



**HAL**  
open science

# Simultaneous Image Segmentation and Medial Structure Estimation - Application to 2D and 3D Vessel Tree Extraction

Sherif Makram-Ebeid, Jean Stawiaski, Guillaume Pizaine

► **To cite this version:**

Sherif Makram-Ebeid, Jean Stawiaski, Guillaume Pizaine. Simultaneous Image Segmentation and Medial Structure Estimation - Application to 2D and 3D Vessel Tree Extraction. SPIE Medical Imaging 2011: Image Processing, Feb 2011, Orlando, FL, United States. pp.79623D, 10.1117/12.878126 . hal-00594714

**HAL Id: hal-00594714**

**<https://hal.science/hal-00594714v1>**

Submitted on 20 May 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Simultaneous Image Segmentation and Medial Structure Estimation - Application to 2D and 3D Vessel Tree Extraction

Sherif Makram-Ebeid <sup>a</sup>, Jean Stawiaski <sup>a</sup> and Guillaume Pizaine<sup>a,b</sup>

<sup>a</sup>Philips Healthcare Research Lab, 33 rue de Verdun, F92156 Suresnes, France;

<sup>b</sup>Institut Telecom, Telecom ParisTech, CNRS LTCI, F75013 Paris, France

## ABSTRACT

We propose a variational approach which combines automatic segmentation and medial structure extraction in a single computationally efficient algorithm. In this paper, we apply our approach to the analysis of vessels in 2D X-ray angiography and 3D X-ray rotational angiography of the brain. Other variational methods proposed in the literature encode the medial structure of vessel trees as a skeleton with associated vessel radii. In contrast, our method provides a dense smooth level set map which sign provides the segmentation. The ridges of this map define the segmented regions skeleton. The differential structure of the smooth map (in particular the Hessian) allows the discrimination between tubular and other structures. In 3D, both circular and non-circular tubular cross-sections and tubular branching can be handled conveniently. This algorithm allows accurate segmentation of complex vessel structures. It also provides key tools for extracting anatomically labeled vessel tree graphs and for dealing with challenging issues like kissing vessel discrimination and separation of entangled 3D vessel trees.

**Keywords:** Variational Techniques, Finite Elements, Medial Structure, Vessel Tree Segmentation

## 1. INTRODUCTION AND SUMMARY

A variational formulation of bi-phasic medical image segmentation problems is proposed. The segmentation is obtained in the form of a signed level set map which is positive inside the object region to extract (foreground) and negative in the remaining image domain (background). In the proposed mathematical setting, one minimizes an objective functional which includes a data fidelity term together with a level-set smoothness term. The fidelity energy term penalizes non-zero level set values at Canny edge locations as well as discrepancies between the level set gradient unit norm and the Canny edge unit normal vector. The smoothness term is a non-negative semi-norm which penalizes spatial changes in the level set gradient. In general, the resulting smooth level set map involves ridges, valleys as well as local extrema and saddle points. These serve to characterize the medial structure of the segmented region. In particular, the ridge crest locus<sup>1</sup> may define the skeleton of the region. The corresponding Hessian eigenvalues signatures provide the class of regions involved (e.g. in the 3D case blob-like, tube like or slab-like etc.). In this paper, our approach is applied to the automatic segmentation and extraction of vessel tree structures in 2D X-ray angiography images and in 3D X-ray rotational brain vessel images.

## 2. METHODS

### 2.1 Extraction of the medial structure

In essence, the proposed algorithm starts with a gray level image as illustrated by the circular ring in the toy example of Fig. 1.a. The Canny edges<sup>2</sup> are extracted keeping only the gradients which are maximal along their direction (See Fig. 1.b). Both the gradient moduli  $|\nabla I|$  and the gradient orientation unit vectors  $\underline{u} = \nabla I / |\nabla I|$  are recorded for later use. One then seeks a level set function  $\varphi(\underline{x})$  of the position  $\underline{x}$  in the image which minimizes the energy

$$E(\varphi) = \sum_{\underline{x}} \rho(\underline{x}) \left( \varphi^2 + \sigma^2 \|\nabla\varphi - \underline{u}\|^2 \right) + \lambda \cdot \text{Reg}(\varphi) \quad (1)$$

---

Further author information: send correspondence to S. Makram-Ebeid: E-mail: sherif.makram-ebeid@philips.com, Tel.: +33 1 47 28 36 65

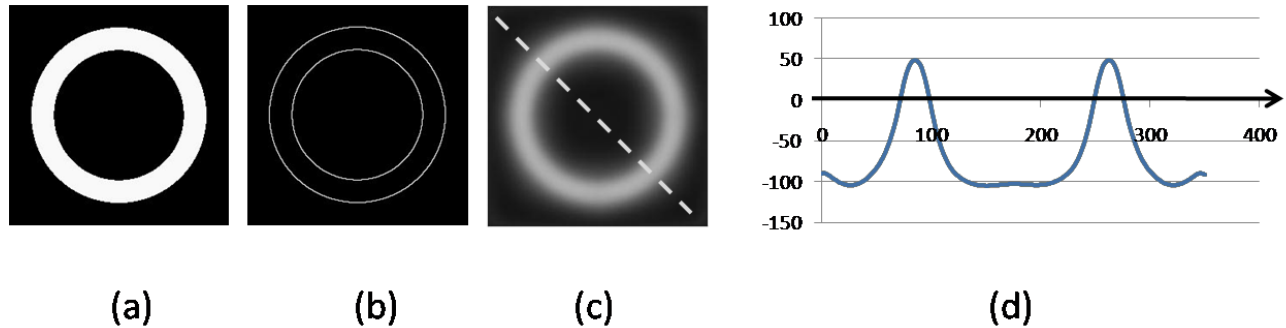


Figure 1. 2D synthetic image illustration: (a) ring data image, (b)  $\rho = |\nabla I|$  of Canny Edges, (c)  $\varphi(\underline{x})$  level-set map, (d)  $\varphi(\underline{x})$  profile versus distance along the dotted line. Note the parabolic shape of the profile for positive  $\varphi$  values  $\varphi(\underline{x}) > 0$

where  $\sigma$  is a scaling factor (e.g.  $\sigma = 1$ );  $\rho(\underline{x})$  is a data fidelity edgeness-weight, which is non-zero only at the positions  $\underline{x}$  of the Canny edges and is a function of the modulus  $|\nabla I|$  of the gradient of data image  $I(\underline{x})$ . In practice, we focus on compact objects and wish to extract and characterize their medial structures, so Canny edges are further paired in such a way as to reduce the influence of edges which are incompatible with the size range or with the contrast of the objects one wishes to segment (see Sect. 2.2).

The role of the first summation term in Eq.1 is clearly to cause  $\varphi$  to approach 0 near Canny edges with its gradient vector  $\nabla\varphi$  approaching the Canny edge unit vector  $\underline{u}$ . The second term of the above energy expression  $Reg(\varphi)$  is a functional which imposes smoothness of  $\varphi$  and  $\lambda$  is a positive constant for tuning the relative roles of fidelity (first term) and regularity (second term).  $Reg(\varphi)$  penalizes partial derivatives of  $\varphi$  of order higher than a given order  $p$ . The use of such smoothness constraints has been advocated in a seminal book by Blake and Zisserman<sup>3</sup> for image segmentation applications. Their thin membrane model makes use of  $Reg(\varphi) = \int_{\Omega} \|\nabla\varphi\|^2$  where  $\Omega$  stands for the image domain; this penalizes partial derivatives of order higher than  $p = 0$  i.e. any non-constant  $\varphi$ . In their thin plate model,  $Reg(\varphi) = \int_{\Omega} \|\nabla\nabla\varphi\|^2$  where  $\|\nabla\nabla\varphi\|^2$  is the Frobenius norm of the Hessian of  $\varphi$  thus penalizing partial derivatives of order higher than  $p = 1$  i.e. any non-affine  $\varphi$ .

In this article, we propose a Partition of Unity Finite Element (PUFEM) multi-resolution pyramid<sup>4</sup> to represent the level set  $\varphi$  as a windowed combination of first or second order polynomials, and to make use of a native regularity semi-norm for  $Reg(\varphi)$ . This regularity term has a long range smoothing effect similar to that of the thin-plate energy as explained in Appendix D. The PUFEM representation has been proposed and deeply studied by Babuška and Melenk<sup>5,6</sup> for solving diverse partial differential equations. In addition to computational efficiency, this approach does not have difficulties with the boundary of the image domain  $\Omega$  as is the case with finite difference approaches which imply the use of natural boundary conditions which can be rather tricky when dealing with an elliptic PDE.<sup>7</sup> Furthermore, the non-zero Canny-edge weights  $\rho(\underline{x})$  appearing in Eq.1 do not represent a continuous function. Again, this can be very conveniently handled within our approach. More technical details are outlined in Appendices A to D.

On Fig. 1, the energy  $Reg(\varphi)$  penalizes all partial derivatives higher than  $p = 2$ ; i.e. it could identically vanish only if  $\varphi$  was a multivariate polynomial of order equal or less than 2. The computed  $\varphi$  map is as shown in Fig. 1.(c). The profile of  $\varphi$  taken along the dotted line of Fig. 1.(c) is depicted in Fig. 1.(d). This profile shows an almost perfect parabolic variation inside the ring region (in which  $\varphi(\underline{x}) > 0$ ). The maxima of  $\varphi(\underline{x})$  along the ring radii occur on the mid circle of the ring (which is also the ridge locus in this case). The zero crossings  $\varphi(\underline{x}) = 0$  coincide with the Canny edges and the corresponding level-set gradient modulus  $\|\nabla\varphi\|$  is near unity (except for a scaling factor in Fig. 1.(d)). A similar parabolic behavior is more generally observed in real situations for vessels with slowly varying cross-sections in 2D and 3D (as well as for uniform width slabs in 3D).

## 2.2 Edge-pair scoring and computation of edgeness weights

In this article, we focus on compact objects and wish to extract and characterize their medial structures. To do so, we start by extracting Canny Edges. It is essential to eliminate edges that do not belong to a compact

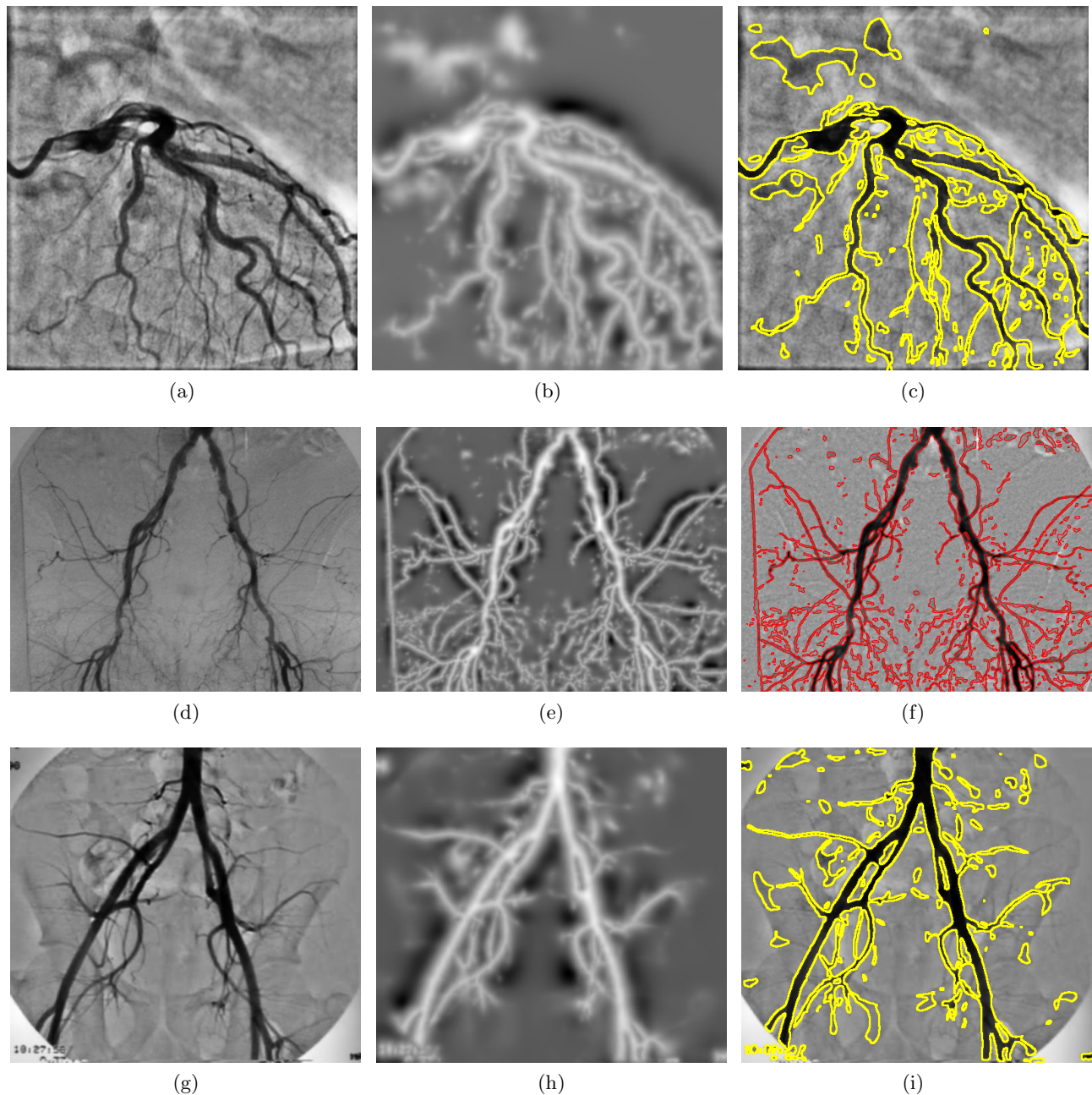


Figure 2. Examples with 2D X-ray images showing the raw image (first column), the resulting medialness map (second column) and the zero-level of the medialness map as yellow contours. (a-c) Coronary Angiography (image size  $492 \times 512$ ), (d-f) Iliac Angiogram (image size  $1360 \times 1025$ ), (g-i) Iliac Angiogram (image size  $528 \times 474$ ).

object. We therefore attribute a score to an edge which depends on how well it can be paired with another edge so that the edge pair represent respectively the entry and exit from the compact object. We also must ensure the segmentation of objects with given contrast signature (*i.e.* either bright relative to darker background or else dark relative to a brighter background). To do so, we systematically scan the whole image on a set of dense scanlines parallel to the main image axes as well as the main diagonals. Each pixel or voxel simultaneously belongs to all those scanlines. There are 4 such sets in 2D and 13 in 3D. For each scanline, the Canny edge gradients are projected in the scan-direction, resulting in a sparse profile of projected gradient values. Along

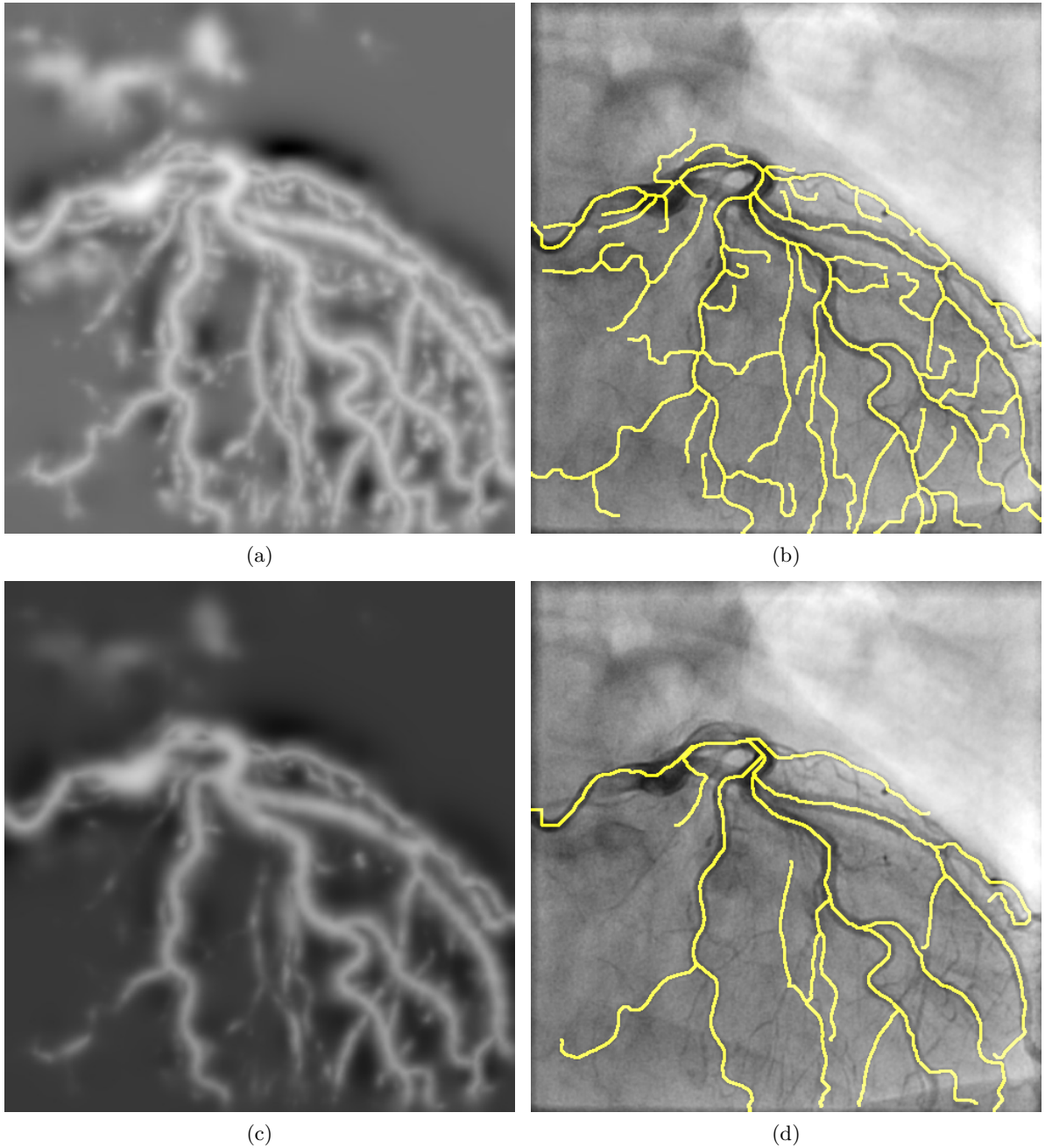


Figure 3. Effect of changing the smoothing weight on Coronary Image for the medialness and the centerline. (a) Medialness with  $\lambda = 20$ , (b) Corresponding centerlines, (c) Medialness with  $\lambda = 150$ , (d) Corresponding centerlines.

each such line we pair the projected gradients (+/- or -/+) in a manner which is compatible with the object contrast we are looking for. In doing so, we only keep pairs with ratios of projected gradient magnitudes which are within prescribed bounds (typically between 1/4 and 4) and separated by less than a predefined maximal distance. Projected gradients of smaller magnitudes must be eliminated if they occur between any two paired gradients. Moreover, a projected gradient may be paired with no more than another one. Each member of a pair is attributed a pairing score equal to the minimum of their projected gradient magnitudes. This process yields a profile of non-negative pairing-scores for each scanline.



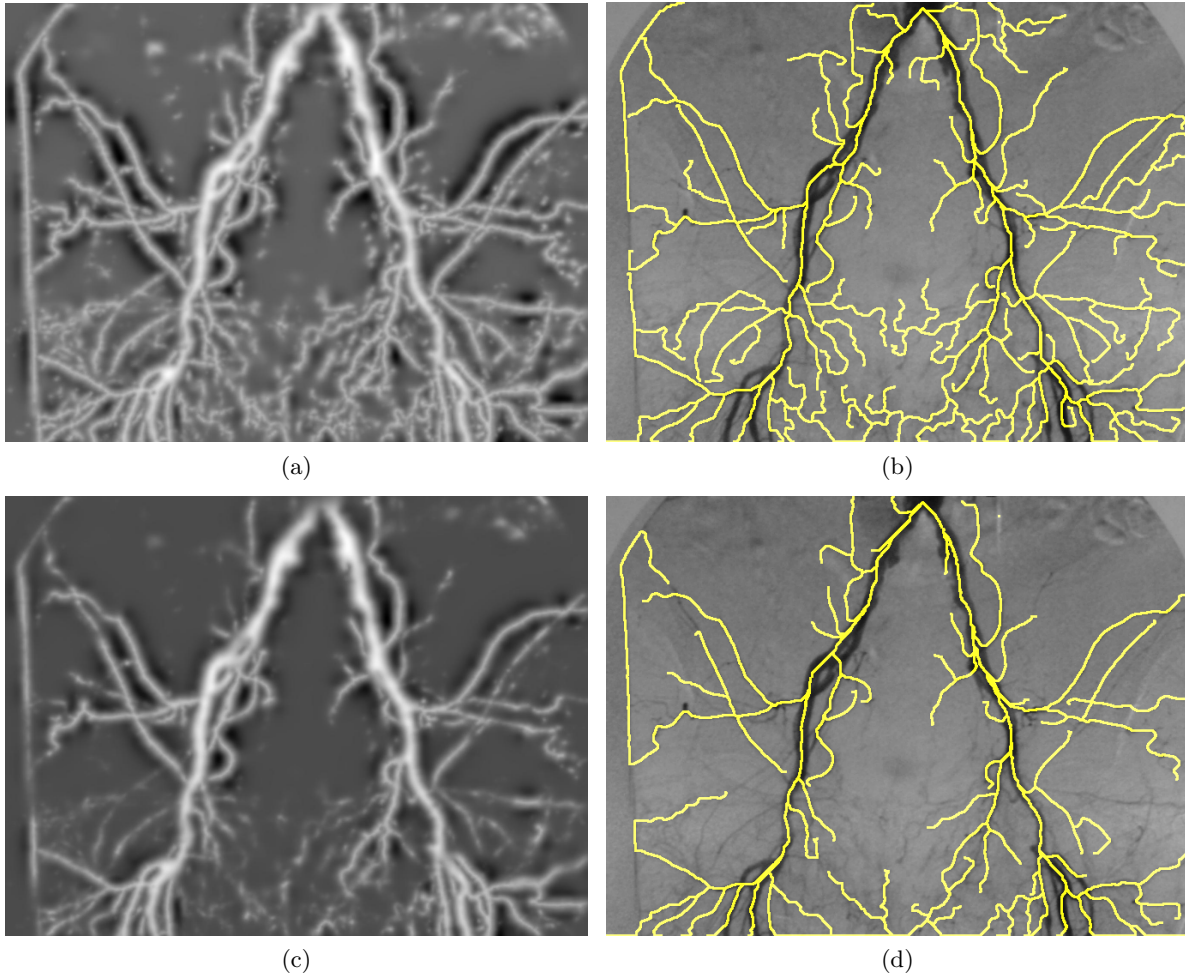


Figure 4. Effect of changing the smoothing weight on Iliac Image (1) for the medialness and the centerline. (a) Medialness with  $\lambda = 20$ , (b) Corresponding centerlines, (c) Medialness with  $\lambda = 150$ , (d) Corresponding centerlines.

An edge-score map is then constituted by taking the maximum of the edge-pairing scores of the different concurrent scanline profiles. This procedure bears some similarity with the maximal gradient flux procedure proposed by Lesage *et al.*<sup>8</sup> Both their approach and ours make use of a pairing procedure to establish a score, the chief difference is that we attribute the resulting score to an edgeness map whereas Lesage *et al.* attribute the score to candidate circular contours of vessel cross-sections. In order to reduce the effect of noise, we found it useful to define the edgeness-weights map  $\rho(\underline{x})$  as a power  $\nu$  of edge-score which may be adjusted according to the application between  $\nu = 1$  to  $\nu = 3$ .

### 3. RESULTS

Figure 2 shows results obtained for 2D X-ray angiography images with  $p = 2$ ,  $\nu = 1$  and  $h = 4$  pixel-units at the finest level of our FEM pyramid. The regularization parameter  $\lambda$  was set at  $\lambda = 20$  to provide a moderate smoothing action. The roles of these parameters are as explained in Appendices REF. Subfigures (a) show the original raw images, (b) the resulting signed medialness maps (zero level is the mid-gray tone) and, in (c), the zero-levels of medialness are displayed as color contours. As discussed in Sect. 2.2, the Canny edges are paired in such a way that ribbon-like structures with wrong contrast (bright) are discarded. Edge pairs which are manifestly not part of a compact object are excluded. In this way, the great majority of vessel sections with a

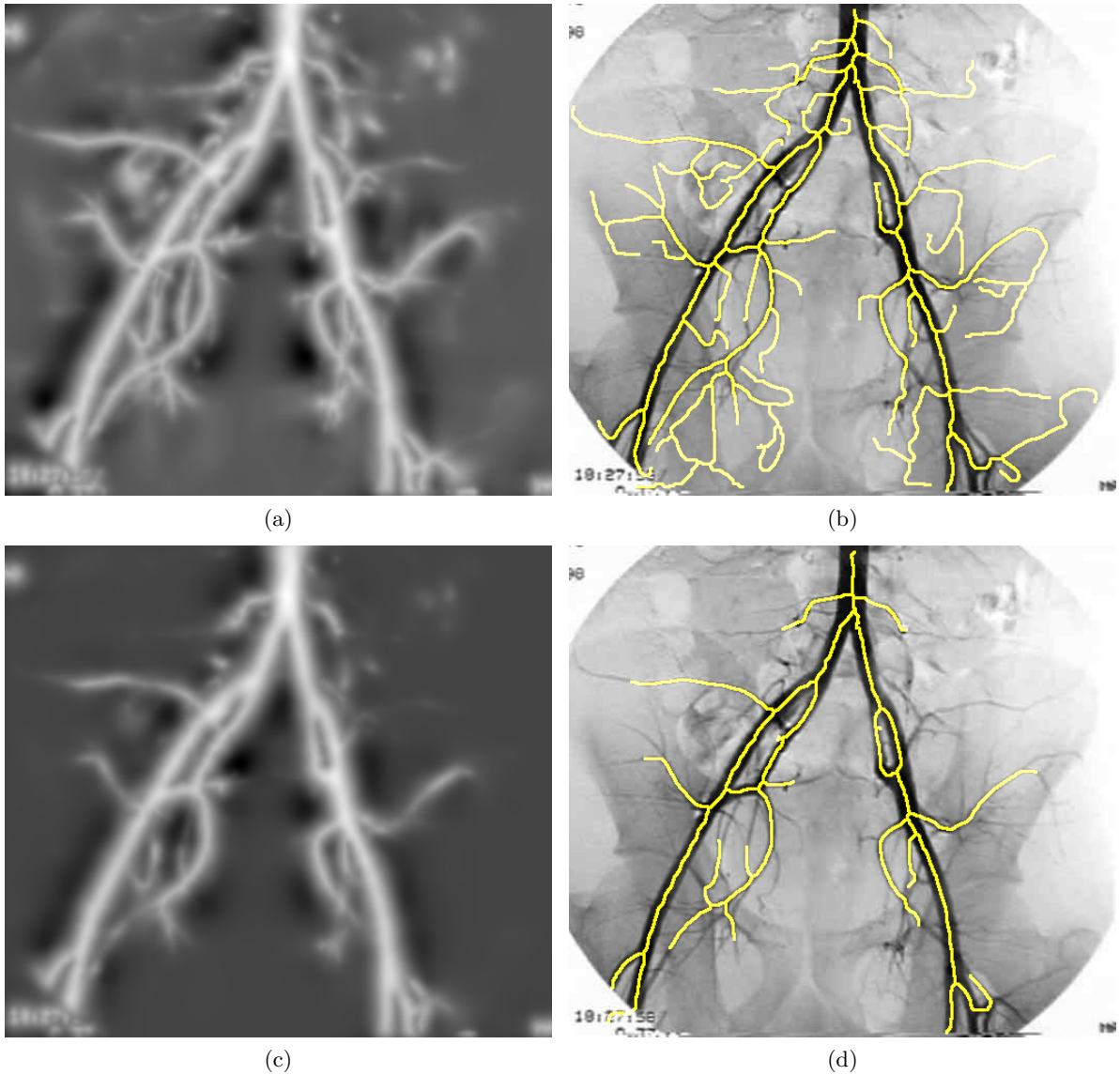


Figure 5. Effect of changing the smoothing weight on Iliac Image (2) for the medialness and the centerline. (a) Medialness with  $\lambda = 20$ , (b) Corresponding centerlines, (c) Medialness with  $\lambda = 150$ , (d) Corresponding centerlines.

not too faint dark contrast and a width larger than one pixel and less than 100 pixels are correctly and accurately delineated.

Figures(3,4,5) show the effect of altering parameters that may have an effect on the degree of smoothing and on the suppression of weakly contrasted details. The upper subfigures (a) and (b) are obtained with  $\lambda = 20, \nu = 1$  as before but the lower subfigures (c) and (d) are obtained with a much larger regularization weight  $\lambda = 150$  and with an edginess exponent  $\nu = 3$  meant to reduce the influence of the edges having a faint contrast. The left column subfigures (a) and (c) are medialness maps whereas the right column subfigures show ridge paths super-imposed on the original images. It can be seen that the effect of enhanced smoothing is to considerably reduce the effect of faintly contrasted vessels together with the suppression of background noise structures. The effect is also to regularize the medialness map appearance and to produce smoother zero-level contours which, however, remain very close to the image edges. The corresponding ridge paths have been tracked using a tree extracting procedure that avoids closed path cycles (this is the main reason why some vessel centerlines are

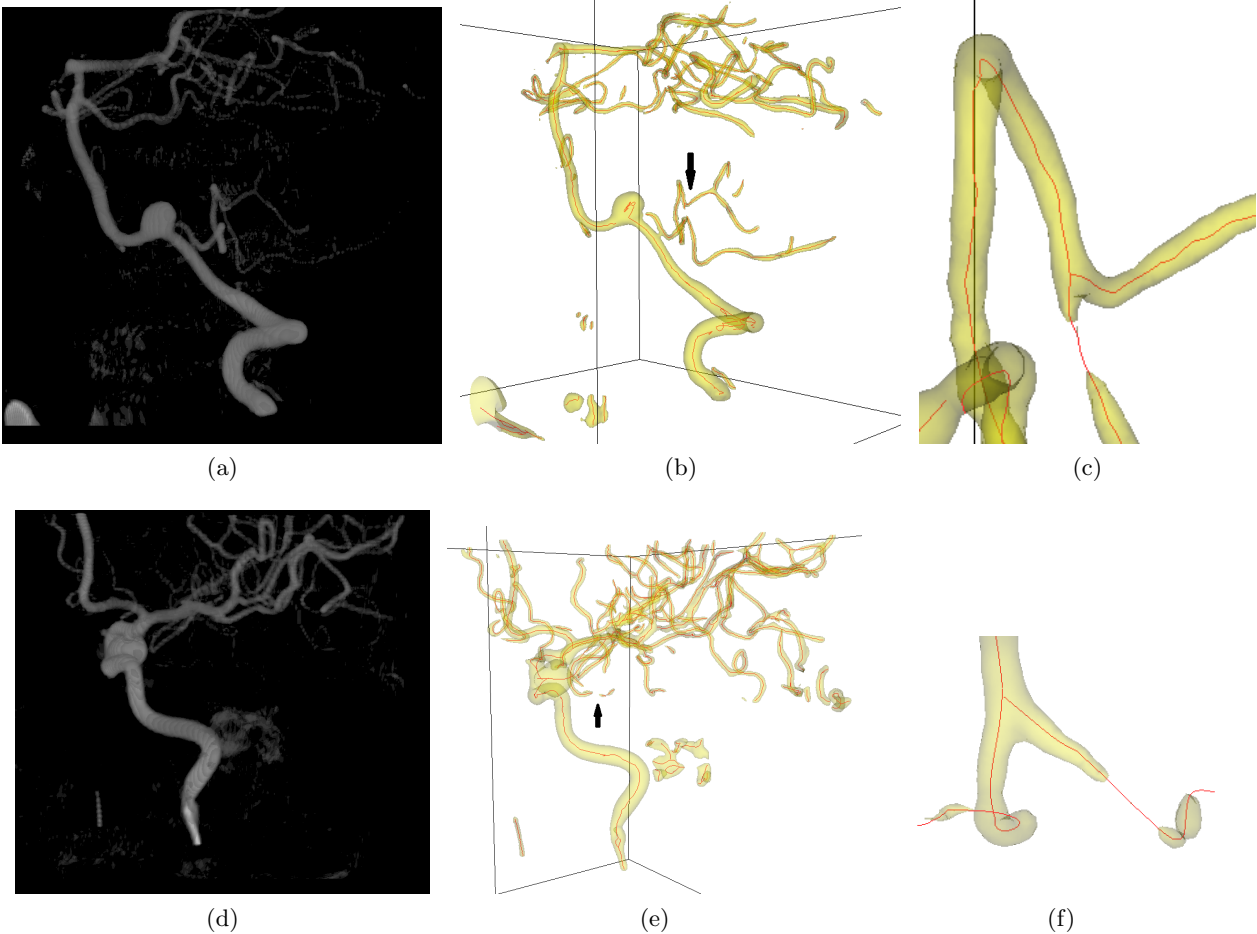


Figure 6. Two examples with 3D X-ray Rotational Angiographies. The raw data ((a) and (d)) has been segmented with the methods described in Sect. 2. (b) and (e) show the zero-level set of the medialness as a 2D surface and the centerlines extracted according to Sect.3. The centerlines extraction can be extended beyond the segmentation boundaries to recover non-connected branches, as shown in (c) and (f).

partly missed in the shown paths).

Unlike approaches based on a multi-scale Hessian analysis, our approach is insensitive to the effect of isolated edges (*e.g.* the diaphragm in Fig. 3.(a) and the image intensifier borders in Fig.5.(a) clearly do not contribute to the medialness map). The medialness ridge paths can be seen to follow closely the vessel centerlines. Work is ongoing to find the best strategy for vessel-branch tracking and labeling in 2D (which is harder than in 3D because of the occurrence of vessel projection overlaps). It is also clear that, even with a perfect medialness extractor, a fine resolution path will track spurious and noise ridges thus necessitating suitable post-processing path-cleaning procedure. This could be done by making use of features related to medialness height fluctuations or to variations of the input image contrast along a path in order to classify candidate paths as true-vessels or related to structural or random noise.

Fig. 6 shows results obtained for 3D X-ray rotational angiography of brain vessels (image size 256x256x256) with  $p = 2$ ,  $\nu = 2$  and  $h = 4$  pixel-units at the finest level of our FEM pyramid. The regularization parameter  $\lambda$  was set at  $\lambda = 10$  to provide a moderate smoothing action. The left image in Fig. 6 shows a rendering of the zero-level of the level set obtained for the segmentation of vessels. The raw data are shown in the viewer in the right image. The corresponding intersection of this zero-level with the orthogonal viewing planes is shown as colored contours. An aneurysm with its connected vessels is extracted in this way (smallest vessel size segmented



is 2-voxels wide). Here again, the medialness map is computed and the corresponding candidate skeleton points extracted and tracked.

In 3D, we define a vessel-ridge or skeleton point  $\underline{x}$  as a point where the medialness  $\varphi$  has a strong local maximum in a planar cross-section with unit normal vector  $\underline{u}_n$ ; let  $(\eta_p < 0)$  be the maximum oriented second derivative of  $\varphi$  in this plane and  $\eta_n$  the second derivative in the  $\underline{u}_n$  direction. The candidate skeleton point is retained provided  $\eta_n$  is greater (less negative) than  $\eta_p$ . It is sufficient to do the selection by examining the discrete local  $3 \times 3 \times 3$  neighborhood of any point  $\underline{x}$ . There are 13 possible normal directions  $\underline{u}_n$  to examine. Each such candidate skeleton point may then be further characterized by the corresponding Hessian tensor. Computation times for the whole procedure are quite reasonable (when sampling down to 128x128x128 resolution, about 15 second on a single core PC processor, 120 seconds for the full 256x256x256 volumes - with a non optimized implementation).

## 4. CONCLUSIONS

We proposed an approach for combining segmentation and medial structure analysis of medical images in a single computationally efficient algorithm. In this paper, we presented applications to vessel tree analysis in 2D and 3D images. Accurate contouring was demonstrated for a wide range of vessel sizes (including connected objects such as aneurysms). The resulting medialness maps provide dense and accurate information on the segmented structures and in particular the medialness ridge provides accurate information on object shape, size and orientation. Work is ongoing to exploit those results for the extraction of semantically labeled trees and for disentangling 2D and 3D vessel trees.

### APPENDIX A. FINITE ELEMENT REPRESENTATION OF LEVEL SET MAP $\varphi$

Let  $\Omega \subset R^d$  be a rectangular image domain of dimension  $d = 2$  or  $3$  over which a set of nodes is regularly distributed. To any node  $m$ , we associate a compact sub-domain  $\Omega_m \subset R^d$ , and a non-negative window function  $\psi^{(m)}(\underline{x})$  which vanishes outside  $\Omega_m$ . A *Partition of Unity* of the domain  $\Omega$  means that:

$$\Omega \subset \bigcup_m \Omega_m \quad \text{and} \quad \forall \underline{x} \in \Omega, \quad \sum_m \psi^{(m)}(\underline{x}) = 1 \quad (2)$$

The level  $\varphi(\underline{x})$  map is expressed as a blending of local polynomials in the form:

$$\varphi(\underline{x}) = \sum_m \psi^{(m)}(\underline{x}) \theta^{(m)}(\underline{x}) \quad (3)$$

where each local polynomial  $\theta^{(m)}(\underline{x})$  is expanded on a local basis  $\{v_r^{(m)}\}$  as:

$$\theta^{(m)}(\underline{x}) = \sum_r a_r^{(m)} v_r^{(m)}(\underline{x}) \quad (4)$$

where the  $a_r^{(m)}$  are the scalar coefficients of the expansion in terms of local basis polynomials  $v_r^{(m)}(\underline{x})$  of degree less or equal than a preset degree  $p$ . For any point  $\underline{x} \in \Omega$ , there are at most  $M$  window functions for which  $\psi^{(m)}(\underline{x}) \neq 0$ . For the sake of computational efficiency, we choose nodes distributed over a regular rectangular array with inter-node spacing  $h$  along each of the coordinate axes ( $i = 1, \dots, d$ ). In our implementation, we have  $M = 2^d$ . Within a window  $m$ , the coordinates  $x_i$  of a point  $\underline{x}$  are replaced by normalized local coordinates  $\zeta_i^{(m)} = (x_i - \xi_i^{(m)})/h$  for which the window center coordinates  $\xi_i^{(m)}$  are taken as origin. Each window function  $\psi^{(m)}(\underline{x})$  is expressed as a product:

$$\psi^{(m)}(\underline{x}) = \prod_{i=1}^d P\left(\left|\zeta_i^{(m)}\right|\right) \quad (5)$$

The function  $P(\zeta)$  for  $\zeta \in [0, 1]$  is taken as a polynomial of degree  $2p+1$  with vanishing derivatives of order 1 to  $p$  at  $\zeta = 0$  and  $\zeta = 1$  and such that  $P(0) = 1$  and  $P(1) = 0$ . It can be easily seen that, for any non negative integer

$p$ , there exists a unique such polynomial and that  $P(\zeta) + P(1 - \zeta) = 1$ . Defining the window functions from Eq.5 is advantageous because they are simple to compute and they automatically satisfy the normalization condition  $\sum_m \psi^{(m)}(\underline{x}) = 1$  with  $M = 2^d$  non-vanishing windows for any  $\underline{x}$ . Moreover, the window function  $\psi^{(m)}(\underline{\xi}^{(m)})$  of Eq.5 has useful flatness properties which derive from the choice of the factors  $P(\zeta)$ . At the center  $\underline{x} = \underline{\xi}^{(m)}$  of the window,  $\psi^{(m)}(\underline{\xi}^{(m)}) = 1$  and all its multivariate partial derivatives of order equal or less than  $p$  are equal to zero. The basis polynomials  $v_r^{(m)}(\underline{x})$  associated with node  $m$  are monomials of all degrees up to a maximum  $p$ , for example, for  $p = 2$  and  $d = 2$ , there are six such monomials ( $r = 1$  to 6) defined by  $\{1, \zeta_1^{(m)}, \zeta_2^{(m)}, (\zeta_1^{(m)})^2, \zeta_1^{(m)}\zeta_2^{(m)}, (\zeta_2^{(m)})^2\}$ . In general, there are  $N = (p + d)!/(p!d!)$  such monomials of degree less or equal than  $p$  for domain dimension  $d$ . Thus each window node  $m$  is associated with a local polynomial  $\theta^{(m)}(\underline{x})$  defined by the  $N$ -dimensional coefficient vector  $\{a_r^{(m)} | r \in [0, N - 1]\}$ . By convention,  $v_r^{(m)}(\underline{x})$  and  $a_r^{(m)}$  are ordered in non descending degree starting with the coefficient  $a_0^{(m)}$  of the zero-degree monomial  $v_0^{(m)} \equiv 1$  for  $r = 0$ .

## APPENDIX B. LOCAL POLYNOMIAL SOLUTION WITHIN A NODAL WINDOW

Making use of PUFEM strategy,<sup>4-6</sup> we replace the data attachment part  $\sum_{\underline{x}} \rho \left( \varphi^2 + \sigma^2 \|\nabla\varphi - \underline{u}\|^2 \right)$  of the energy in Eq.1 by an upper bound in the form  $\sum_m F^{(m)}$  where  $F^{(m)}$  is the windowed data attachment quadratic energy in node  $m$

$$F^{(m)}(\theta^{(m)}) = \sum_{\underline{x}} \rho \psi^{(m)} \left( (\theta^{(m)})^2 + \sigma^2 \|\nabla\theta^{(m)} - \underline{u}\|^2 \right). \quad (6)$$

By substituting the expression of Eq.4 for  $\theta^{(m)}$  in the above equation, we get  $F^{(m)}$  as a function of the coefficients  $a_r^{(m)}$

$$F^{(m)}(a_r^{(m)}) = \sum_{r,s} M_{r,s}^{(m)} a_r^{(m)} a_s^{(m)} - 2 \sum_r G_r^{(m)} a_r^{(m)} \quad (7)$$

where

$$M_{r,s}^{(m)} = \sum_{\underline{x}} \rho \psi^{(m)} \left( v_r^{(m)} \cdot v_s^{(m)} + \left( \frac{\sigma}{h} \right)^2 \sum_i \left( \partial_i v_r^{(m)} \cdot \partial_i v_s^{(m)} \right) \right) \quad (8)$$

$$G_r^{(m)} = \sum_{\underline{x}} \rho \psi^{(m)} \left( \frac{\sigma^2}{h} \sum_i \left( \partial_i v_r^{(m)} \cdot u_i \right) \right) \quad (9)$$

where it is recalled that  $h$  is the inter-node distance,  $i$  is the index of coordinate  $x_i$  and  $\partial_i$  stands for the partial derivative operator  $\frac{\partial}{\partial x_i}$  and  $u_i$  the  $i^{th}$  component of the Canny Edge unit normal vector. The nodal data attachment energy  $F^{(m)}$  can now be rewritten in matrix notation

$$F^{(m)}(X^{(m)}) = \left( X^{(m)} \right)^T M^{(m)} X^{(m)} - 2 \left( G^{(m)} \right)^T X^{(m)} \quad (10)$$

where  $X^{(m)}$  is the column vector of coefficients  $\{a_r^{(m)} | r \in [0, N - 1]\}$  while the  $N \times N$  matrix  $M^{(m)}$  and the  $N$  vector  $G^{(m)}$  have components defined above in Eqs.8 and 9 respectively. A local solution of the variational problem can now be determined by minimizing the above quadratic form of Eq.10 i.e. by solving the the linear system  $M^{(m)} X^{(m)} = G^{(m)}$ . In practice, we add a very small Tikhonov term (typically  $10^{-5} tr(M^{(m)})$ ) to the diagonal entries of matrix  $M^{(m)}$  which will ensure that the resulting symmetric matrix is positive definite (hence non-singular). This will always work unless  $\rho$  is identically zero within the support of window  $m$ . Let the solution for the local window  $m$  be  $Y^{(m)}$ . The local energy  $F^{(m)}$  can now be rewritten by completing the quadratic form so that it will facilitate its integration in the global energy minimization scheme involving regularization as discussed in the next section.

$$F^{(m)}(X^{(m)}) = \left( X^{(m)} - Y^{(m)} \right)^T M^{(m)} \left( X^{(m)} - Y^{(m)} \right) + C^{(m)} \quad (11)$$

where the scalar  $C^{(m)}$  does not depend on  $X^{(m)}$ . The quadratic part of this expression is an anisotropic  $L_2$  norm of the difference vector  $X^{(m)} - Y^{(m)}$ . When experimenting with the global optimization scheme practically identical results at convergence are obtained by replacing matrix  $M^{(m)}$  by the scalar  $\mu^{(m)}$  equal to the mean value of the diagonal elements i.e.  $\mu^{(m)} = \text{tr}(M^{(m)})/N$ . The above expression for  $F^{(m)}$  now simplifies to  $F^{(m)} = \mu^{(m)} \|X^{(m)} - Y^{(m)}\|^2$ . The final nodal coefficient vectors  $X^{(m)}$  will be determined by optimizing the total energy of Eq.1. The nodes for which  $\mu^{(m)}$  is very small will have almost no influence on the global optimum of the coefficient vector  $X^{(m)}$ . In the particular case where  $\rho$  is identically zero within the window,  $\mu^{(m)} = 0$  but this can be easily detected early on so that further computations within the corresponding window may be skipped and a zero weight  $\mu^{(m)} = 0$  attributed to the node. In general,  $\mu^{(m)}$  are data attachment weights playing a role, at nodal level, similar to the role that the edgeness weights  $\rho(\underline{x})$  play at pixel or voxel level.

### APPENDIX C. INTRODUCING THE NON-CONFORMITY REGULARIZATION ENERGY

This is the natural regularization energy of our FEM representation. We define it as a sum over all pairs of nodes  $m, n$  which have non empty overlap of their window support i.e.  $\Omega_{m,n} \equiv \{\underline{x} | \psi^{(m)}(\underline{x})\psi^{(n)}(\underline{x}) > 0\}$  is not empty

$$Reg(\varphi) = \sum_{m < n} \int_{\Omega_{m,n}} \psi^{(m)}(\underline{x})\psi^{(n)}(\underline{x}) \left( \theta^{(n)}(\underline{x}) - \theta^{(m)}(\underline{x}) \right)^2 \quad (12)$$

which is manifestly a quadratic semi-norm vanishing if and only if all windows have identical local polynomials  $\theta^{(m)}(\underline{x})$  of degree  $p$  and by virtue of the partition of unity property,  $\varphi(\underline{x})$  would then be equal to the same polynomial. The contribution  $NC(m, n)$  of any overlapping pair of windows can readily be expressed as a quadratic function of their coefficients  $\{a_r^{(m)} | r \in [0, N - 1]\}$  and  $\{a_r^{(n)} | r \in [0, N - 1]\}$  in the form

$$\begin{aligned} NC^{(m,n)}(X^{(m)}, X^{(n)}) &= \sum_{r,s} K_{r,s}^{m,n} a_r^{(m)} a_s^{(m)} - 2 \sum_{r,s} C_{r,s}^{m,n} a_r^{(m)} a_s^{(n)} + \sum_{r,s} L_{r,s}^{m,n} a_r^{(n)} a_s^{(n)} \\ &= a^{(m)T} K^{m,n} a^{(m)} - 2a^{(m)T} C^{m,n} a^{(n)} + a^{(n)T} L^{m,n} a^{(n)} \end{aligned} \quad (13)$$

where the entries of the matrices  $K^{m,n}$ ,  $C^{m,n}$  and  $L^{m,n}$  are:

$$K_{r,s}^{m,n} = \int_{\Omega_{m,n}} \psi^{(m)}(\underline{x})\psi^{(n)}(\underline{x}) v_r^{(m)}(\underline{x}) \cdot v_s^{(m)}(\underline{x}) \quad (14)$$

$$C_{r,s}^{m,n} = \int_{\Omega_{m,n}} \psi^{(m)}(\underline{x})\psi^{(n)}(\underline{x}) v_r^{(m)}(\underline{x}) \cdot v_s^{(n)}(\underline{x}) \quad (15)$$

$$L_{r,s}^{m,n} = \int_{\Omega_{m,n}} \psi^{(m)}(\underline{x})\psi^{(n)}(\underline{x}) v_r^{(n)}(\underline{x}) \cdot v_s^{(n)}(\underline{x}) \quad (16)$$

All the above square matrix entries are readily evaluated analytically because they depend on integrals of dimensionally separable polynomials over a rectangular domain.

We can now put together the overall energy to optimize in our FEM model:

$$E(X) = \sum_m \mu^{(m)} \|X^{(m)} - Y^{(m)}\|^2 + \lambda \cdot \sum_{m,n} NC^{(m,n)}(X^{(m)}, X^{(n)}) \quad (17)$$

We use the conjugate gradient algorithm<sup>9</sup> to find the set of node coefficient vectors  $X^{(m)}$  which minimize  $E$ . We can see that the nodal weight  $\mu^{(m)}$  introduced in the previous section define the confidence to give to a local solution  $Y^{(m)}$ . When the number of nodes is large, total convergence of the conjugate gradient would require a large number of iterations (equal to the number of unknowns). To alleviate this problem, we make use of a PUFEM multi-resolution dyadic pyramid<sup>4</sup> with its prolongation operator allowing to accurately reproduce

the map  $\phi(\underline{x})$  obtained from a coarse PUFEM level with inter-node separation  $2h$  to a finer level with inter-node separation  $h$ . We therefore start at a coarse level (with a corresponding coarse scale image) for which minimization can be fast and then use the prolongation operator to initialize the algorithm at the next finer PUFEM level and image scale and proceed likewise up the finest level and scale of the pyramid.

The nodal data attachment weights  $\mu^{(m)}$  are dependent on the dynamic range of the raw image. The smoothness weight  $\lambda$  must be adjusted accordingly. A convenient way for doing this is just to normalize, at each level and scale, the  $\mu^{(m)}$  weights by a soft-peak value  $\mu^{(peak)}$  defined by  $\mu^{(peak)} = \sum_m (\mu^{(m)})^2 / \sum_m (\mu^{(m)})$ . It is important to note that in our approach, the local optimization step of Appendix B which is needed to compute  $\mu^{(m)}$  and  $Y^{(m)}$  is not iterative and is just done once for each level and and scale. Only the minimization of the node-wise discretized energy defined by Eq.17 requires conjugate gradient iterations. It must be recalled that these iterations have moderate computational cost because there are, in general, much fewer nodes than elements in the images.

#### APPENDIX D. RELATION BETWEEN THE PROPOSED SMOOTHNESS ENERGY WITH POPULAR DIFFERENTIAL SMOOTHNESS ENERGIES

It is interesting to analyze the long range (low spatial frequency) smoothing effect of our FEM native regularization energy defined in Eq.12. To do this, we note that, by virtue of the flatness property of the window functions  $\psi^{(m)}(\underline{x})$  when  $\underline{x}$  coincides with the centroid of node  $m$   $\xi^{(m)}$  (see Appendix A), the value of  $\varphi(\underline{x})$  and its multivariate partial derivatives of order less or equal to  $p$  are equal to those of the local polynomial  $\theta^{(m)}(\underline{x})$ . The corresponding partial derivatives of  $\theta^{(m)}(\underline{x})$  are in turn linearly related to the coefficients  $a_r^{(m)}$ . We use for  $\varphi(\underline{x})$  a test function and express it in a Taylor expansion around any node centroid  $\underline{\xi}^{(m)}$  up to an order larger than  $p$ , the coefficients  $a_r^{(m)}$  of this nodes as well as those  $a_r^{(n)}$  of all neighboring nodes having a non empty overlap sub-domain  $\Omega_{m,n}$  can be expressed in terms of  $\varphi$  values and its partial derivatives at  $\underline{\xi}^{(m)}$ . We can now evaluate the nonconformity energy contribution of the sub-domain  $\Omega_m$  by making use of Eqs.(13-16) as a non-negative quadratic function (by construction) of the partial derivatives of  $\varphi$  at  $\underline{\xi}^{(m)}$  which can be written as a sum of terms of the form

$$C_{(\underline{\alpha}, \underline{\delta})} \cdot h^{d+|\underline{\alpha}|+|\underline{\delta}|} (D^{\underline{\alpha}}\varphi)_{(m)} (D^{\underline{\delta}}\varphi)_{(m)} \quad (18)$$

where  $(D^{\underline{\alpha}}\varphi)_{(m)}$  stands for the multivariate partial derivative of  $\varphi$  at  $\underline{\xi}^{(m)}$  for the multiplicity vector  $\underline{\alpha} = (\alpha_1, \dots, \alpha_d)$  and  $|\underline{\alpha}| = \sum_{i=1}^d \alpha_i$  i.e.

$$D^{\underline{\alpha}}\varphi \equiv \frac{\partial^{|\underline{\alpha}|}\varphi}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}} \quad (19)$$

likewise for  $D^{\underline{\delta}}\varphi$ . Each such term involves a power  $d + |\underline{\alpha}| + |\underline{\delta}|$  of the inter-node separation  $h$  that depends on the sum of differentiation orders of the term's factors. As expected, no partial derivative of order less or equal than  $p$  appears in the result. The expansion corresponding to  $Reg(\varphi)$  can be deduced by taking a sum over all sub-domain  $\Omega_m$  contributions. Retaining only the lowest power of  $h$  in the expansion yields in 2D for  $p = 1$  ( $d = 3, p = 1$ ) the expression

$$E_2 \cong \sum_m \gamma h^6 \left( \left( \frac{\partial^2 \varphi}{\partial x_1^2} \right)^2 + \left( \frac{\partial^2 \varphi}{\partial x_2^2} \right)^2 + \beta \left( \frac{\partial^2 \varphi}{\partial x_1 \partial x_2} \right)^2 \right)_m \quad (20)$$

and in 3D ( $d = 3, p = 1$ )

$$E_3 \cong \sum_m \gamma h^7 \left( \left( \frac{\partial^2 \varphi}{\partial x_1^2} \right)^2 + \left( \frac{\partial^2 \varphi}{\partial x_2^2} \right)^2 + \left( \frac{\partial^2 \varphi}{\partial x_3^2} \right)^2 + \beta \left( \frac{\partial^2 \varphi}{\partial x_1 \partial x_2} \right)^2 + \beta \left( \frac{\partial^2 \varphi}{\partial x_1 \partial x_3} \right)^2 + \beta \left( \frac{\partial^2 \varphi}{\partial x_2 \partial x_3} \right)^2 \right)_m \quad (21)$$

where  $\gamma = 0.002183$  and  $\beta = 4.0675$ . The above analysis leading to the discrete integrals over the FEM domain (20,21) provides some insight on similarities between our native FEM regularization energy with the very popular thin-plate smoothness energy<sup>3</sup>  $\int_{\Omega} \left\{ \sum_i \left( \frac{\partial^2 \varphi}{\partial x_i^2} \right)^2 + 2 \sum_{i < j} \left( \frac{\partial^2 \varphi}{\partial x_i \partial x_j} \right)^2 \right\}$ .

Retaining the lowest power of  $h$  makes sense in particular within the coarse to fine FEM setting that we use (recall that  $h$  is halved when going from an FEM level to the next finer). Also, note that the full expansion of  $Reg(\varphi)$  can be expressed in Fourier domain in which case, the terms involving a factor  $h^{d+k}$  in the above expansion with  $k > 0$  will involve spatial frequency factors in the form  $\prod_{i=1}^d \omega_i^{\beta_i}$  where  $\beta_i \geq 0$  for all  $i$  with a total degree  $\sum_{i=1}^d \beta_i = k$ . Therefore, the long-range smoothing will be dictated by the lowest degree terms in this  $\omega_i$  expansion which correspond to the lowest power of  $h$  (i.e.  $h^{d+2p+2}$ ).

## REFERENCES

- [1] A.M. Lopez, F. Lumbreras, J. S. and Villanueva, J., "Evaluation of methods for ridge and valley detection," *IEEE Transaction on Pattern Analysis and Machine Intelligence* **21(4)**, 327–335 (1999).
- [2] Canny, J., "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence* **8(6)**, 679–698 (1986).
- [3] Blake, A. and Zisserman, A., [*Visual Reconstruction*], MIT Press, Cambridge, Mass. (1987).
- [4] Makram-Ebeid, S. and Somphone, O., "Non-rigid image registration using a hierarchical partition of unity finite element method," in [*Proceedings of the 2007 International Conference on Computer Vision*], (2007).
- [5] Babuška, I. and Melenk, J. M., "The partition of unity method," *International Journal of Numerical Methods in Engineering* **40(4)**, 727–758 (1997).
- [6] Melenk, J. M. and Babuška, I., "The partition of unity finite element method: Basic theory and applications," *Computer Methods in Applied Mechanical Engineering* **139(1-4)**, 289–314 (1996).
- [7] Croll, J. G. A., "The treatment of natural boundary conditions in the finite element and finite difference methods," *International Journal for Numerical Methods in Engineering* **5(3)**, 443–446 (1973).
- [8] D. Lesage, E. Angelini, I. B. and Funka-Lea, G., "Design and study of flux-based features for 3d vascular tracking," in [*IEEE International Symposium on Biomedical Imaging (ISBI)*], 286–289 (2009).
- [9] Barrett, R., [*Templates for the Solution of Linear Systems, 2nd ed.*], SIAM, Philadelphia, PA (1994).