

Learning Fully-Connected CRFs for Blood Vessel Segmentation in Retinal Images

José Ignacio Orlando^{1,2,3} and Matthew Blaschko^{1,4}

¹ Équipe Galen, INRIA Saclay, Île-de-France, France

² Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Argentina

³ Pladema Institute, UNCPBA, Argentina

⁴ Center for Learning and Visual Computing, École Centrale Paris, France

Abstract. In this work, we present a novel method for blood vessel segmentation in fundus images based on a discriminatively trained, fully connected conditional random field model. Retinal image analysis is greatly aided by blood vessel segmentation as the vessel structure may be considered both a key source of signal, e.g. in the diagnosis of diabetic retinopathy, or a nuisance, e.g. in the analysis of pigment epithelium or choroid related abnormalities. Blood vessel segmentation in fundus images has been considered extensively in the literature, but remains a challenge largely due to the desired structures being thin and elongated, a setting that performs particularly poorly using standard segmentation priors such as a Potts model or total variation. In this work, we overcome this difficulty using a discriminatively trained conditional random field model with more expressive potentials. In particular, we employ recent results enabling extremely fast inference in a fully connected model. We find that this rich but computationally efficient model family, combined with principled discriminative training based on a structured output support vector machine yields a fully automated system that achieves results statistically indistinguishable from an expert human annotator. Implementation details are available at <http://pages.saclay.inria.fr/matthew.blaschko/projects/retina/>.

Keywords: Blood vessel segmentation, Fundus imaging, Conditional Random Fields, Structured Output SVM.

1 Introduction

Retinal images—also known as fundus images or retinographies—are projective color images of the inner surface of the human eye (Figure 1). They allow physicians to observe the retina and its internal parts, including the vascular tree, the optic disc, the fovea, etc. [1]. Retinal blood vessels provide useful information for several applications, including medical diagnosis and screening of ophthalmological and cardiovascular diseases. The vessels tortuosity, for instance, can be used to characterize hypertensive retinopathy, and the measurement of vessels diameter is utilized to diagnose hypertension [2]. Changes in the vasculature distribution are also interpreted as a possible symptom of diabetic retinopathy [1]. Moreover, vessels are used to assist laser surgeries [3] and as landmarks for image registration [4]. All these applications require the segmentation of the retinal vasculature. In current best practice, trained specialists delineate

the vessels manually, although this is a particularly tedious and time-consuming task. Difficulties in the imaging process—such as the inadequate contrast between vessels and background, and uneven background illumination—increase the variability among segmentations performed by different human observers. These facts motivate the development of automatic strategies for blood vessel segmentation without human intervention.

In the last two decades, several approaches have been introduced to solve this problem, but none of them have yet proved to be accurate enough to be assumed as a standard by the medical community [1]. These segmentation algorithms can be classified into two general categories, supervised and unsupervised. Supervised methods require a set of labeled training samples to learn a model. These samples are composed of pixels with known annotations and their features. Most of the effort in supervised segmentation involves finding new features for training, or better classifiers to perform the pixel classification task. For instance, Ricci and Perfetti [5] combine a linear support vector machine (SVM) with a line detector feature. Becker *et al.* [6] learn the features in a supervised way, based on a gradient boosting framework. Unsupervised methods usually involve systems based on clustering, region-oriented approaches [7, 8] or thresholding after vessel enhancement with filters [2, 9], line detectors [5, 10] or morphological operators [11]. More complex image processing operations have been combined with vessel centerline detection at different scales to obtain the final segmentation [11, 12]. The main advantage of unsupervised methods is that they do not need to be trained using manual annotations. However, they have reported worse results than supervised approaches [10].

Conditional Random Fields (CRFs) have been widely utilized to solve segmentation problems in several applications [13]. To the best of our knowledge, however, they have not been yet applied to retinal blood vessel segmentation in fundus images. This is likely due to that standard pairwise potentials such as in a Potts model assign a low prior to the elongated structures that comprise a vessel segmentation. We overcome this by using a much richer class of potentials.

The main contribution of our paper consists of an automatic method for blood vessels segmentation in retinal images based on fully-connected CRFs. We follow the efficient inference approach proposed in [14], and we learn its configuration in a supervised way, using a Structured Output SVM (SOSVM) [15]. We validate the approach on the main benchmark data set for vessel segmentation in fundus images, the DRIVE data set [16]. In contrast with several published works, no test data was utilized to adjust the parameters of the method. Instead, we randomly split the original training data into two subsets, one for training and the other for validation. Once we found the best configuration using the validation set, we applied it over the images in the test set. In this statistically principled setting, we report a fully-automated system that achieves a performance statistically indistinguishable from an expert human annotator.

The remainder of this paper is organized as follows: the formulation of the fully-connected CRF is described in Section 2; in Section 3 we explain how we learn the CRF model using a SOSVM. In Section 4 we include our results and a comparison against other state-of-the-art works. Finally, Section 5 concludes the paper.

2 Fully-Connected CRF Segmentation

We pose the segmentation task as an energy minimization problem in a fully-connected conditional random field (CRF). In the original definition of CRFs, images are mapped to graphs, where each pixel represents a node, and every node is connected with an edge to its immediate neighbors [13]. In the fully-connected version, each node is assumed to be a neighbor of every other. Following this approach the method is able to take into account not only neighboring information but also long-range interactions between pixels. This property improves the segmentation accuracy, but makes the inference process computationally expensive. Recently, however, Krähenbühl and Koltun [14] have introduced an efficient inference approach under the restriction that the pairwise potentials are a linear combination of Gaussian kernels over a Euclidean feature space. This approach, which is based on taking a mean field approximation of the original CRF, is able to provide accurate segmentations in less than a second.

We denote $\mathbf{y} = \{y_i\}$ as a labeling over all pixels of the image in the label space $\mathcal{L} = \{-1, 1\}$, where 1 is associated to blood vessels and -1 to any other class. Its corresponding energy function is given by the sum of its unary energy ψ_u and its pairwise energy ψ_p :

$$E(\mathbf{y}) = \sum_i \psi_u(y_i, \mathbf{x}_i) + \sum_{i < j} \psi_p(y_i, y_j, \mathbf{f}_i, \mathbf{f}_j) \quad (1)$$

where \mathbf{x}_i and \mathbf{f}_i are the unary and pairwise features, respectively. Unary potentials define a log-likelihood over the label assignment \mathbf{y} , and they are traditionally computed by a classifier [14]. Pairwise potentials define a similar distribution but considering only the interactions between pixels features and their labels. Parameters for both potentials are learned using a Structured Output SVM, as we explain in detail in Section 3.

Unary Potentials: We obtain the unary potentials according to the following expression:

$$\psi_u(y_i, \mathbf{x}_i) = -\langle \mathbf{w}_{u_{y_i}}, \mathbf{x}_i \rangle - \beta_{y_i} \quad (2)$$

where $\mathbf{w}_{u_{y_i}}$ is a weight vector and β_{y_i} is a bias term, respectively, both associated to the label y_i .

Pairwise Potentials: Following the restriction imposed by the inference approach we use, our pairwise potentials are obtained as follows:

$$\psi_p(y_i, y_j, \mathbf{f}_i, \mathbf{f}_j) = \mu(y_i, y_j) \sum_{m=1}^M w_p^{(m)} k^{(m)}(f_i^{(m)}, f_j^{(m)}) \quad (3)$$

where each $k^{(m)}$ is a Gaussian kernel over an arbitrary feature $f^{(m)}$, $w_p^{(m)}$ is a linear combination weight, and $\mu(y_i, y_j)$ represents a label compatibility function. In general, our pairwise kernels have the following form:

$$k^{(m)}(f_i^{(m)}, f_j^{(m)}) = \exp\left(-\frac{|\mathbf{p}_i - \mathbf{p}_j|^2}{2\theta_p^2} - \frac{|f_i^{(m)} - f_j^{(m)}|^2}{2\theta_{(m)}^2}\right) \quad (4)$$

where \mathbf{p}_i and \mathbf{p}_j are the coordinate vectors of pixels i and j . We include positions in order to increase the effect of close pixels over distant ones. Parameters θ_p and $\theta_{(m)}$ control the degree of relevance of the two parts of the kernels into the expression. The scale value $\theta_{(m)}$ is estimated as the median of the distances over pairs $(f_i^{(m)}, f_j^{(m)})$ [17]. In the same way, θ_p is obtained using $f^{(m)} = \mathbf{p}$. The compatibility function μ is given by the Potts model, $\mu(y_i, y_j) = [y_i \neq y_j]$. It penalizes nearby similar pixels that are assigned to different labels.

3 Learning CRF’s Parameters Using Structured Output SVM

Our goal is to learn a vector $\mathbf{w} = (\mathbf{w}_u, \mathbf{w}_\beta, \mathbf{w}_p)$, where \mathbf{w}_u , \mathbf{w}_β and \mathbf{w}_p are the weights for the unary features, for the bias term and for the pairwise kernels, respectively. We obtain it in a supervised way, using the 1-slack formulation of the Structured Output SVM with margin-rescaling presented in [15].

Let the training set $S = \{(s^{(1)}, y^{(1)}), \dots, (s^{(n)}, y^{(n)})\}$, where n is the number of training images. Each $y^{(i)}$ corresponds to the ground truth of the i -th image in the training set. Each set $s^{(i)} = \{x^{(i)}, B, f^{(i)}\}$ contains the set $x^{(i)}$ of unary feature vectors, a bias constant B , and the set $f^{(i)}$ of pairwise features for every pixel in the image.

The weights \mathbf{w} are obtained by solving:

$$\begin{aligned} \min_{\mathbf{w}, \xi \geq 0} & \frac{1}{2} \|\mathbf{w}\|^2 + C\xi & (5) \\ \text{s.t.} & \forall \left(\bar{y}^{(1)}, \dots, \bar{y}^{(n)}\right) : \sum_{i=1}^n \langle \mathbf{w}, \varphi(s^{(i)}, y^{(i)}) - \varphi(s^{(i)}, \bar{y}^{(i)}) \rangle \geq \sum_{i=1}^n \Delta(y^{(i)}, \bar{y}^{(i)}) - \xi \end{aligned}$$

where C is a regularization constant; ξ is a slack variable shared across all the constraints $\bar{y}^{(i)}$; $\varphi(s, y)$ is a feature map function that relates a given set s with a given labeling y ; and $\Delta(y, \bar{y})$ is a loss function that evaluates the difference between a ground truth y and a constraint \bar{y} . In this work, we define Δ as the typical Hamming loss:

$$\Delta(y, \bar{y}) = \sum_i [y_i \neq \bar{y}_i] \tag{6}$$

Our feature map is defined as follows:

$$\varphi(s, y) = \left(\sum_k \varphi_u(\mathbf{x}_k, y_k), \sum_k \varphi_\beta(B, y_k), \sum_k \sum_{j < k} \varphi_p(y_k, y_j, \mathbf{f}_k, \mathbf{f}_j) \right) \tag{7}$$

where the components represent the sum of the unary feature map, the bias feature map and the pairwise feature map, respectively, for all the pixels in the image. We make precise the definitions of φ_u , φ_β , and φ_p in the sequel.

We define a binary vector $\varphi_y(y_i) \in \{0, 1\}^{|\mathcal{L}|}$ such that:

$$\varphi_y(y_i) \begin{cases} (1, 0) & \text{if } y_i = -1 \\ (0, 1) & \text{if } y_i = 1 \end{cases} \tag{8}$$

The individual feature maps are obtained as follows:

$$\varphi_u(\mathbf{x}_k, y_k) = \mathbf{x}_k \otimes \varphi_y(y_k) \quad (9)$$

$$\varphi_\beta(B, y_i) = B\varphi_y(y_i) \quad (10)$$

$$\forall m : [\varphi_p(y_k, y_j, \mathbf{f}_k, \mathbf{f}_j)]_m = \mu(y_i, y_j)k^{(m)}(f_i^{(m)}, f_j^{(m)}) \quad (11)$$

where \otimes is the Kronecker product. Eq. (5) is solved using a cutting-plane approach [15].

4 Validation and Results

We validate our method using the publicly available data set DRIVE [16]. DRIVE is widely utilized in the state-of-the-art to quantify blood vessel segmentation performance. The data set is divided into two sets: one for training and one for test, each of them containing 20 images. The test set provides two manual segmentations generated by two different experts for each image. The selection of the first observer is accepted as ground truth and used for performance evaluation in literature. The training set includes a set of manual segmentations made by the first observer.

We performed the evaluation of our method in terms of sensitivity (Se) and specificity (Sp):

$$Se = \frac{TP}{TP + FN}, \quad Sp = \frac{TN}{TN + FP}, \quad (12)$$

where TP , TN , FP and FN represent the number of true positives, true negatives, false positives and false negatives, respectively. Sensitivity and specificity are indicators of the capability of the algorithm to identify blood vessels and to exclude all other classes, respectively.

The value of C and the combination of features utilized to compute the unary and pairwise energies of the CRF were selected without using test data. We have divided the original training set into two new subsets, *training** and *validation*, containing 75% and 25% of the images, respectively. We utilized *training** to train the model, and *validation* to optimize the combination of parameters. We then used the selected parameters to re-train on the *training** set, and we evaluated the final model on the test set. During model selection, we used a criterion that minimized the distance with respect to the Se and Sp of the second human observer.

Several state-of-the-art features were considered, including wavelets [18], line [5, 10] and ridge detectors [8], matched filter responses [19], vessel enhancers [20–24], gradient module, Huffman's entropy and image intensities. A forward selection process was followed to select the best combination. All features are obtained over preprocessed images. We first select the green band of the original color image. Then we extend the borders of the FOV [18], and finally we subtract an estimated background for bias correction using a median filter with square windows of length 25. For each combination of features, values of $C = 10^i$, with $i = \{0, 1, \dots, 5\}$, were explored. The best configuration we found for the unary features utilize 2-D Gabor wavelets [18], line detectors [5]

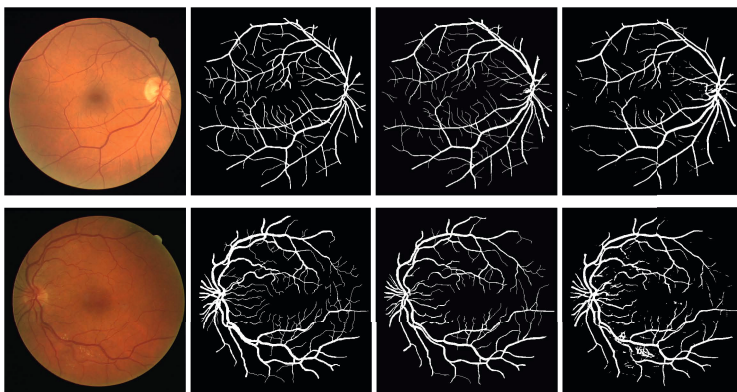


Fig. 1. Examples of results obtained on healthy (top) and pathological (bottom) images from DRIVE. From left to right: original images, ground truth labelings, 2nd human observer labelings, our segmentations.

and the vessel enhancement process proposed in [22]. The vessel enhanced image obtained after applying the process proposed in [23] is utilized as a pairwise feature.

Two segmentation examples, obtained over a healthy and a pathological image, are included in Figure 1. We also include the ground truth labeling and the labeling performed by the second observer for qualitative comparison.

We have performed an extensive comparison of our method with state-of-the-art automated methods that report their results in terms of Se and Sp , and that are not trained on the test set. Figure 2 shows a scatter plot of the competing methods as well as the performance of the human observer, in terms of Se and Sp . We also include results obtained following the same principled discriminative training but using a local-neighborhood based CRF. The numerical values are listed in Table 1.

5 Conclusions

In this work, we have presented a discriminatively trained segmentation model based on a fully connected CRF for the purpose of blood vessel segmentation in fundus images. Conditional random fields have not been previously applied in this setting, likely due to standard pairwise potentials (such as in a Potts model) leading to poor performance in the presence of thin, branching structures. In contrast, our approach is able to achieve high accuracy by considering a much more expressive, fully connected graph structure. The benefits of this approach are demonstrated empirically.

Extensive comparison with state-of-the-art methods has shown that our approach performed best and is a fully-automated segmentation algorithm that achieves results on par with a human annotator as measured by sensitivity and specificity. Our method is statistically tied with the expert annotator for both measures. This is in part due to previous studies focusing on raw pixel accuracy, which ignores the fact that the number of pixels occupied by blood vessels is a relatively small fraction of the image. As a result, competing methods suffer as measured by sensitivity. In contrast, our method closely

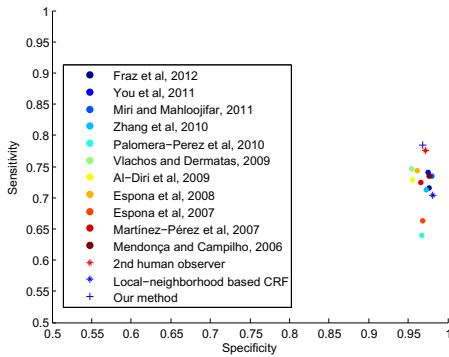


Fig. 2. Scatter-plot of Se vs. Sp , comparing existing methods, local-neighborhood based CRF, the proposed method, and the human annotator, based on DRIVE data set. Our method is statistically tied with the performance of the expert annotator and achieves a much higher sensitivity than all other segmentation systems.

Table 1. Comparison of Se and Sp of our method with respect to the existing blood vessel segmentation algorithms and the human observer, based on DRIVE data set.

Method	Sensitivity	Specificity
Fraz et al., 2012 [11]	0.7152	0.9769
You et al., 2011 [25]	0.741	0.9751
Miri and Mahloojifar, 2011 [2]	0.7352	0.9795
Zhang et al., 2010 [9]	0.712	0.9724
Palomera-Pérez et al., 2010 [26]	0.64	0.967
Vlachos and Dermatas, 2009 [12]	0.747	0.955
Al-Diri et al., 2009 [27]	0.7282	0.9551
Espona et al., 2008 [28]	0.7436	0.9615
Espona et al., 2007 [29]	0.6634	0.9682
Martínez-Pérez et al., 2007 [8]	0.7246	0.9655
Mendonca and Campilho, 2006 [7]	0.7344	0.9764
Local-neighborhood based CRF	0.7042±0.049	0.9815±0.005
Our method	0.785±0.045	0.967±0.007
2nd human observer	0.776±0.059	0.973±0.008

matches the performance of an expert human both in sensitivity and specificity, and achieves a substantially higher sensitivity than all other competing methods. Sensitivity is particularly important as it reflects an accurate estimation of the vessel pixels, the primary goal in vessel segmentation for fundus image analysis. Additional implementation details are available at <http://pages.saclay.inria.fr/matthew.blaschko/projects/retina/>.

Acknowledgements. This work is partially funded by ERC Grant 259112, and FP7-MC-CIG 334380.

References

1. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: Blood vessel segmentation methodologies in retinal images—a survey. *Computer Methods and Programs in Biomedicine* 108(1), 407–433 (2012)
2. Miri, M.S., Mahloojifar, A.: Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction. *IEEE T-BME* 58(5), 1183–1192 (2011)
3. Kanski, J.J., Bowling, B.: *Synopsis of Clinical Ophthalmology*. Saunders Limited (2012)
4. Li, Y., Gregori, G., Knighton, R.W., Lujan, B.J., Rosenfeld, P.J.: Registration of OCT fundus images with color fundus photographs based on blood vessel ridges. *Optics Express* 19(1), 7 (2011)
5. Ricci, E., Perfetti, R.: Retinal blood vessel segmentation using line operators and support vector classification. *IEEE T-MI* 26(10), 1357–1365 (2007)
6. Becker, C., Rigamonti, R., Lepetit, V., Fua, P.: Supervised feature learning for curvilinear structure segmentation. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) *MIC-CAI 2013, Part I. LNCS*, vol. 8149, pp. 526–533. Springer, Heidelberg (2013)
7. Mendonca, A.M., Campilho, A.: Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE T-MI* 25(9) (2006)

8. Martinez-Perez, M.E., Hughes, A.D., Thom, S.A., Bharath, A.A., Parker, K.H.: Segmentation of blood vessels from red-free and fluorescein retinal images. *Medical Image Analysis* 11(1), 47–61 (2007)
9. Zhang, B., Zhang, L., Zhang, L., Karray, F.: Retinal vessel extraction by matched filter with first-order derivative of Gaussian. *Computers in Biology and Medicine* 40(4), 438–445 (2010)
10. Nguyen, U.T., Bhuiyan, A., Park, L.A., Ramamohanarao, K.: An effective retinal blood vessel segmentation method using multi-scale line detection. *Pattern Recognition* (2012)
11. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: Ensemble classification system applied for retinal vessel segmentation on child images containing various vessel profiles. *Image Analysis and Recognition* (2012)
12. Vlachos, M., Dermatas, E.: Multi-scale retinal vessel segmentation using line tracking. *Computerized Medical Imaging and Graphics* 34(3), 213–227 (2010)
13. Li, S.Z.: *Markov Random Field Modeling in Image Analysis*, 3rd edn. Springer (2009)
14. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with Gaussian edge potentials. In: *NIPS* (2012)
15. Joachims, T., Finley, T., Yu, C.N.J.: Cutting-plane training of structural SVMs. *Machine Learning* 77(1), 27–59 (2009)
16. Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B.: Ridge based vessel segmentation in color images of the retina. *IEEE T-MI* 23(4), 501–509 (2004)
17. Schölkopf, B.: *Support Vector Learning*. PhD thesis, Oldenbourg Verlag, Munich (1997)
18. Soares, J.V., Leandro, J.J., Cesar, R.M., Jelinek, H.F., Cree, M.J.: Retinal vessel segmentation using the 2-d Gabor wavelet and supervised classification. *IEEE T-MI* 25(9) (2006)
19. Al-Rawi, M., Qutaishat, M., Arrar, M.: An improved matched filter for blood vessel detection of digital retinal images. *Computers in Biology and Medicine* 37(2), 262–267 (2007)
20. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) *MICCAI 1998*. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
21. Marín, D., Aquino, A., Gegúndez-Arias, M.E., Bravo, J.M.: A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE T-MI* 30(1), 146–158 (2011)
22. Sinthanayothin, C., Boyce, J.F., Cook, H.L., Williamson, T.H.: Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *British Journal of Ophthalmology* 83(8), 902–910 (1999)
23. Saleh, M.D., Eswaran, C.: An efficient algorithm for retinal blood vessel segmentation using h-maxima transform and multilevel thresholding. *Computer Methods in Biomechanics and Biomedical Engineering* 15(5), 517–525 (2012)
24. Zana, F., Klein, J.-C.: Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE TIP* 10(7), 1010–1019 (2001)
25. You, X., Peng, Q., Yuan, Y., Cheung, Y.-M., Lei, J.: Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. *Pattern Recognition* 44(10) (2011)
26. Palomera-Pérez, M.A., Martínez-Pérez, M.E., Benítez-Pérez, H., Ortega-Arjona, J.L.: Parallel multiscale feature extraction and region growing: application in retinal blood vessel detection. *IEEE T-ITB* 14(2), 500–506 (2010)
27. Al-Diri, B., Hunter, A., Steel, D.: An active contour model for segmenting and measuring retinal vessels. *IEEE T-MI* 28(9), 1488–1497 (2009)
28. Espona, L., Carreira, M.J., Penedo, M.G., Ortega, M.: Retinal vessel tree segmentation using a deformable contour model. In: *ICPR* (2008)
29. Espona, L., Carreira, M.J., Ortega, M., Penedo, M.G.: A snake for retinal vessel segmentation. In: Martí, J., Benedí, J.M., Mendonça, A.M., Serrat, J. (eds.) *IbPRIA 2007*. LNCS, vol. 4478, pp. 178–185. Springer, Heidelberg (2007)