

Adaptive Bayesian Shrinkage Model for Spherical Wavelet Based Denoising and Compression of Hippocampus Shapes

Xavier Le Faucheur¹, Brani Vidakovic², Delphine Nain¹, and Allen Tannenbaum¹

¹ School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA, (xavier@gatech.edu),

² School of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Abstract. This paper presents a novel wavelet-based denoising and compression statistical model for 3D hippocampus shapes. Shapes are encoded using spherical wavelets and the objective is to remove noisy coefficients while keeping significant shape information. To do so, we develop a non-linear wavelet shrinkage model based on a data-driven Bayesian framework. We threshold wavelet coefficients by locally taking into account shape curvature and interscale dependencies between neighboring wavelet coefficients. Our validation shows how this new wavelet shrinkage framework outperforms classical compression and denoising methods for shape representation. We apply our method to the denoising of the left hippocampus from MRI brain data.

1 Introduction

Shape analysis and modeling of the hippocampus structure has become a topic of interest in medical imaging since local, and sometimes complex, variations of the hippocampus could carry relevant information about its role in neural disorders, like Schizophrenia or other diseases. In [1, 2], the authors proposed to encode shape signal using spherical wavelets, which were originally introduced in the computer graphics area [3]. These “second generation” wavelets allow for a multiscale shape representation by decomposing the data both in scale and space using a multiresolution mesh. This representation also allows for efficient compression by discarding wavelet coefficients with low values that correspond to irrelevant shape information and high frequency coefficients that represent noisy artifacts. This process is called hard wavelet shrinkage and has been widely researched for traditional types of wavelets, but not much explored for second generation wavelets. During a wavelet shrinkage process, significant variations of the shape have to be preserved, while noisy terms should be removed. In classical wavelet shrinkage, different methods offer a selective denoising model by developing the best uniform threshold in order to remove low-valued wavelet coefficients

[4–6], however results may become unsatisfactory when noise increases, gets irregularly distributed over the signal or when the coefficients are very sparse. To address these issues, Bayesian approaches are capable of incorporating interesting prior information that allows for more robust and adaptive thresholding [7–10].

The objective of this work is to obtain a compressed three-dimensional shape representation using Bayesian shrinkage framework applied to second generation wavelets that combine smoothness and accuracy, where smoothness will be defined later. In [11] the authors present such a Bayesian shrinkage framework for spherical wavelets with interscale dependency and intrascale smoothing considerations and show how local consistency can help outperform uniform shrinkage rules.

In this paper, we build on that model by proposing a more complete, spatially adaptive, data-driven and more robust Bayesian shrinkage framework for spherical wavelets by taking into account shape curvature in addition to interscale information. Curvature estimation is based on the coarse structure of the shape and interscale information is brought by wavelet coefficients of neighboring vertices from coarser levels. When applying our shrinkage rule to a certain coefficient, we consider the location of the vertex at which the coefficient is defined by looking at those two distinct elements, which may have different influences on the shrinkage. For instance, in low curvature regions of the shape, coefficients from fine levels have a higher likelihood to be shrunk, since they would be considered noise, unless all their direct neighbors have significantly high values. Our results exhibit a higher compression rate while preserving a low average reconstruction error. Therefore, smoothness, consistency and accuracy are combined an optimal way.

We now summarize the remainder of our paper. In Section 2, we give an overview of the shape representation using spherical wavelets. Our Bayesian shrinkage model is described in Section 3. In Section 4, we present our results on left hippocampus dataset, and in Section 5 we comment on our results and outline several further research directions.

2 Shape Representation

2.1 Spherical Wavelets

This work is based on the shape encoding described in [1, 2], where the shape signal is approximated using a set of wavelet basis functions defined on the sphere and modulated by coefficients. We used datasets of left hippocampus shapes obtained from MR imaging, which were manually segmented by a radiologist. Each 3D surface was then extracted as a triangulated surface; see Section 4 below.

A regular multiresolution mesh with spherical topology is first built by recursively subdividing an initial polyhedron, where new vertices are introduced at each new subdivision level. A hierarchy between vertices of the different resolution levels is therefore created, and a “parent” vertex from level j has a given

number of “children” from level $j + 1$. At each resolution level, we build spherical wavelet functions consisting of biorthogonal scalar functions with decreasing support as the resolution increases [12]. Then, with a one-to-one mapping from our surfaces to the unit sphere we equip our shapes with that triangulated multiresolution mesh. The coordinates of the original 3D surface at the N vertices of the regular multiresolution mesh are found by interpolation. The shape signal is represented by a $3N$ -long vector containing the (x, y, z) coordinates of each one of the mesh vertices. We then encode that signal by projecting it onto the set of spherical wavelet functions previously built. Thus, at each resolution level, we obtain wavelet coefficients (d_x, d_y, d_z) for the coordinates (x, y, z) of each one of the vertices belonging to that level.

2.2 Shape Model

Both shape signal and wavelet coefficients are represented by $3N$ -long vectors. In the vectors of coefficients the first N entries correspond to the x-coordinates of the shape, ranked from coarse resolution to high resolution. The following N entries correspond to the y-coordinates while the last N entries correspond to the z coordinates. The goal of our method is to remove wavelet coefficients that most likely correspond to irrelevant shape information and that are considered noise. The proposed model of shape representation is then:

$$\mathbf{s} = \mathbf{f} + \boldsymbol{\varepsilon} \quad (1)$$

where \mathbf{s} is the observed shape signal, \mathbf{f} the $3N$ -vector of the “noiseless” shape signal (x, y, z) we want to estimate at the N vertices, and $\boldsymbol{\varepsilon}$ is i.i.d. normal noise.

From the linearity of the wavelet transform [1], we obtain a similar additive representation in the wavelet domain:

$$\mathbf{d} = \boldsymbol{\theta} + \boldsymbol{\eta} \quad (2)$$

where \mathbf{d} , $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are the wavelet coefficient vectors obtained after applying the wavelet transform to \mathbf{s} , \mathbf{f} and $\boldsymbol{\varepsilon}$ respectively. Given a set of noisy coefficients \mathbf{d} , we want to estimate the wavelet coefficients belonging to the noiseless signal $\boldsymbol{\theta}$. Even if our wavelet transform is not orthogonal, but instead biorthogonal, experiments and statistical tests show that we can reasonably consider the components of $\boldsymbol{\eta}$ as i.i.d. normal random variables.

Given the set of coefficients used to represent shapes in the wavelet domain, we can assume that signal part has a sparse representation. This is shown in Figure 1. This sparsity motivates the wavelet shrinkage paradigm.

3 Hypothesis Testing and Bayesian Shrinkage

3.1 Bayesian Framework

The coefficients are clustered and analyzed by resolution level. We consider the 3 coordinates of our shape signal as independent from one another, and process

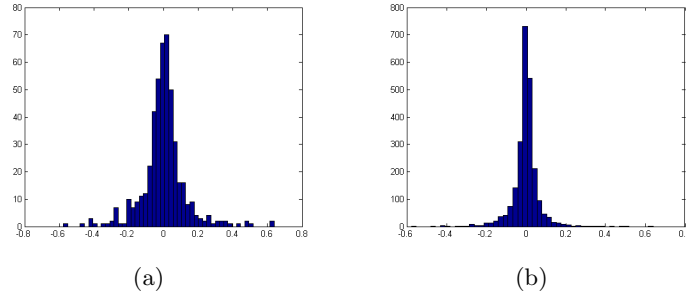


Fig. 1. Distribution of the X-signal wavelet coefficients from the fourth (a) and fifth (b) levels of resolution. We can see the sparsity of the representation.

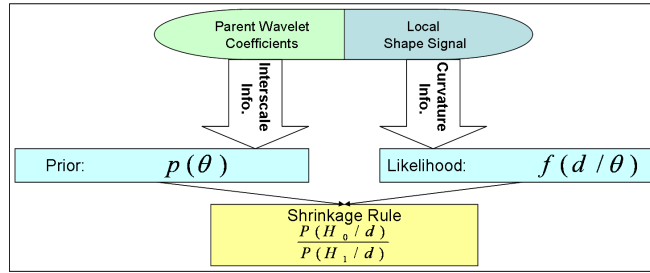


Fig. 2. Curvature information and interscale dependencies are entries of our Bayesian wavelet shrinkage framework.

each one separately. Thus, each coefficient carries three indices, which are the index of the level it belongs to, its index in that level and the coordinate $(x,y$ or $z)$ it encodes. For notational convenience the indices are dropped and θ - or d - will refer to an arbitrary coefficient from the detail spaces. The proposed selection model in the wavelet domain is based on a Bayesian hypotheses testing in which signal part θ is of interest. We set the null hypothesis as follows: $H_0 : \theta = 0$. For each wavelet coefficient subject to shrinkage, we test the hypothesis H_0 versus $H_1 : \theta \neq 0$ by estimating the posterior odds in favor of H_0 :

$$R_{Bayesian} = \frac{P(H_0 | d)}{P(H_1 | d)} = \frac{P(\theta = 0 | d)}{P(\theta \neq 0 | d)} \tag{3}$$

If $R_{Bayesian} > 1$, we set the coefficient to zero. Otherwise, the null hypothesis is rejected and we set $\theta = d$ as the signal part coefficient.

The different components of our Bayesian framework, prior and likelihood distributions, enable us to compute posterior probabilities $P(\theta | d)$ [7] and will be described below. As mentioned in section 1, interscale model and curvature estimation are both entries of our shrinkage rule as presented in Figure 2.

3.2 Prior Distribution and the Interscale Model

In order to stay consistent with the wavelet coefficient sparsity, we model the prior distribution as a weighted mixture of a point mass at zero δ_0 with a “spread” distribution ξ from the double exponential family $\xi(\theta) = \frac{1}{2\tau}e^{-|\theta|/\tau}$ as follows:

$$p(\theta) = \pi_0\delta_0 + \pi_1\xi(\theta), \quad \pi_0 + \pi_1 = 1. \quad (4)$$

This type of prior had been proposed in [7] for classical wavelets.

Here, we define π_0 as a function of neighboring wavelet coefficients from coarser resolution levels. This enables one to put more weight on the removal of noisy artifacts that are characterized by lonely high-valued coefficients with no neighboring coefficient relevance. On the other hand, if a coefficient likely to be shrunk in an individual analysis has a high magnitude neighborhood, its chance to be retained in the model increases. In practice, for a wavelet coefficient d , we select its neighboring points (typically its direct parents from the multiscale mesh), compute the average value c of their wavelet coefficients, which represents the “strength” of the neighborhood, and define π_0 as a continuous function of c , $\pi_0 = \pi_0(\text{neighbors}) = \pi_0(c) = 1 - \pi_1(c)$. We propose to model it as follows:

$$\begin{cases} \pi_0(c) = K & \text{if } |c| \leq T \\ \pi_0(c) = \alpha \cdot e^{-\beta \cdot |c|} & \text{if } |c| > T \end{cases} \quad (5)$$

where T, β are parameters to be set and $\alpha = K \cdot \exp(\beta \cdot T)$. Since T constitutes a threshold for wavelet coefficients, we estimate it using the universal thresholding rule [4]. K is typically .95, whereas β is chosen by bounding π_0 between 0.2 and 0.95.

3.3 Likelihood and Curvature Information

Likelihood Definition This section is the main contribution of our work. We chose the likelihood function $f(d | \theta)$ as a means to incorporate information on the curvature and the coarse variations of the surface into our shrinkage model. We want our final thresholding rule to be as follows: the more regular/flat a region of the surface is, the more likely a coefficient will be shrunk, unless it has very strong neighboring coefficients. In Figure 3, we show how this information can enhance our shrinkage results by adding curvature information to the interscale-based shrinkage model (as presented in section 3.2). Lower curvature regions of the shape undergo a much stricter shrinkage than the others, and it enables us to recover a smoother signal without losing significant information. The underlying idea is that the interscale model (via the prior distribution) has to bring much more convincing information for keeping a coefficient when curvature is low.

Since the noise is assumed to be Gaussian, we can write the likelihood function as a normal distribution centered in θ . In order to incorporate curvature information into the expression of the likelihood function we propose to weight the value of the observed coefficient d by a normalized curvature term κ that

will be close to 0 (*resp.* 1) for low (*resp.* very high) curvatures. This process constitutes a sort of pre-modeling stage where observed coefficients are multiplied by our curvature term:

$$f(d|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{|d \cdot \kappa - \theta|^2}{\sigma^2}\right) \quad (6)$$

where the noise variance σ^2 is estimated for each level of decomposition (see below).

Noise Variance Estimation We estimate the noise variance in a level-wise manner using power spectrum analysis in the wavelet domain. We derive a reasonable estimate of the noiseless spectrum for each fine level k by looking at the *log level energy*. This function decreases linearly with the levels of resolution for a noiseless signal and stays the same for noisy and noiseless signals at the coarse levels since the noise does not affect low frequency terms. Thus, we estimate the slope of the best linear fit for a range of levels k , which gives us an approximation of the noiseless *log level energy* function. Then, the difference between noisy and noiseless spectra gives an estimate of noise variance for each fine level. Finally, we rely on empirical relationships found between successive resolution levels in order to estimate noise variance at coarser levels, but again, this part of the signal does not capture high frequencies of noise.

Curvature Term The term κ is computed at each vertex and consists of the maximum absolute value of the two principal curvatures ($\max(|\kappa_1|, |\kappa_2|)$), which seems to perfectly fit our smoothing strategy. Indeed, we want our curvature term to influence shrinkage only in flat regions of the shape. Then, if and only if both principal curvatures are low valued, the coefficient will have a higher potential to be shrunk. We also tested our method with mean and Gaussian curvatures. However, mean curvature (*resp.* Gaussian curvature) would tend to oversmooth the signal if κ_1 and κ_2 had opposite signs (*resp.* if only one of the two was almost zero).

Now it remains to determine how to obtain such information from a noisy shape signal. The multiscale structure provides us with a powerful tool to do so. By temporarily completely removing the finer levels of resolution (via *linear wavelet shrinkage*), we obtain a smooth version of each surface and are then able to estimate local curvature with robustness and accuracy. The method used to compute the principal curvatures from a polyhedral mesh approximation was proposed in [13].

3.4 Hypothesis Testing

With the same notations, we incorporate equations (4), (5) and (6) into our posterior odds defined in section 3.1 and obtain:

$$R_{Bayesian} = \frac{\pi_0(c) e^{-(|d \cdot \kappa|^2 / \sigma^2)}}{\pi_1(c) \int e^{-(|d \cdot \kappa - \theta|^2 / \sigma^2)} \xi(\theta) d\theta} \quad (7)$$

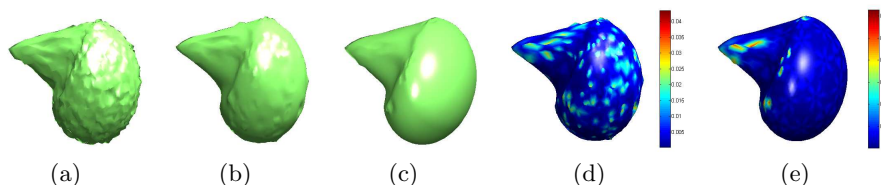


Fig. 3. Effect of curvature information on shrinkage rule for synthetic data. (a) represents a noisy synthetic shape. In (b) and (d), we show the recovered shape after shrinkage and the pointwise distance with the original shape when curvature is not considered, and (c) and (e) present the improvement brought into the model by the information on local curvature.

4 Experiments

We have applied our model to the denoising of 25 left hippocampus shapes obtained from MR images. As noted above, the shapes were hand-segmented by a radiologist and presented to us as triangulated surfaces. We added Gaussian noise of variable variance σ_s^2 (see Table 1) to the original surfaces. Surfaces are equipped with a multiscale mesh defined by 5 levels of resolution and 2562 vertices. The coordinates x , y and z are treated as independent signals and we apply our shrinkage to the 2 finest resolution levels. In the prior distribution, we bound π_0 with values 0.1 and 0.95, which sets up our parameters.

In Figure 4, results of denoising are presented, where our data-driven method is compared to the inter- and intra-scale based model [11] and to the best results we could obtain with classical thresholding models with a threshold of the type $\lambda = \sqrt{k \cdot \log N} \cdot \sigma$ (where $k = 2$ is an upper bound, corresponding to universal thresholding defined in [4]). Estimation of optimal thresholds for non-Bayesian models is hard to accomplish and very optimistic when true signal is unknown but we want our model to be better than that optimal solution. Our method outperforms the others in terms of smoothness while keeping the reconstruction error low. We have noticed that this becomes more true as noise gets stronger and more irregularly distributed, or when the original shape exhibits very low curvature regions. In those cases, regular thresholding methods lose consistency and miss some context adaptive consideration.

One should note that we reach a very satisfying compression rate while keeping it under control. Our model allows for a very well-driven thresholding, without unreasonably oversmoothing the shape signal whereas uniform thresholding with overestimated threshold value would do so. Also, we observe that the error variance is reduced with our method and it confirms the robustness of the proposed scheme.

5 Discussion, Conclusions and Future Work

The statistical wavelet shrinkage method presented in this paper helps one obtain a regular, compressed and accurate surface by removing noisy coefficients while

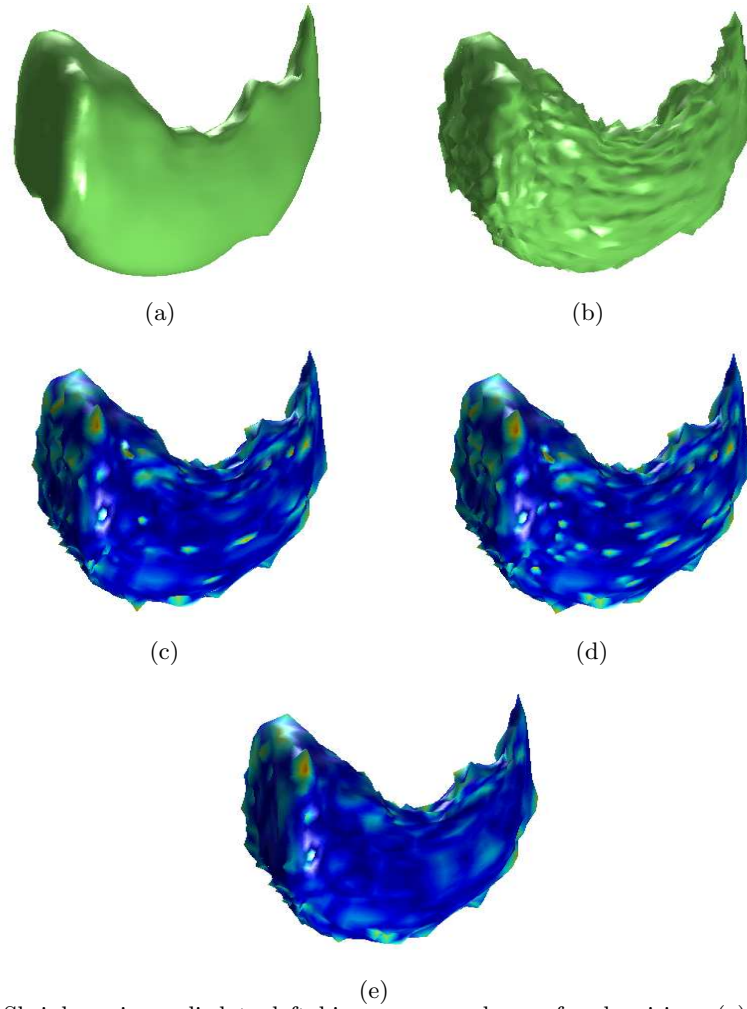


Fig. 4. Shrinkage is applied to left hippocampus shapes for denoising: (a) original shape, (b) noisy shape, (c) results with traditional thresholding, (d) inter-/intra-scale and (e) proposed Bayesian method (in (c),(d),(e) color is normalized reconstruction error at each vertex (in % of the shape bounding box) from blue (lowest) to red)

preserving intrinsic shape features. It allows for spatial consistency in the reconstruction and brings real improvement in case of very noisy data. This shrinkage method is very well adapted to hippocampus shape compression since it allows us to capture the main variations of the shape and its folding anatomy, while removing the irrelevant artifacts from the low curvature regions. We thus obtain a very good rate of compression through a pretty robust multiscale shrinkage.

Our overall algorithm depends upon several parameters. Specifically, noise variance σ^2 appears to have a relevant role in this scheme and we have developed our estimation procedure in 3.3.

Also, a part of our model requires shape curvature estimation. The robustness of such a computation is a natural issue since meshes are only approximations of real surfaces. We tested the robustness of the method used in the paper [13] by comparing results obtained for different meshing resolutions and concluded positively on the accuracy of the algorithm for our surfaces.

More generally, the mixture of information in our model reduces the sensitivity to parameter estimation whereas classical thresholding does not always guarantee to get an optimal threshold value.

Finally, even though the experimental results that are shown in this paper are obtained with Gaussian noise, the combination of our neighbor dependent model with the curvature driven term seems to exhibit some robustness under deviation from normality. That is a part of our on-going research. We plan to statistically analyze the clinical significance of such a model for hippocampus by evaluating how the proposed shrinkage allows to exhibit the relevant shape variations. We want to make sure that the signal the model captures contains all the information required for classification and diagnosis purposes.

		Left Hippocampus	
σ_s^2	Method	compression(%)	mean error(%)
0.06	Classical Threshold	86	1.0
	Inter/Intra-scale	86	1.0
	Proposed Model	88	0.9
0.1	Classical Threshold	87	1.1
	Inter/Intra-scale	87	1.1
	Proposed Model	89	1.0
0.12	Classical Threshold	89	1.2
	Inter/Intra-scale	87	1.3
	Proposed Model	90	1.1

Table 1. Results for hippocampus denoising. We present the compression rate (number of coefficients shrunk over total number of coefficients) and the mean reconstruction error averaged over all vertices (in % of the shape bounding box) for 3 different noise variances σ_s^2 . Comparison of the 3 methods described in Sec.4.

Acknowledgements

Allen Tannenbaum is also with the Department of Electrical Engineering, Technion, Israel where he is supported by a Marie Curie grant through the EU. This work was supported in part by grants from NSF, AFOSR, ARO, MURI, as well as by a grant from NIH (NAC P41 RR-13218) through Brigham and Women's Hospital. This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from <http://nihroadmap.nih.gov/bioinformatics>.

References

1. Nain, D., Haker, S., Bobick, A., Tannenbaum, A.: Multiscale 3d shape analysis using spherical wavelets. In: Proc. of MICCAI. (2005)
2. Yu, P., Grant, P.E., Qi, Y., Han, X., Ségonne, F., Pienaar, R., Busa, E., Pacheco, J., Makris, N., Buckner, R.L., Golland, P., Fischl, B.: Cortical surface shape analysis based on spherical wavelets. *IEEE Trans. Med. Imaging* **26**(4) (2007) 582–597
3. Schroeder, P., Sweldens, W.: Spherical wavelets: Efficiently representing functions on the sphere. *Computer Graphics* **29** (1995) 161–172
4. Donoho, D., Johnstone, I.: Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81** (1994) 425–455
5. Donoho, D., Johnstone, I.: Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association* **90**(432) (1995) 1200–1224
6. Abramovic, F., Benjamini, Y.: Adaptive thresholding of wavelet coefficients. *Computational Statistics and Data Analysis* **22** (1996) 351–361
7. Vidakovic, B.: Nonlinear wavelet shrinkage with bayes rules and bayes factors. *Journal of the American Statistical Association* **93** (1998) 173–179
8. F. Abramovic, T.S., Silverman, B.: Wavelet thresholding via a bayesian approach. *Journal of the Royal Statistical Society, Ser. B* **60** (1998) 725–749
9. Malfait, M., Roose, D.: Wavelet-based image denoising using a markov random field a priori model. *IEEE Trans. on Image Proc.* **6** (1997) 549–565
10. A. Pizurica, W. Philips, I.L., Acheroy, M.: A joint inter- and intrascale statistical model for bayesian wavelet based image denoising. *IEEE Trans. on Image Processing* **11** (2002) 545–557
11. X. Le Faucheur, B. Vidakovic, D.N., Tannenbaum, A.: Bayesian spherical wavelet shrinkage: Applications to shape analysis. In: Proc. of SPIE Optics East. (2007)
12. Sweldens, W.: The lifting scheme: a construction of second generation wavelets. *SIAM Journal on Mathematical Analysis* **29** (March 1998) 511–546
13. Taubin, G.: Estimating the tensor of curvature of a surface from a polyhedral approximation. In: Proc. of ICCV. (1995)