Anisotropic Regularization of Posterior Probability Maps Using Vector Space Projections. Application to MRI Segmentation

Conference Paper *in* Lecture Notes in Computer Science · February 2003

DOI: 10.1007/978-3-540-45210-2_54 · Source: DBLP

Anisotropic Regularization of Posterior Probability Maps Using Vector Space Projections. Application to MRI Segmentation

M.A. Rodriguez-Florido[1], R. Cárdenes[1], C.-F. Westin[2], C. Alberola[3], and J. Ruiz-Alzola[1,2]

[1] Centro de Tecnología Médica
Universidad de Las Palmas de Gran Canaria and
Hospital de Gran Canaria Dr. Negrín
Barranco de La Ballena s/n - 35020 - Canary Islands - Spain
{marf,ruben,jruiz}@ctm.ulpgc.es
http://www.ctm.ulpgc.es

[2] Laboratory of Mathematics in Imaging
Brigham and Women's Hospital,
Harvard Medical School - Boston (MA) - USA
westin@bwh.harvard.edu

[3] Laboratorio de Procesado de Imagen
Universidad de Valladolid
Valladolid - Spain
caralb@yllera.tel.uva.es
http://www.lpi.tel.uva.es

Abstract. In this paper we address the problem of regularized data classification. To this extent we propose to regularize spatially the class-posterior probability maps, to be used by a MAP classification rule, by applying a non-iterative anisotropic filter to each of the class-posterior maps. Since the filter cannot guarantee that the smoothed maps preserve their probabilities meaning (i.e., probabilities must be in the range $[0, 1]$ and the class-probabilities must sum up to one), we project the smoothed maps onto a probability subspace. Promising results are presented for synthetic and real MRI datasets.

1 Introduction

Classical approaches to optimal statistical classification usually resort to a filtering step followed by the classifier itself. Ideally the filter should whiten the data in order to achieve optimality with rules classifying independently every voxel. The complexity of images, even more with medical images, makes it unfeasible to design practical whitening filters. Therefore, independent voxel classification leads to suboptimal results with scattered miss-classifications. Several approaches to regularize the resulting classification can be thought of. For example, just to name some of them, it is possible to prefilter the images, to post-regularize the

R. Moreno-Díaz and F. Pichler (Eds.): EUROCAST 2003, LNCS 2809, pp. 597–606, 2003.

independent classification using relaxation labeling or mathematical morphology, or to use the more formal Markov Random Fields approach.

Teo et al. [13] have proposed a new segmentation technique where the class-conditional posterior probabilities maps obtained from the raw data, i.e. without any pre-filtering step, are smoothed using the well-known Perona-Malik diffusion scheme [8]. The classification is then carried out by using independent MAP (Maximum a Posteriori) rules on the smoothed posterior maps.

The results obtained with this technique are better than with other schemes because the regularizing diffusion produces piece-wise-constant posterior probability maps, which yield piece-wise "constant" MAP classifications. However, it presents some difficulties: Each posterior probability (a scalar field for every class) is smoothed independently by non-linear diffusion, ignoring the intrinsic relationship among them, and therefore not guaranteeing the posteriors to sum up to one for every voxel in each diffusion step. To resolve this, the authors normalize the addition of posterior probabilities after each discrete iteration, assuming that every class posterior probability is positive or zero.

Some extensions of this classification framework have been provided, specially in the fields of MRI (Magnetic Resonance Imaging) and SAR (Synthetic Aperture Radar). In particular, [12] has shown that the framework can be seen as a MAP solution of a discrete Markov Random Field with a non-interacting, analog discontinuity field, and [7] extended it by using a system of coupled PDE's in order to guarantee for the diffusion process that the posterior probabilities are positive and sum up to one in each voxel.

In this work we propose an alternative approach to regularize the posterior probability fields. The major differences to previous schemes are:

1. Our approach is strictly anisotropic. Notice that the conventional Perona-Malik diffusion scheme is only homogeneous [15], since it uses a scalar function of the norm of the gradient as diffusivity.
2. It is not based on coupled PDE's, but it uses anisotropic-adaptive multidimensional filtering on every class-posterior probability map. A local structure tensor, extracted from the input image, shapes the filter response adaptively, and the solution is obtained in a single step by linear combination of a basis filters outputs [3] [10].
3. In order for the filtered probabilities maps to sum up to one and be positive, a vector spaces projection approach POCS (Projection on Convex Sets) algorithm is proposed, i.e, the filtered maps are projected onto the bounded hyper-plane defined by this condition.

The paper is organized as follows: First, after this introduction to the problem, we review briefly the concepts of *regularization* and *vector spaces projections*, and their relationship. Second, we introduce the POCS approach to MAP segmentation. Then we review the multidimensional anisotropic filter used to smooth the posterior maps. And finally, we discuss our results with both synthetic and real datasets (MRI volumes), and propose some future lines of work.

2 Regularization and Vector Space Projections

There are several problems in image processing that are considered as *ill-posed* in the sense of Hadamard [1] and Tikhonov [14]: A problem is *ill-posed* if there is no guarantee of the solution existence, the solution is not unique, or the solution does not depend continuously on the input data (small variations of initial conditions of the problem load to big variations on the solution). Typical examples of ill-posedness are: edge detection, visual interpolation, structure from stereo, shape from shading, computational of optical flow, and typically inverse problems.

Regularization theory provides a convenient way to solve ill-posed problems and to compute solutions that satisfy prescribed smoothness constraints. In fact, *regularization* can be seen as the restriction of the space of admissible solutions by introducing *a priori* input data information. As Poggio et al. [9] say in their classical paper, to solve an *ill-posed* problem expressed like $\mathbf{Hx} = \mathbf{y}$, is to find the value of $\mathbf{x}$ that minimize the expression:

$$\| \mathbf{Hx} - \mathbf{y} \|^2 + \lambda \| \mathbf{Rx} \|^2 \tag{1}$$

where λ is the regularization coefficient, $\mathbf{R}$ the regularization stabilizing operator, and $\mathbf{H}$ is a linear operator.

Basically, we can distinguish two ways to regularize the solution:

- Variational Regularization ([14]). An energy functional of the regularized solution $f_r(\mathbf{x})$ is minimized with respect to the input data $f_i(\mathbf{x})$:

$$E[f_r] = \frac{1}{2} \int ((f_i - f_r)^2 + \sum_{k=1}^{\infty} \lambda_k (\frac{\partial^k f_r}{\partial \mathbf{x}^k})^2) d\mathbf{x} \tag{2}$$

  with $\lambda_k \geq 0$, such as if $\lambda_j = 0 \; \forall \; j{>}n$ the regularization is called *nth* order. If the energy functional can be expressed with only sums of quadratic terms of the derivatives of the solution, it is equivalent to a linear filtering given by a smoothing operator.
- Statistical approach ([5]) that fixes the statistical properties of the solution space. The Bayes estimation taking the appropriate probability distributions is reduced to an expression similar to the regularization in the sense of Tikhonov, and if the *a priori* information is modeled in terms of a Markov Random Field (MRF) the maximum a posteriori estimation of the MRF is equivalent to a variational principle of the form of (2), as we mentioned in the introduction.

Signals with different properties form different subspaces in the general Hilbert signal space. These subspaces can be overlapped or disjointed, depending on the constraints. Finding the best approximation of a signal in a subspace can be achieved by projecting orthogonally the signal onto it. For example, the traditional sampling paradigm with ideal prefiltering could be seen as an orthogonal projection of the signals onto the subspace of band-limited signals.

Therefore, projections onto *vector spaces* are useful tools to find a subspace formed by signals that satisfy multiple constraints. Then, an interesting question arises: *Could the regularization be seen as a "projection" of the initial signal onto a subspace that satisfies the imposed regularity properties?* A partially answer to this question is possible if we consider the Sobolev spaces [1], because nth order linear regularization is a mapping between Sobolev spaces, where the target space is included into the starting space: $\mathbf{R} : S_i \to S_r$ with $S_r \subset S_i$. However, the projection is a idempotent operator, and regularization can not be strictly called a projection. Hence, we can only give a partial affirmative answer to the previous question.

3 POCS-Based MAP Classification

The goal of any segmentation scheme is to label every voxel as belonging to one out of N possible classes $w_1 \ldots w_N$. We will assume that a posterior probability model $p(w_i/z(\mathbf{x}))$ of the class w_i conditioned on the image intensity z is known for every voxel $\mathbf{x}$. Therefore there are N scalar maps, corresponding to the posterior probabilities of each class for every voxel, which can be stacked in a vector map of posterior probabilities. It is obvious that the components of the vector map must be all greater than or equal to zero and sum up to one, i.e., the L_1-norm is equal to one for every voxel.

Conventional MAP proceeds by selecting the class with the maximum posterior probability, i.e. by choosing the greater component of the vector probability map at each voxel.

Now, let us consider that the vector probability map is regularized before any decision is made, in order to obtain a smooth behavior of the posterior probabilities inside regions and steep changes in their boundaries. To this extent the structure of the original image must be taken into account in order to obtain a structure tensor to steer the smoothing process: probabilities must be smoothed inside regions belonging to the same class and along their boundaries, but not across them. It is clear that the filtered vector map must also have all its components greater than or equal to zero and, for every voxel, the L_1-norm must be equal to one. This constraint can be seen at every voxel as the hyper-plane $P_1 + \cdots + P_N = 1$, where P_i corresponds to the $i-th$ component of the posterior probability vector, ie., to the posterior probability of the class i. See figure 1 for the three classes case.

This condition is not easy to meet for adaptive anisotropic smoothers in general. A first possibility to overcome this problem is to normalize the posterior vector map using a partition function, i.e, by dividing each vector component by the L_1-norm at that voxel. This normalization does not change the result of the MAP rule though it becomes crucial if more than one iteration (smoothing and normalization) are carried out before the MAP rule is applied.

[1] A Sobolev space of order n is the space of all functions which are square integrable, and have derivatives well-defined and square integrable up to order n.

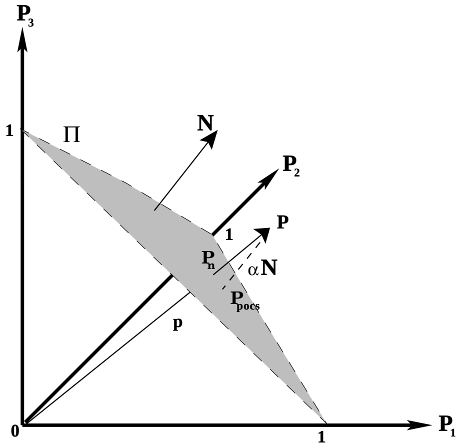


Fig. 1. An example of a three classes case with a posterior probability vector $\{P_1, P_2, P_3\}$. On the hyper-plane of interest Π, the points P_n and P_{pocs} result from the normalization of components after processing, and the filtered from POCS projection, respectively. **N** denotes the normal vector to the plane Π, **p** every filtered vector, and α a scalar factor.

Another alternative comes from projecting orthogonally the filtered class posteriors onto the constraint hyper-plane. This gives the closest class-posterior probabilities to the filtered ones that satisfy the constraint of summing up to one. To enforce that every component (every class-posterior probability) of the posterior probability vector being greater than or equal to zero, a new orthogonally projection must be done onto the hyper-plane restricted to the first hyper-quadrant. See figure 2 for the three classes case.

This technique is known as *POCS*, from *Projection Onto Convex Sets* [6]. POCS is a special case of the composite mapping algorithm that has been widely used in a variety of settings such as tomography and image restoration. The main idea of this technique consists of projecting the space of solutions onto a set attached to some *convex* constraints (the projection into the associated convex set is unique).

The projection of the filtered posterior vector onto the constraint hyper-plane can be easily obtained as follows:

$$\mathbf{P_{POCS}} = \mathbf{P} + \frac{1- < \mathbf{P}, \mathbf{N} >}{\|\mathbf{N}\|^2}\mathbf{N} \tag{3}$$

where **P** and **N** denote the posterior probability vector and the normal vector to the hyper-plane of interest, respectively.

4 The Smoothing Filter

The application of a smoothing operator is a well known technique of regularization, and as we mentioned before, this is a special case of the variational linear

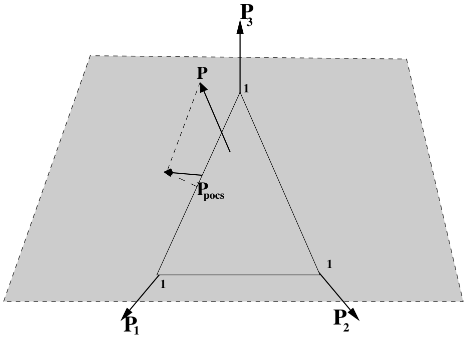


Fig. 2. Constraint of posterior probability greater than or equal to zero of the three classes case. The notation is the same than in figure 1. Note the second projection onto the line $P_3 + P_1 = 1$.

regularization approach. Traditionally, a PDE-based anisotropic diffusion filter is used in order to smooth the inside regions while preserving steep discontinuities across the borders. Here, we propose to use a non-iterative anisotropic filter which renders very good results.

A few words about *unsharp-masking*, a conventional enhance technique used by photographers to increase the importance of details in their pictures, will help us to introduce our anisotropic filter. Unsharp-masking is based on a decomposition of the image in a low pass component and a high-pass one. The high pass component is obtained from the difference between the image and the low pass component. If the weight of the high-pass is increased, the level of detail will be increased too. Then, we can consider that:

- The amount of detail could be adaptive at every point, weighting the high-pass component locally.
- The anisotropy of the structures should be taken in account to control in which orientations the high-pass component is important, and leaving the rest with a low pass component. This allows to smooth in the orientations of less structure (homogeneous regions, along edges, etc), preserving high structure elements.

Following these ideas, Knutsson et al. [4], introduced the local anisotropy information to process 2D grey-scale images. This was later generalized to N-dimensional images [3], developing an anisotropic N-dimensional adaptive filter, weighted by the local complexity of the dataset. Basically, this technique makes a linear combination of the output of a set of basis filters: one low pass (ideally a Wiener filter) and a set of high-pass filters in different orientations. The low pass filter fixes the scale of the dataset, and the high-pass ones the detail in each orientation. The contribution of each high basis component is related to the local dataset structure, and it is given by some coefficients. These coefficients should convey the importance of the information in every orientation with respect to the local structure, giving more weight to that orientation closest to the major direction of variation. Then, since in our approach the local complexity is coded by the local structure tensor $\mathbf{T}$,[2], to obtain the importance of every orientation, we have to estimate the projection of the eigenvectors of the local structure tensor at every point onto every vector orientation $\hat{\mathbf{n}}_k$, i.e., the inner product of the normalized local structure tensor $\mathbf{C}$ and a basis tensor $\{\mathbf{M}^k\}$ associated to $\hat{\mathbf{n}}_k$. The basis $\{\mathbf{M}^k\}$ is the *dual basis* of $\mathbf{N}_k = \hat{\mathbf{n}}_k \hat{\mathbf{n}}_k^T$, unique and obtained by the biorthogonality relation:

$$<\mathbf{M}^k, \mathbf{N}_l>= \delta_l^k \tag{4}$$

with δ_l^k the *Kronecker's delta.*

The minimum number of coefficients is the minimum number of high-pass filters, and it will be given by the relation between the data dimension N and the rank of this local structure tensor: $\frac{N(N+1)}{2}$.
The process is described by the equation 5, where $S_{lp}(\mathbf{x})$ and $S_{hp_k}(\mathbf{x})$, denote the outputs of the low-pass and high-pass basis filters, respectively.

$$S_{AAF}(\mathbf{x}) = S_{lp}(\mathbf{x}) + \sum_{k=1}^{N(N+1)/2} <\mathbf{C}, \mathbf{M}^k> S_{hp_k}(\mathbf{x}) \tag{5}$$

where $< \cdot, \cdot >$ denotes the inner product of tensors.

Figure 3 shows the flux diagram of our filtering approach is shown for a 2D grey-level image. The local structure is represented with an ellipsoid associated to the quadratic form of the structure tensor. The parameters used has been: $\rho_{l_p} = \frac{\pi}{2}$ and $\eta = 2$ (a Gaussian filter).

In the Fourier domain, the basis filter $\{Lp, Hp_1, .., Hp_{N(N+1)/2}\}$ used, is given by the expresion:

$$\begin{aligned} Lp(\boldsymbol{\nu}) &= \begin{cases} e^{-(\rho^\eta \cdot \Delta)} & if\ 0 \leq \rho \leq \rho_{l_p} \\ 0 & \text{otherwise} \end{cases} \\ Hp_k(\boldsymbol{\nu}) &= (1 - Lp(\boldsymbol{\nu})) \cdot (\hat{\boldsymbol{\nu}}^T \, \hat{\mathbf{n}}_k)^2 \end{aligned} \tag{6}$$

with $\boldsymbol{\nu}$ the N-dimensional frequency vector, $\rho = \|\boldsymbol{\nu}\|$, ρ_{l_p} the magnitude of the frequency where the Lp filter is $\frac{1}{2}$, and Δ and η filters shape parameters, related by: $\Delta = \frac{ln2}{\rho_{l_p}^\eta}$.

A comparison of our filtering approach with an anisotropic diffusion technique, in the context of 3D medical imaging, can be found in [10].

5 Results

In order to illustrate the proposed approach we show its results on a simple synthetic image with two vertical regions (see Fig. 4.a). Each region is generated from IID Gaussian distributions with identical typical deviations (30) and different mean values (10 on the left region and 70 on the right one). Figure 4.b shows the results achieved with the optimal MAP classifier considering homogenous priors (0.5 each). Figure 4.c shows the results achieved by applying the

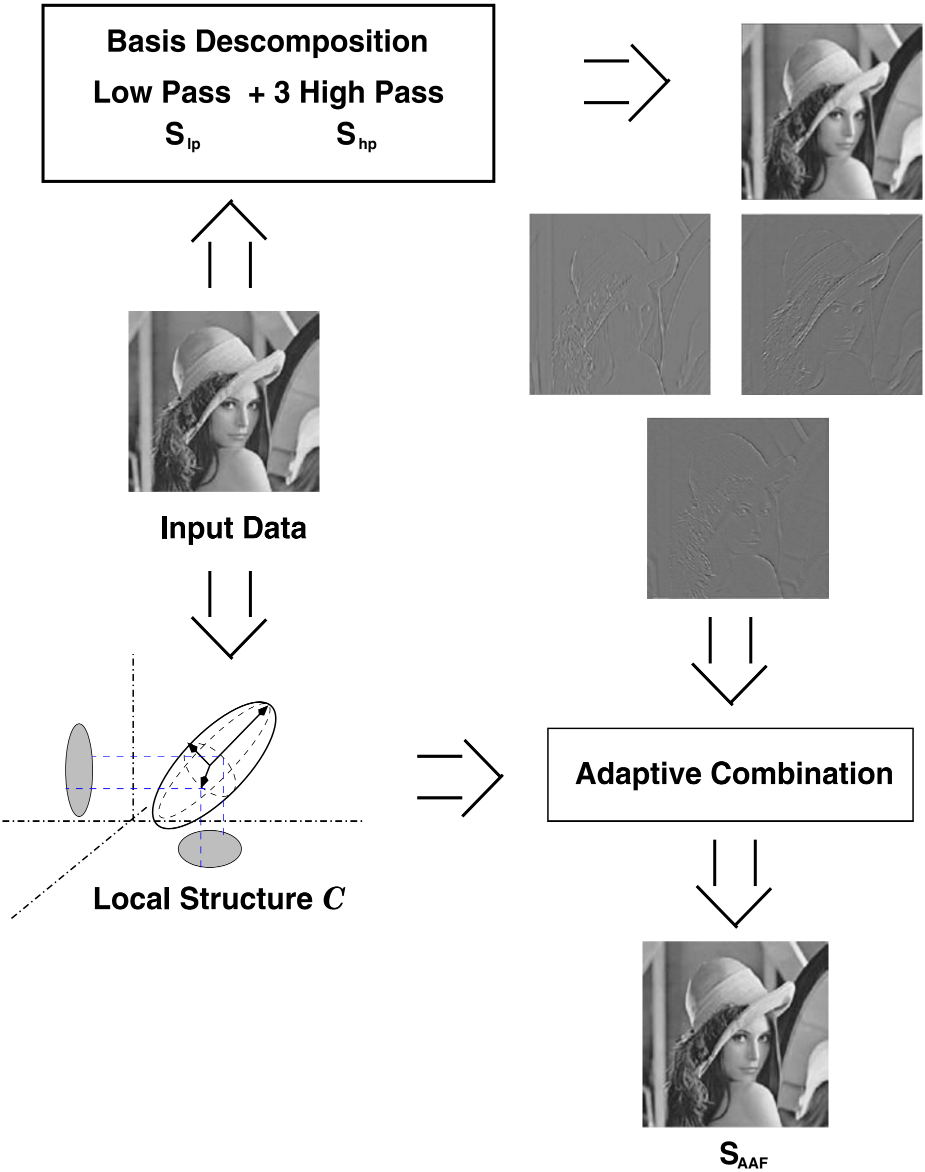
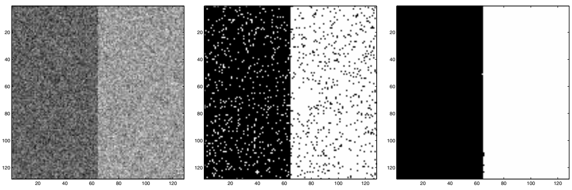


Fig. 3. Flux diagram of our filtering approach in a 2D scalar field.

MAP rule after the posteriors have been anisotropically smoothed and POCS-projected. Notice how inside each region the probability smoothing provides the correct classification and how the edge between both of them is almost optimally preserved.

To illustrate the performance of the method, in the case of real data, we apply it to the classification of the three main brain classes: gray matter (GM),



(a) (b) (c)

Fig. 4. a) Synthetic image with two regions **b**) MAP classification **c**) Regularized MAP classification

(a) (b) (c)

Fig. 5. a) Zoomed MRI brain image **b**) MAP classification **c**) Regularized MAP. Black: Background, Dark Gray: CSF, Gray: GM, and White: WM.

white matter (WM) and cerebrospinal fluid (CSF), from a 3D ($256 \times 256 \times 160$) MRI data. The brain is first extracted using a skull stripping automatic algorithm (thresholding followed by a light erosion, a hard opening, and region growing in order to obtain a brain mask that is applied to the original MRI data). Figure 5.a shows a zoomed area of a slice from the original MRI volume, Fig. 5.b shows the MAP classification (the parameters are provided from a ground-truth segmentation) and Fig. 5.c shows the regularized MAP segmentation using the same probabilistic characterization. The labels for the segmentation are white for WM, gray for GM, dark gray for CSF, and black for the background. Notice how the regularized MAP provides much more spatially coherent classifications while preserving the borders.

6 Discussion and Future Directions

In this paper we have proposed an approach to improve conventional independent voxel-wise MAP classifications. The approach is non-iterative and fast, and produces piecewise segmentations while preserving the class-borders. Our preliminary results are encouraging,though much more validation is still needed to introduce it in routine clinical segmentations. Extensions of the approach include its use in arbitrary vector fields with differents constraints. In fact, recently, we have developed a new approach to regularize tensor fields [11], using a variation of this technique.

Acknowledgments. The first author is funded by a FPU grant at the University of Las Palmas de Gran Canaria. This work is partially funded by the research project TIC2001-3808, by the Castilla y León Goverment project joint to the research grant VA91/01, and the US grants NIH P41-RR13218, and CIMIT.

References

1. Hadamard J. Lectures on the Cauchy Problem in Linear Partial Differential Equations. Yale University Press, 1923.
2. Knutsson, H. Representing local structure using tensors. In 6th Scandinavian Conference on Image Analysis. Oulu, Finland, pages 244–251, 1989.
3. Knutsson H., Haglund L., Bårman H., Granlund G. H. A framework for anisotropic adaptive filtering and analysis of image sequences and volumes. In Proceedings ICASSP-92, San Fransisco, CA, USA, March 1992. IEEE.
4. Knutsson, H., Wilson, R. and Granlund, G.H. Anisotropic non-stationary image estimation and its applications-part I: Restoration of noisy images. IEEE Trans. on Communications. COM-31, 3:388–397, 1983.
5. Marroquin J.L. Probabilistic Solution of Inverse Problems. PhD thesis, M.I.T., Boston (MA) – USA, 1985.
6. Moon T.K. and Stirling W.C. Mathematical Methods and Algorithms for Signal Processing. Prentice-Hall, 2000.
7. Pardo A. and Sapiro G. Vector probability diffusion. IEEE Signal Processing Letters, 8(4):106–109, 2001.
8. Perona P., Malik J. Scale-Space and edge detection using anisotropic diffusion. IEEE Trans. on Pattern Analysis and Machine Intel., 12(7):629–639, July 1990.
9. Poggio T., Torre V. and Koch C. Computational Vision and Regularization Theory. Nature, 317:314–319, 1985.
10. Rodriguez-Florido M.A., Krissian K., Ruiz-Alzola J., Westin C.-F. Comparison between two restoration techniques in the context of 3d medical imaging. Lecture Notes in Computer Science - Springer-Verlag Berlin Heidelberg, 2208:1031–1039, 2001.
11. Rodriguez-Florido M.A., Westin C.-F., Casta no C. and Ruiz-Alzola J. DT-MRI Regularization Using Anisotropic Tensor Field Filtering. submitted to MICCAI03, 2003.
12. Teo P.C., Sapiro G.,Wandell B.A. Anisotropic smoothing of posterior probabilities. In 1997 IEEE International Conference on Image Processing – October 26–29 – Washintong, D.C., pages 675–678, 1997.
13. Teo P.C., Sapiro G., Wandell B.A. A method of creating connected representations of cortical gray matter for functional mri visualization. IEEE Trans. Medical Imaging, 16:852–863, 1997.
14. Tikhonov A., Arsenin V. Solutions of ill-posed problems. Wiley, New-York, 1977.
15. Weickert, J. Anisotropic Diffusion in image processing. Teubner-Verlag, Stuttgart, 1998.