECONDS FOR 50 DIFFERENT SYNTHETIC IMAGES (DIFFERENT COSINE WAVE PATCHES). TESTS ARE RUN ON INTEL CORE2 DUO CPU 2.67 GHZ AND 2.00 GB OF RAM.

	DWT	L_2	L_1 **LP**	L_1 **IRWLS**
mean	0.1507	6.884	477.0	171.7
median	0.1503	6.880	448.5	169.4

6. CONCLUSION

In this paper, a novel adaptive wavelet image decomposition using the least absolute deviation criterion is presented. Results on synthetic and real images show that proposed method outperforms fixed wavelets in the sense of information compaction and sparseness. The sparseness is very important for many applications: compression, denoising, compressive sensing, sparse component analysis and many others. We have shown the advantages of the L_1 criterion, namely robustness to outliers and shorter transitions between the areas of different statistical properties. Good analytical properties of the proposed LAD adaptive scheme are paid by increased numerical complexity, which become less and less important along with advance of computers and algorithms.

REFERENCES

[1] M. Tomić, D. Seršić, M. Vrankić, "Edge-preserving adaptive wavelet denoising using ICI rule," *IET Electronics Letters*, 44, 11; pp. 698-699, 2008.

[2] M. Vrankić, D. Seršić, "Image denoising based on adaptive quincunx wavelets," *In Proc. of the IEEE 6th Workshop on MSP*, Omnipress, Italy, pp. 251-254, 2004.

[3] M. Vrankić, D. Seršić, V. Sučić, "Adaptive 2-D wavelet transform based on the lifting scheme with preserved vanishing moments," *IEEE Transactions on Image Processing*, 19, 8; pp. 1987-2004, 2010.

[4] W. Sweldens and I. Daubechies, "Factoring wavelet transforms into lifting steps," *J. Fourier Analysis and Applications,* vol. 29, pp. 511-546, 1997.

[5] R. G. Baraniuk. "Compressive sensing," *IEEE Signal Processing Mag*., vol. 24, no. 4, pp. 118–120, 124. 2007.

[6] Y. Li, S. Amari, A. Cichocki, and C. Guan, "Underdetermined blind source separation based on sparse representation," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 3139–3152, Feb. 2006.

[7] R.J. Bošković, „De literaria expeditione per pontificiam ditionem et synopsis amplioris operis," *Bononiensi Scientiarum et Artum Instituto atque Academia Commentarii*, 1757.

[8] W. Sweldens, "The lifting scheme: A construction of second generation wavelets". *SIAM J. Math. Anal*., vol 29, no.2, pp. 511-546, 1997.

[9] D. Seršić, "A realization of wavelet filter bank with adaptive filter parameters," *In Proc. EUSPICO 2000. X European Signal Processing Conference Tampere*, vol. 3, pp. 1733-1736, 2000.

[10] K. Sabo, R. Scitovski, "The best least absolute deviations line-properties and two efficient methods for its derivation," *ANZIAM Journal*. 50, 2; 185-198, 2008.

[11] C. J. Adcock and N. Meade. "A Comparison of two LP solvers and a new IRLS algorithm for L_1 estimation," *L_1-Statist. Procedures and Related Topics*, vol. 31, pp. 119-132, 1997.

[12] J.A. Cadzow. "Minimum L_1, L_2 and L_∞ norm approximate solutions to an overdetermined system of linear equations," *Digital Signal Processing*, vol. 12, pp. 524-560, 2002.