FROM LOCAL POLYNOMIAL APPROXIMATION TO POINTWISE SHAPE-ADAPTIVE TRANSFORMS: AN EVOLUTIONARY NONPARAMETRIC REGRESSION PERSPECTIVE

Alessandro Foi and Vladimir Katkovnik

Institute of Signal Processing, Tampere University of Technology,
P.O. Box 553, Tampere 33101, FINLAND - firstname.lastname@tut.fi

ABSTRACT

In this paper we review and discuss some of the theoretical and practical aspects, the problems, and the considerations that pushed our research from the one-dimensional LPA-ICI (local polynomial approximation - interesection of confidence intervals) algorithm [27] to the development of powerful transform-based methods for anisotropic image processing [18, 15, 19, 20, 8, 21].

In this paper we do not present a new algorithm. Instead, we propose a different and *more general* interpretation of our recently-developed image denoising algorithms. In particular, we show how they can be treated as nonparametric estimators based on aggregation of a family of local estimates which are pointwise-adaptive in terms of both shape and order.

1. INTRODUCTION

The path that we follow begins from conventional local-polynomial estimators and their spatially adaptive anisotropic versions [29, 16, 12]. Despite the outstanding success demonstrated by these adaptive techniques, it must be observed that, at a local level, such estimators are characterized by a tightly restricted number of parameters (i.e. polynomial order). While the intrinsic stability of these simple models makes their spatial adaptation effective and robust, it also poses a significant limit to their fitting ability. Low-order polynomial models appear often too rigid for satisfactorily approximating all the various local behaviors of natural images.

Improved approximation can be achieved by means of higher-order models. Unfortunately, direct application of adaptive scale-selection procedures for higher-order models fails, because of the very large variance inherent of the higher-order estimates [25]. A number of compromises have been proposed in the literature: for example, exploiting a lower-order (usually zero-order) model in the spatial-adaptation step and a higher-order one (first- or second- order) for the approximation and estimation; or separately producing few spatially-adaptive approximations of different fixed orders (e.g. zero, first, and second order) and then combining them, with some (adaptive or non-adaptive) weights (see, e.g., [6], [23]). However, the choice of the final (i.e. maximum) order remains crucial. It should be noted that if this order is unrestricted there is no smoothing, since the model attains a perfect fit to the noisy observations. It turns out that it is necessary not only to adapt with respect to the spatial features in the image, but also with respect to the different orders which are required to model such various features. We solve this dilemma by decomposing the higher-order models in several orthogonal complements and adaptively compose the most appropriate model from the corresponding subspaces using conventional estimators such as thresholding. This is done at a local level, on an adaptive starshaped neighborhood defined via a reliable zero-/first-order multidirectional estimator, namely the anisotropic LPA-ICI [16, 29]. In this way, we come to pointwise spatially- and order-adaptive anisotropic polynomial estimators.

We make one step further, and relax the procedure by replacing polynomial models with more general transforms. In particular, we concentrate on the shape-adaptive variants of the discrete cosine transform (DCT) because of their near-optimal decorrelation ability for natural images. Low-complexity versions of such transforms exist and can be exploited to develop fast algorithms based on the above paradigm. We remark that these are not simply some speculations of abstract interest. A breakthrough in image restoration performance corresponds to each progressive step taken from local polynomial estimators towards pointwise shape-adaptive transforms. In particular, the pointwise shape-adaptive DCT algorithms for denoising [21, 19], deblurring [15], deblocking [21, 20], and inverse-halftoning [8] deliver estimates of an unprecedented quality. Such algorithms typically outperform in terms of perceptual and objective criteria all the other techniques existing in the open literature.

2. BACKGROUND

The pointwise local approximations were intensively investigated and discussed during the last years. We refer to the books [10] and [36] for a nice and detailed overview of local polynomial modeling and related literature. In the maximum-likelihood framework these approximations can be applied to many technical and scientific problems. Important issues for local modeling is the choice of localization (window-size, bandwidth, smoothing) parameters. Different types of model-selection techniques based on the asymptotic expansion of the local-likelihood are mentioned in [11] and [25], which include global and variable

This work was partially supported by the Academy of Finland, project No. 213462 (Finnish Centre of Excellence program 2006 - 2011).

bandwidth selection.

The problem of pointwise-adaptive estimation has received a new powerful impetus in connection with a number of new methods developed for adaptive scale/bandwidth selection. Various developments of this idea and various statistical rules are the subject of a thorough study in the papers [34], [23], [43]. This sort of methods can be treated as a quality-of-fit statistics applied locally in pointwise manner and are known as "Lepski's method" [32]. There is a number of modifications of this approach and different adaptive algorithms with successive applications reported in many publications.

However, multiple studies produced in statistics show that the finite sample performance of estimators based on bandwidth or model selection is often rather unstable (e.g. [4]). This point is particularly critical for the pointwise model selection procedures like Lepski's method, [33]. In spite of the nice theoretical properties (see [34], [33], or [43]), the resulting estimates suffer from a high variability due to a pointwise model choice, especially for a large noise level.

This suggests that in some cases, the attempt to identify the true model maybe is not necessarily the right thing to do. One approach to reduce the variability in adaptive estimation is known as "model mixing" or "aggregation". Let us clarify this idea. In local approximation, the estimation neighborhood is selected adaptively. In order to do it, a number of estimates is calculated and compared. This is how Lepski's method works. Mixing or aggregation assumes to use a combination (in particular weighted mean of the estimates) instead of attempts to select the best estimate. A number of aggregation procedures has been developed (e.g. [48], [49], [47], [3]). It was proved that, with a proper combination, the mixed estimate achieves the minimal estimation risks, i.e. its performance is not worse than the performance of the best of the individual estimates. Thus, aggregation can be used instead of selection, and provides estimates that are more stable with respect to the random noise (e.g. [48], [49], [5]). In the regression framework, optimal aggregation procedures have been proposed and their properties studied (e.g. [26], [47], [2], [35]).

In image processing a lot of our work has been done in the line of the best-model selection in order to produce efficient adaptive algorithms. Success has been achieved due two important elements of the developed approach: starshaped anisotropic neighborhoods used for estimation and an adaptive scale-selection procedure equipped with special non-symmetric directional filters.

For the scale selection, the intersection of confidence intervals (ICI) – a particular instance of Lepski's approach – has proved to be extremely successful. The developed technique has been applied to many standard image processing problems: denoising [29, 16], deblurring [29, 28], inverse-halftoning [17], restoration from Poissonian and signal-dependent noise [30, 14, 13].

However, overall our expertise in the field says that the potential of these adaptive pointwise approximations is practically exhausted. Even though some further improvements and optimization can be produced, the principal limits of the approach have been achieved. Further development, in order to have a practical and concrete significance, requires novel ideas.

Our recently proposed algorithms differ from the conventional anisotropic LPA-ICI in a number of fundamental aspects.

Firstly, the *pointwise* approximations of the signal are replaced by *local* approximations over neighborhoods (we remind that in LPA-ICI algorithms the estimates are always pointwise, meaning that each LPA is exploited to estimate a unique pixel only).

Further, the *fixed* polynomial order is replaced by an *adaptive* order (which is restricted only by the size of the neighborhood where the local model is fitted).

Finally, we exploit *aggregation*, in order to compensate to the variability of these local estimates and thus to improve the quality of the final estimate.

3. PROCEDURE

Suppose we are given noisy data in the form $z(x) = y(x) + n(x)$, $x \in X$, where y is a signal of interest and n is some additive noise. Our approach to estimation for a point x_0 can be roughly described as the following four stage procedure.

Stage I (*spatial adaptation*)**:** For every $x \in X$, define a neighborhood $\tilde{U}_x^+$ of x where a simple (e.g. constant or linear) signal model fits the data;

Stage II (*order selection*)**:** apply some localized transform (parametric series model) to the data on the set $\tilde{U}_x^+$, use a thresholding operator (model selection procedure) in order to identify the significant (i.e. non-zero) elements of the transform (and thus the order of the parametric model).

Stage III (*approximation*)**:** Calculate, by inverse-transformation of the significant elements only, the corresponding estimates $\hat{y}_{\tilde{U}_x^+}(v)$ of the signal for all $v \in \tilde{U}_x^+$. These $\hat{y}_{\tilde{U}_x}$ are calculated for all $x \in X$.

Stage IV (*aggregation*)**:** Let $x_0 \in X$ and $I_{x_0} = \left\{x \in X : x_0 \in \tilde{U}_x^+\right\}$ be the set of the centers of the neighborhoods which have x_0 as a common point. The final estimate $\hat{y}(x_0)$ is calculated as an aggregate of $\left\{\hat{y}_{\tilde{U}_x^+}(x_0)\right\}_{x \in I_{x_0}}$.

Let us discuss more in detail these four stages.

3.1 Adaptive neighborhood (scale/size/shape) selection

The use of neighborhoods in which the signal is nearly constant or otherwise very homogeneous provides a dramatic improvement to the effectiveness of our method. In these neighborhoods there are no strong irregularities such as jumps, change-points, or change-lines. Therefore the signal allows for very sparse representations in the transform domain. It means that the true signal is transformed into few high-magnitude coefficients which are significant. All other coefficients have small amplitude and convey little information about the signal of interest.

To adaptively select these neighborhoods we exploit a simplified version of the Anisotropic LPA-ICI ([15, 21]).

It is interesting to observe that various kinds of preprocessing have been recently proposed in combination with wavelets (e.g. [45], [9], [37]) and harmonic transforms ([41], and references therein) in order to suppress the disturbing effects (e.g. ringing) that commonly-used transforms produce in correspondence with strong edges. In our approach preprocessing is unnecessary, since the transforms are used on supports which are pointwise adaptive with respect to the signal.

3.2 Adaptive order selection

Stage II is a typical model-selection procedure looking for the proper order of a polynomial series expansion. Let us consider an orthonormal basis $\left\{\psi^{(i)}_{\tilde{U}^+_x}\right\}_{i=1}^{\left|\tilde{U}^+_x\right|}$ for $L^2\left(\tilde{U}^+_x\right)$, where $\left|\tilde{U}^+_x\right|$ denotes the cardinality (number of elements) of $\tilde{U}^+_x$. In particular, we employ orthonormal polynomials and bases obtained from harmonic functions.

Let $W^{(i)}_{\tilde{U}^+_x}$ be the eigenspace generated by $\psi^{(i)}_{\tilde{U}^+_x}$. The space $L^2\left(\tilde{U}^+_x\right)$ can be trivially decomposed as the direct sum of these orthogonal eigenspaces

$$L^2\left(\tilde{U}^+_x\right) = \oplus_{i=1}^{\left|\tilde{U}^+_x\right|} W^{(i)}_{\tilde{U}^+_x}. \qquad (1)$$

Our adaptive model-selection aims at identifying a subspace $V_x \subseteq L^2\left(\tilde{U}^+_x\right)$ of the form

$$V_x = \oplus_{i\in S_x} W^{(i)}_{\tilde{U}^+_x} \qquad (2)$$

which is suitable for approximating y on $\tilde{U}^+_x$. The set S_x in (2) indicates the significant (or active) spectral components of y with respect to the decomposition (1).

Hard-thresholding appears as the most natural approach to identify the set S_x and thus the subspace V_x. It defines

$$S_x = \left\{ i : \left| \left\langle z_{|\tilde{U}^+_x}, \psi^{(i)}_{\tilde{U}^+_x} \right\rangle_{L^2\left(\tilde{U}^+_x\right)} \right| \geq \tau_x \right\},$$

where $z_{|\tilde{U}^+_x}$ is the restriction of z to $\tilde{U}^+_x$, $\langle\cdot,\cdot\rangle_{L^2\left(\tilde{U}^+_x\right)}$ is the inner-product in $L^2\left(\tilde{U}^+_x\right)$, and τ_x is a threshold parameter[1].

Many different and more sofisticated techniques can be applied here in the place of hard-thresholding.

Observe that V_x does not necessarily contain all $W^{(i)}_{\tilde{U}^+_x}$ with $i < \max(S_x)$, i.e. S_x is not necessarily a continuous spectrum. On the contrary, it is desirable (and typically happens) to have only very few significant components, or, in other words, a sparse representation.

[1]In our algorithms, τ_x is a function of $\left|\tilde{U}^+_x\right|$, similarly to Donoho's universal threshold.

3.3 Approximation

Once the adaptive model has been identified, we can proceed to construct the corresponding local approximation of the signal. Similarly to the LPA, we look for the best approximation in terms of L^2-norm. In our case, the best approximation is given as the projection of $z_{|\tilde{U}^+_x}$ onto V_x. Obviously, it is given by the reconstruction formula

$$\begin{aligned}\hat{y}_{\tilde{U}^+_x} &= \operatorname{argmin}_{\psi\in V_x} \left\| z_{|\tilde{U}^+_x} - \psi \right\|_{L^2\left(\tilde{U}^+_x\right)} = \\ &= \sum_{i\in S_x} \left\langle z_{|\tilde{U}^+_x}, \psi^{(i)}_{\tilde{U}^+_x} \right\rangle_{L^2\left(\tilde{U}^+_x\right)} \psi^{(i)}_{\tilde{U}^+_x}.\end{aligned}$$

Choice of the basis

Let us give some remarks on the choice of the orthonormal basis used in Stage II and Stage III. All the basis elements are supported on the adaptive neighborhood $\tilde{U}^+_x$, and orthonormality holds with respect to the norm of $L^2\left(\tilde{U}^+_x\right)$. It means that to each differently-shaped neighborhood corresponds a different adapted basis. The practical design of bases with arbitrarily-shaped supports has been subject of extensive research. The most naive, and arguably also the most effective and computationally demanding approach is essentially the successive orthonormalization (i.e. Gram-Schmidt algorithm). In principle, any basis defined on a superset of $\tilde{U}^+_x$ can be orthonormalized onto $\tilde{U}^+_x$. Nevertheless, traditionally only polynomials and harmonic bases (Fourier, cosine, etc.) are used for this scope. The application of such bases for image compression had been proposed almost two decades ago [22] but never found concrete use because of the excessive computational burden and some technical complications [38]. Other approaches based on partial differential equations have been suggested (e.g. [40]). However, the most successful family of orthonormal bases defined on arbitrarily-shaped supports is the shape-adaptive DCT (SA-DCT) [42, 31], which is a low-complexity transform implemented in a separable-like manner. It is shown in Section 5 that the mentioned bases and transforms eventually produce very similar results. Therefore, when concrete denoising applications are considered, the SA-DCT is without any doubt the transform of choice.

Since transforms obtained by orthogonalization require precalculation, we shall refer to them as "shape-adapted" transforms (a term proposed by Gilge [22]). The low-complexity SA-DCT does not require precalculation and is used directly. It is thus truly a "shape-adaptive" transform. We note that unfortunately in the literature there is no complete agreement on this terminology.

3.4 Aggregation

Aggregation is possibly the most important element in our approach and it strongly influences the design of all preceding stages.

The aggregation implies that the final estimate $\hat{y}(x_0)$ depends on data given in the union of the neighborhoods $\bigcup_{x\in I_{x_0}} \tilde{U}^+_x$, as opposed to the anisotropic LPA-ICI, where

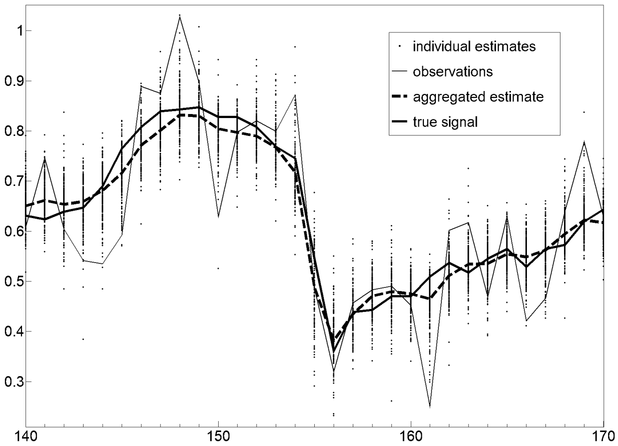

Figure 1: Aggregation in action. Details of a cross-section of length 31 pixels from the *Peppers* image (σ=25): the dots show all the individual estimates which are aggregated in order to obtained the final estimates at each position. For each pixel there are about 200 individual estimates.

the final estimate for a point depends on the data given in the corresponding neighborhood only. In this sense, in our aggregation the initial local estimate $\hat{y}_{\tilde{U}^+_{x_0}}$ is supplanted by a non-local one, with much larger areas of the data involved[2].

The stabilizing effect of aggregation is clearly depicted in Figure 1, for a cross-section of the noisy *Peppers* image. One can observe that the individual estimates are extremely volatile, and that the majority of them is not a particularly good estimate of the true signal. Numerous outliers can be seen. Aggregation aims at combining all these estimates in order to produce an "aggregated estimate" which is not worse (in terms of estimation risk) than the best of them. However, as it can be seen in the figure, all of the estimates $\left\{\hat{y}_{\tilde{U}^+_x}\right\}_{x\in I_{x_0}}$ are usually noisy and the averaging effect involved in the aggregation gives a remarkable improvement of the accuracy. In practice, one can observe that the aggregated estimate $\hat{y}\left(x_0\right)$ is often significantly better than any of the individual estimates $\left\{\hat{y}_{\tilde{U}^+_x}\left(x_0\right)\right\}_{x\in I_{x_0}}$.

Experimental study has been produced in order to identify the most effective and efficient way to combine all such individual estimates. For the sake of computational simplicity, it is desirable to use a convex combinations of $\left\{\hat{y}_{\tilde{U}^+_x}\right\}_{x\in I_{x_0}}$ where the factors for $\hat{y}_{\tilde{U}^+_x}$ can be calculated independently of $\left\{\hat{y}_{\tilde{U}^+_v}\right\}_{v\neq x}$. This allows to easily implement the aggregation by means of two separate buffers (a "numerator" for the weighted estimates and a "denominator" for the weights). However, beside the conventional average and weighted-averages, also robust stastistics such as trimmed means and median estimators were considered as possible choices for "aggregators". It was observed that, using adaptive weights w_x which depend simultaneously on the size of the neighborhood and on the number of significant or active elements of the local model[3], a simple weighted-average with weights constant on $\tilde{U}^+_x$ provides the best aggregation. An experimental analysis in [24] confirms our conclusion, at least for the simpler case of DCT transform with non-adaptive support of size 8×8. As a matter of fact, such weights are very natural, and have thus been used extensively within sliding-transform filters already for more than a decade (see [50], and references therein). Further considerations about these adaptive weights can be found in [21].

Overall, it follows that the individual estimates of Stage III are allowed to be quite noisy, and the final efficient noise removal can be enabled by the aggregation procedure. The biasedness of the individual estimates is much more damaging than their noisiness, as it cannot be easily removed by the aggregation. Thus, selection of the local neighborhoods $\tilde{U}^+_x$ on Stage I and the model selection (significant spectrum S_x and subspace V_x) on Stage II require a special care. We note that the ringing artifacts which can appear as a result of transform-based filtering are essentially bias-errors. These artifacts are typical when the transforms are applied on non-adaptive supports (loss of sparsity), or when thresholding (i.e. model selection) is too aggressive (loss of significant components).

4. COMMENTS

We wish to note that the term "model selection" is used in a different sense in different areas. In classical statistics it usually denotes basic function selection from a given set or from the library of these sets. In the nonparametric regression model-selection refers to the selection of the size of conventionally symmetric neighborhoods used for estimation. In our approach we use the polynomial and cosine transforms defined on neighborhoods with varying size and shape. Thus, here we approach to the model-selection problem in generalized form, assuming selection of the active elements of the transform as well as the size and the shape of the support of this transform.

5. EXAMPLES

We developed a number of novel and well-tested algorithms [18, 15, 19, 20, 8, 21] implementing the above ideas of the novel approach to nonparametric regression estimation. The proposed algorithms are quite general and can be applied in various areas. Our work concerns with image processing problems as one of the most competitive area of the signal processing. Overall, the proposed

[2] It is interesting to observe that the *recursive* anisotropic LPA-ICI [16, 12] presents a very similar behaviour, where the data used for estimation at a point x_0 is progressively enlarged by incorporating other neighborhoods which have x_0 as a common point. This happens at every iteration, and starting from the second iteration even neighborhoods which do not contain x_0 might be used for the estimation at this point.

It is also because of this property that the recursive LPA-ICI provides better performance than the non-recursive anisotropic LPA-ICI (see Table 1).

[3] i.e. the number of coefficients retained by thresholding, and thus the average variance of the reconstructed local estimate.

technique	*Cameraman*	*Peppers*	*Lena*	*Boats*
Anisotropic LPA-ICI (zero order)	28.16	28.34	29.67	27.92
Recursive Anisotropic LPA-ICI (zero-first order mixture)	28.68	29.16	30.43	28.58
Pointwise "shape-adapted" orthonormal polynomials	28.89	29.61	31.34	29.08
Pointwise "shape-adapted" DCT	**28.99**	**29.69**	**31.43**	**29.20**
Pointwise shape-adaptive DCT (SA-DCT) (low-complexity)	28.90	29.52	31.27	29.09

Table 1: *PSNR* (dB) results for denoising from additive Gaussian white noise (σ=25) for a few test images with different algorithms for anisotropic nonparametric estimation.

technique (two-stage, with empirical Wiener filtering)	*Cameraman*	*Peppers*	*Lena*	*Boats*
Pointwise "shape-adapted" orthonormal polynomials	29.13	30.03	31.68	29.43
Pointwise "shape-adapted" DCT	**29.19**	**30.05**	**31.76**	**29.51**
Pointwise shape-adaptive DCT (SA-DCT) (low-complexity)	29.11	29.92	31.66	29.47

Table 2: *PSNR* (dB) results for denoising with transform-domain empirical Wiener filtering. The reference estimate used in Wiener filtering is the one obtained by the respective algorithm, as shown in Table 1. The estimates obtained by this two-stage procedure are significantly better than those obtained by just a single stage.

algorithms demonstrate breakthrough results, essentially outperforming the previous generation of the algorithms based on pointwise approximation with a fixed polynomial order. On many occasions, the results by the new pointwise SA-DCT algorithms overcome the best achievements in the field. We refer the reader to [15] and [21] for technical details about the implementation of these algorithms and for extensive comparative analyses. Here our scope is different and we present only a limited set of experiments, mainly to highlight the improvement of these new methods over the anisotropic LPA-ICI and to show that – in terms of quality of results – the differences between orthonormal polynomials and SA-DCT is not significant.

Table 1 gives *PNSR* results for the denoising of a few standard images. We consider the basic and recursive anisotropic LPA-ICI and three realizations of the presented approach using three different transforms for arbitrarily-shaped supports: "shape-adapted" orthonormal polynomials and cosines bases (obtained by successive orthogonalization), and the low-complexity shape-adaptive DCT (SA-DCT). The "shape-adapted" DCT provides the best results in all experiments, thus demonstrating a better approximation ability for natural images than the "shape-adapted" orthonormal polynomials. The low-complexity SA-DCT achieves a performance which is extremely close to that of the more complex transforms, but at a much lower computational cost[4].

Table 2 shows results obtained by implementing the proposed approach with a second stage, where empirical Wiener filtering is used instead of hard-thresholding. For this Wiener filtering the reference estimate is the one obtained from the first stage (whose *PSNR* is given in Table 1). The two-stage implementation further improves the denoising performance.

We conclude with a visual illustration of some of the concepts discussed in the previous sections. Figure 2 shows, for a given noisy image (σ=15, *PSNR*=24.58dB), several quantities corresponding to individual locations $x \in X$. Orthonormal polynomials are used for this simulation.

The size of the adaptive neighborhood of the point x, $\left|\tilde{U}_x^+\right|$, is the most basic concept of the whole approach. In this example, the largest allowed neighborhood is – for computational reasons – a 17×17 square block, which corresponds to $\left|\tilde{U}_x^+\right|$=289, shown as white in the figure. The smallest neighborhood is a single pixel, $\left|\tilde{U}_x^+\right|$=1, shown as black. Observe that the neighborhoods are large where the image is smooth, whereas they shrink in the vicinity of edges. The variance of the eight directional adaptive-scales which define the actual shape of the neighborhood, $\mathrm{var}\left\{\left\{h_{\theta_i}^+\right\}_{i=1}^8\right\}$, give a good indication of its anisotropy. Narrow, elongated neighborhoods have scales of distinctively different values, whereas rounder neighborhoods have more homogeneous scales (roughly corresponding to a "radius" of the neighborhood). The most anisotropic neighborhoods (shown as light gray) are usually found next to edges, where the neighborhoods are asymmetrical, with larger scales for those directions opposite to the edge and shorter ones in the direction across the edge. The varying size and anisotropy clearly demonstrate the spatial adaptivity of the neighborhoods $\tilde{U}_x^+$.

As in Equations (1)-(2), for a neighborhood of size $\left|\tilde{U}_x^+\right|$ we can have an adaptive local model with $|S_x| \leq \left|\tilde{U}_x^+\right|$ dimensions (i.e. number of active components, dimension of V_x). It is interesting to see that, even where the image is rather smooth, high-order models are very frequently used: in this example, the adaptive polynomial order (i.e. maximum degree of the polynomial components of $\left\{\psi_{\tilde{U}_x^+}^{(i)}\right\}_{i\in S_x}$) goes up to 30 (shown in white). Nevertheless, the number of active compoments is usually much smaller (and thus the spectrum is highly fragmented). The sparsity achieved by the local estimate $\hat{y}_{\tilde{U}_x^+}$

[4]Complexity and implementation issues related to these approaches are discussed in [15] and [21].

Figure 2: A visualization of some of the elements and concepts which appear in the proposed approach (see text).

can be expressed by the ratio $\frac{|S_x|}{\left|\tilde{U}_x^+\right|}$ between the dimension of V_x and the size of the neighborhood. Sparser representations correspond to smaller ratios (shown as darker grays). Smoother areas of the image allow to achieve very sparse local estimates, regardless of the degree of their polynomial building elements $\left\{\psi_{\tilde{U}_x^+}^{(i)}\right\}_{i\in S_x}$. Sharper details of the image, require instead a larger proportion of basis elements into the estimate.

The number of aggregated estimates $|I_x|$ is a direct indicator of the overcompleteness of the method. This number depends only on the configuration of the adaptive neighborhoods which is the result of the first stage of the procedure (Section 3). It is not influenced by the outcome of the three subsequent stages. On the contrary, the sum of the weights $\{w_v\}_{v\in I_x}$ used in aggregation can be treated as a pointwise goodness of the final aggregated estimate. This estimate (PSNR=32.87dB) is visually good, with well reconstructed details and no significant artifacts.

6. FUTURE DEVELOPMENTS

We aim at the development of an algorithmically efficient technique for spatial adaptation which shall be based on a *local* comparison of *adaptive higher-order* estimates rather than on the *pointwise* comparison of *lower-order* ones, currently used. Statistical criteria to enable such comparisons are known. However, the calculation of all estimates to be compared represents a computationally unacceptable task. Exploiting essential simplifications seems an unavoidable necessity. Techniques based on block-similarity (e.g. [7]) suggest alternative ways to simultaneosly incorporate spatial-adaptation and aggregation into simple transform-domain operations.

REFERENCES

[1] Abramovich, F., and B.W. Silverman, "Wavelet Decomposition Approaches to Statistical Inverse Problems", *Biometrika*, vol. 85, no. 1., pp. 115-129, March, 1998.

[2] Belomestny, D., and V. Spokoiny, "Spatially adaptive local likelihood modelling via stagewise aggregation", Preprint no. 1000, Weierstrass Institute, Berlin, 2004.

[3] Belomestny, D., and V. Spokoiny, "Spatial aggregation of local likelihood estimates with applications to classification", submitted, 2006.

[4] Breiman, L., "Stacked regression", *Machine Learning*, vol. 24, pp. 49-64, 1996.

[5] Catoni, O., *Statistical learning theory and stochastic optimization*, Lecture Notes in Mathematics, 1851, Springer-Verlag, Berlin, 2004.

[6] Cleveland, W.S., and C. Loader, "Smoothing by local regression: principles and methods", *Statistical theory and computational aspects of smoothing*, Springer, New York, pp. 10-49, 1996.

[7] Dabov, K., A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering", *Proc. SPIE El. Imaging 2006, Image Process.: Algorithms and Systems V*, 6064A-30, San Jose, California USA, 2006.

[8] Dabov, K., A. Foi, V. Katkovnik, and K. Egiazarian, "Inverse halftoning by pointwise shape-adaptive DCT regularized deconvolution", *Proc. 2006 Int. TICSP Workshop Spectral Meth. Multirate Signal Process., SMMSP 2006*, Florence, September 2006.

[9] Daubechies, I., and G. Teschke, "Variational image restoration by means of wavelets: Simultaneous decomposition, deblurring, and denoising", *Applied and Computational Harmonic Analysis*, vol. 19, no. 1, pp. 1-16, July 2005.

[10] Fan, J., and I. Gijbels, *Local polynomial modelling and its application*. London, Chapman and Hall, 1996.

[11] Fan, J., M. Farmen, and I. Gijbels, "Local maximum likelihood estimation and inference", *J. Royal Statist. Soc.*, ser. B, 60, pp. 591-608, 1998.

[12] Foi, A., *Anisotropic nonparametric image processing: theory, algorithms and applications,* Ph.D. Thesis, Dipartimento di Matematica, Politecnico di Milano, Italy, ERLTDD-D01290, April 2005.

[13] Foi, A., S. Alenius, M. Trimeche, V. Katkovnik, and K. Egiazarian, "A spatially adaptive Poissonian image deblurring", *Proc. IEEE 2005 Int. Conf. Image Processing, ICIP 2005,* pp. 925-928, September 2005.

[14] Foi, A., R. Bilcu, V. Katkovnik, and K. Egiazarian, "Anisotropic local approximations for pointwise adaptive signal-dependent noise removal", *Proc. of XIII European Signal Proc. Conf., EUSIPCO 2005*, Antalya, September 2005.

[15] Foi, A., K. Dabov, V. Katkovnik, and K. Egiazarian, "Shape-adaptive DCT for denoising and image reconstruction", *Proc. SPIE El. Imaging 2006, Image Process.: Algorithms and Systems V*, 6064A-18, January 2006.

[16] Foi, A., V. Katkovnik, K. Egiazarian, and J. Astola, "A novel anisotropic local polynomial estimator based on directional multiscale optimizations", *Proc. 6th IMA Int. Conf. Math. in Signal Processing*, Cirencester (UK), pp. 79-82, 2004.

[17] Foi, A., V. Katkovnik, K. Egiazarian, and J. Astola, "Inverse halftoning based on the anisotropic LPA-ICI deconvolution", *Proc. 2004 Int. TICSP Workshop Spectral Meth. Multirate Signal Proc., SMMSP 2004,* pp. 49-56, Vienna, September 2004.

[18] Foi, A., V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT as an overcomplete denoising tool", *Proc. 2005 Int. TICSP Workshop Spectral Meth. Multirate Signal Process., SMMSP 2005*, pp. 164-170, Riga, June 2005.

[19] Foi, A., V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT denoising with structure preservation in luminance-chrominance space", *Proc. of the 2nd International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM2006*, Scottsdale, AZ, January 2006.

[20] Foi, A., V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality deblocking of compressed color images", *Proc. 14th European Signal Process. Conf., EUSIPCO 2006*, Florence, September 2006.

[21] Foi, A., V. Katkovnik, and K. Egiazarian, "Pointwise Shape-Adaptive DCT for High-Quality Denoising and Deblocking of Grayscale and Color Images", (in review) *IEEE Trans. Image Process.*, April 2006.

[22] Gilge, M., T. Engelhardt, and R. Mehlan, "Coding of arbitrarily shaped image segments based on a generalized orthogonal transform", *Signal Processing: Image Communication*, vol. 1, no. 2, pp. 153-180, October 1989.

[23] Goldenshluger, A., and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression", *Math. Meth. Statistics*, vol. 6, pp. 135-170, 1997.

[24] Guleryuz, O.G., "Weighted overcomplete denoising", *Proc. Asilomar Conference on Signals and Systems*, Pacific Grove, CA, November 2003.

[25] Hart, J.D., *Nonparametric smoothing and lack-of-fit tests*, Springer-Verlag, New York, 1997.

[26] Juditsky, A., and A. Nemirovski, "Functional aggregation for nonparametric estimation", *Annals of Statistics*, vol. 28 pp. 682-712, 2000.

[27] Katkovnik, V., "A new method for varying adaptive bandwidth selection", *IEEE Trans. Signal Proc.,* vol. 47, no. 9, pp. 2567-2571, 1999.

[28] Katkovnik, V., K. Egiazarian, and J. Astola, "A spatially adaptive nonparametric regression image deblurring", *IEEE Trans. on Image Processing*, vol. 14, no. 10, pp. 1469-1478, October 2005.

[29] Katkovnik, V., A. Foi, K. Egiazarian, and J. Astola, "Directional varying scale approximations for anisotropic signal processing", *Proc. XII European Signal Proc. Conf., EUSIPCO 2004,* Vienna, pp. 101-104, September 2004.

[30] Katkovnik, V., A. Foi, K. Egiazarian, and J. Astola, "Anisotropic local likelihood approximations", *Proc. of Electronic Imaging 2005*, 5672-19, 2005.

[31] Kauff, P., and K. Schuur, "Shape-adaptive DCT with block-based DC separation and ΔDC correction", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 3, pp. 237-242, 1998.

[32] Lepski, O., "On one problem of adaptive estimation in Gaussian white noise", *Theory Probab. Appl*., vol. 35, no. 3, pp. 454-466, 1990.

[33] Lepski, O., and V. Spokoiny, "Optimal pointwise adaptive methods in nonparametric estimation", *Annals of Statistics,* vol. 25, no. 6, pp. 2512-2546, 1997.

[34] Lepski, O., E. Mammen, and V. Spokoiny, "Ideal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selection", *Annals of Statistic*s, vol. 25, no. 3, pp. 929-947, 1997.

[35] Leung, G., and A.R. Barron, "Information Theory and Mixing Least-Squares Regressions", *IEEE Trans. Inf. Theory*, vol. 52., no. 8, pp. 3396-3410, August 2006.

[36] Loader, C.R., *Local likelihood density estimation*, Academic Press, 1996.

[37] Peotta, L., L. Granai, and P. Vandergheynst, "Image compression using an edge adapted redundant dictionary and wavelets", *Signal Processing*, vol. 86, no. 3., pp. 444-456, March 2006.

[38] Philips, W., "Orthogonal base functions on a discrete two-dimensional region", *ELIS Technical Report DG 91-20*, Dept. of Electronics and Inform. Syst., Universiteit Gent, Belgium*,* November 1993.

[39] Polzehl, J., and V. Spokoiny, "Propagation-Separation Approach for Local Likelihood Estimation", *Probab. Theory Related Fields*, vol. 135, no. 3, pp. 335-362, July 2006

[40] Saito, N., "Geometric harmonics as a statistical image processing tool for images defined on irregularly-shaped domains", *Proc. IEEE Statistical Signal Process. Workshop*, Bordeaux, France, July 2005.

[41] Saito, N., and J.F. Remy, "The polyharmonic local sine transform: A new tool for local image analysis and synthesis without edge effect", *Applied and Computational Harmonic Analysis*, vol. 20, no. 1, pp. 41-73, January 2006.

[42] Sikora, T., "Low complexity shape-adaptive DCT for coding of arbitrarily shaped image segments", *Signal Processing: Image Communication*, vol. 7, pp. 381-395, 1995.

[43] Spokoiny, V., "Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice", *Annals of Statistics*, vol. 26, no. 4, pp. 1356-1378, 1998.

[44] Staniswalis, J.C., "The kernel estimate of a regression function in likelihood-based models", *Journal of the American Statistical Association*, vol. 84, pp. 276-283, 1989.

[45] Starck, J.-L., M. Elad, and D.L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach", *IEEE Trans. Image Processing*, vol. 14, no. 10, pp. 1570-1582, Oct. 2005.

[46] Tibshirani, J.R., and T.J. Hastie, "Local likelihood estimation", *Journal of the American Statistical Association*, vol. 82, pp. 559-567, 1987.

[47] Tsybakov, A., "Optimal rates of aggregation", *Computational Learning Theory and Kernel Machines* (B. Scholkopf and M. Warmuth, eds.), Lecture Notes in Artificial Intelligence, 2777, Springer, pp. 303-313, 2003.

[48] Yang, Y., "Regression with multiple candidate models: selecting or mixing?", *Statistica Sinica*, vol. 13, pp 783-809, 2003.

[49] Yang, Y., "Aggregating regression procedures to improve performance", *Bernoulli*, vol. 10, pp. 25-47, 2004.

[50] Yaroslavsky, L., K. Egiazarian, and J. Astola, *Transform Domain Image Restoration Methods, Review, Comparison and Interpretation*, TICSP Series no. 9, Tampere, 2000.