Segmentation of Retinal Blood Vessels Using Dictionary Learning Techniques

Taibou Birgui-Sekou, Moncef Hidane, Julien Olivier, Hubert Cardot

Segmentation of Retinal Blood Vessels Using Dictionary Learning Techniques

Taibou Birgui Sekou[1,3], Moncef Hidane[1,3], Julien Olivier[1,3], and Hubert Cardot[2,3]

[1] Institut National des Sciences Appliquées Centre Val de Loire, Blois France,
[2] Université de Tours, Tours, France
[3] LI EA 6300, Tours, France

Abstract. In this paper, we aim at proving the effectiveness of dictionary learning techniques on the task of retinal blood vessel segmentation. We present three different methods based on dictionary learning and sparse coding that reach state-of-the-art results. Our methods are tested on two, well-known, publicly available datasets: DRIVE and STARE. The methods are compared to many state-of-the-art approaches and turn out to be very promising.

Keywords: retinal blood vessel segmentation, medical image segmentation, dictionary learning, sparse coding, linear classification, random forests

1 Introduction

Retinal fundus images are now widely used for the diagnosis of various pathologies, including age-related macular degeneration, diabetic retinopathy and glaucoma. As a part of the central nervous system, the retina, and in particular its vasculature, is also used as a biomarker for early detection of neurodegenerative diseases.

Manual analysis of retinal images by ophthalmologists is a tedious task and, as for other manual delineation tasks, is subject to inter- and intra-operator variability. Thus, automatic and semi-automatic tools have been proposed, in particular, for retinal blood vessel segmentation (RBVS). While these tools are now starting to pervade clinical practice, the low image quality and the scale variation of the vessels still represent major challenges to most recent methods.

We examine in this paper a family of *supervised* RBVS methods based on *sparse representations* in *learned dictionaries*. These methods make use of the sparse representation of a patch around a pixel to determine its label. The differences between the proposed methods depend mainly on the number of learned dictionaries and the way the classifier is trained.

The general framework is composed of three stages: data preparation, segmentation, and post-processing. Data preparation consists in patches extraction and normalization. Then, in the segmentation phase, a patch around each pixel is used by a dictionary learning method to assign its label. The output of the

previous step leads to impulsive segmentation errors and we propose a robust regularization approach to tackle this problem. The remainder of this paper is outlined as follows. We review in Section 2 some recent state-of-the-art methods for RBVS. Section 3 reviews the supervised and unsupervised dictionary learning approaches and explains how both paradigms yield three different methods for RBVS. In Section 4, the experimental setup and the results are exposed. We conclude the paper in Section 5, pointing to possible directions for future work.

2 Related Work

RBVS has attracted a number of researchers for over two decades. Fraz et al. [1] presented a global review of the proposed methods in the field, up to 2012. They divided the methods into 6 classes: machine learning methods, matched filtering methods, morphological processing methods, vessel tracing/tracking methods, multi-scale methods, and model-based methods. It turns out that machine learning methods, especially the supervised ones, are in general the best. A more specific review focusing on computer-aided diagnosis for diabetic retinopathy is presented in [2].

In recent machine learning based approaches, a Lattice Neural Network with Dendritic Processing (LNNDP) framework is presented in [3]. Each pixel is classified using a $5-$dimensional feature vector extracted from an enhanced version of the green channel of the original RGB images. In [4], a discriminative dictionary learning technique is used. Image patches are extracted from an enhanced version of the green channel image to learn a specific dictionary per class. Two classes of patches are considered: patches containing a blood vessel and the others. Given a test image, overlapping patches are first extracted. The class of each patch is attributed according to the dictionary that best represents it. Then a segmentation map of each patch is obtained by thresholding the blood vessel patches and setting to zero the non-vessel ones.

The numerical results reported in [5] and [6] indicate that these works constitute the current state-of-the-art in RBVS. Wang et al. [5] proposed to follow two steps: a hierarchical feature extraction followed by an ensemble classification. The hierarchical features are obtained from different layers of a Convolutional Neural Network (CNN). Then Random Forest classifiers are trained on some levels of the CNN. The final class is obtained with a winner-take-all strategy. Liskowski et al. [6] used a CNN both as a feature extractor and a classifier.

3 RBVS Using Dictionary Learning

In what follows, we write patches as vectors. Let $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ be the input dataset consisting of n patches $\mathbf{x}_i \in \mathbb{R}^m$. In the supervised setting, one has also access to a vector of labels $\mathbf{y} \in \mathbb{R}^n$, with y_i denoting the label associated with the sample $\mathbf{x}_i$. This section first introduces a general view of dictionary learning, then, presents the methods proposed for RBVS.

3.1 Sparse Coding and Dictionary Learning

By choosing an overcomplete family $\{\mathbf{d}_k\}_{k=1}^{p}$ of vectors in $\mathbb{R}^m$ one can decompose an input patch $\mathbf{x} \in \mathbb{R}^m$ as a linear combination $\mathbf{x} = \sum_{k=1}^{p} a_k \mathbf{d}_k = \mathbf{D}\mathbf{a}$, where $\mathbf{D} \in \mathbb{R}^{m \times p}$ is called a *dictionary*, and $\mathbf{a} \in \mathbb{R}^p$ is the vector of coefficients. Due to overcompleteness, the previous decomposition is not unique. In order to enforce sparsity, the following formulation has been widely adopted:

$$\mathbf{a}^* \leftarrow \min_{\mathbf{a} \in \mathbb{R}^p} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_1, \tag{1}$$

where λ balances the trade-off between sparsity and reconstruction error, $\|.\|_q$ is the ℓ_q-norm[1]. The first term ensures a good reconstruction of the patch $\mathbf{x}$ from the dictionary, and the last term encourages the vector to be sparse. In general, increasing the value of λ yields sparser solutions.

The general idea of dictionary learning is to learn $\mathbf{D}$ from the dataset $\mathbf{X}$ by ensuring that each patch $\mathbf{x}_i$ is decomposed in a parsimonious manner (*see* [7] and references therein for more details). The formulation we retain for our present work is the following [8]

$$\mathbf{D}^*, \mathbf{A}^* \leftarrow \min_{\mathbf{D} \in \mathbb{R}^{m \times p}, \mathbf{A} \in \mathbb{R}^{p \times n}} \Big[\mathcal{R}(\mathbf{X}, \mathbf{D}, \mathbf{A}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|_2^2 + \lambda \|\mathbf{a}_i\|_1 \Big], \tag{2}$$

where $\mathbf{A} = [\mathbf{a}_1, ..., \mathbf{a}_n]$ is the matrix of sparse coefficients. To resolve scale ambiguity, the columns of $\mathbf{D}$ are further constrained to be in the unit Euclidean ball. This constraint is applied to all subsequent dictionary learning variants.

3.2 RBVS by Sparse Coding then Classifier (SCTC)

This method first learns a dictionary that best represents the entire training set (without class discrimination). Then, a random forest (RF) classifier is trained on the generated sparse codes. The method uses the dictionary learning phase as a high dimensional feature extractor. The training is done in the two following steps:

– Step 1: Solve a classical dictionary learning problem (Eq. (2)).

– Step 2: Train a random forest classifier of 50 trees using the matrix of sparse coefficients produced in Step 1 as input.

Given a patch query $\mathbf{z}$, its label l is computed by first solving a sparse coding problem using the learned dictionary. Then, the classifier is applied on the produced vector of coefficients to predict the label.

3.3 RBVS by Joint Dictionary and Classifier Learning (JDCL)

This method consists in learning jointly a dictionary and a linear classifier instead of separating them as done in SCTC. Introduced in [9], the method is formulated as follows

$$\mathbf{D}^*, \mathbf{A}^*, \mathbf{W}^* \leftarrow \min_{\mathbf{D}, \mathbf{A}, \mathbf{W}} \alpha \|\mathbf{L} - \mathbf{W}\mathbf{A}\|_F^2 + \beta \|\mathbf{W}\|_F^2 + \mathcal{R}(\mathbf{X}, \mathbf{D}, \mathbf{A}), \tag{3}$$

[1] The ℓ_q-norm ($q \geq 1$) of a vector $\mathbf{x}$ is: $\|\mathbf{x}\|_q = [\sum_i \mid x[i] \mid^q]^{1/q}$

where $\mathcal{R}$ is as defined in Eq. (2), $\mathbf{W} \in \mathbb{R}^{N \times p}$ contains the linear classifier's parameters, N is the number of classes, α and β are weight parameters, $\|.\|_F$ is the Frobenius norm[2], and $\mathbf{L} = [\mathbf{l}_1, ..., \mathbf{l}_n]$ is the label matrix where the vector $\mathbf{l}_i \in \mathbb{R}^N$ is 1 at the index corresponding to the class of the sample $\mathbf{x}_i$ and 0 elsewhere. Eq. (3) can be solved efficiently using standard dictionary learning techniques (*see* [9] for more optimization details).

The label l of a query patch $\mathbf{z}$ is obtained using the following equation:

$$l = \underset{i=1,...,N}{\arg\max} \, [e_i = \mathbf{W}^* \mathbf{a}^*], \tag{4}$$

where $\mathbf{W}^*$ is the previously learned classifier of Eq. (3) and $\mathbf{a}^*$ is obtained by applying Eq. (1) using the query patch $\mathbf{z}$ and the learned dictionary $\mathbf{D}^*$.

3.4 RBVS by One Dictionary per Class (DPC)

Let $\mathbf{X} = [\mathbf{X}_1, ..., \mathbf{X}_N]$ be the division of the dataset into sub-matrices, where each sub-matrix $\mathbf{X}_i \in \mathbb{R}^{m \times n_i}$ contains only the n_i samples that belong to the class i, and N is the number of classes ($N = 2$ in our setting).

In this method, we learn independently one dictionary on each sub-matrix. A query patch is classified by selecting the associated class of the dictionary that best reconstructs it, similarly to [10].

A dictionary $\mathbf{D}_i^*$ associated with the class i is learned using the corresponding sub-matrix by solving the problem in Eq. (2).

The label l of a query patch $\mathbf{z}$ is obtained by first computing the associated sparse coefficient $\mathbf{a}_i^*$ (using Eq. (1)) on each learned dictionary. Then, the following equation is used to predict the label:

$$l = \underset{i=1,...,N}{\arg\min} \Big[e_i = \|\mathbf{z} - \mathbf{D}_i^* \mathbf{a}_i^*\|_2^2 + \lambda \|\mathbf{a}_i^*\|_1 \Big]. \tag{5}$$

3.5 Post-Processing: Total Variation with ℓ_1 Fidelity Norm (TV$-\ell_1$)

The three proposed methods produce systematic errors in the form of *impulse* noise. This is a common issue encountered in most pixel classification methods. The image c in Fig. 1 shows a typical example, here obtained after applying the SCTC method. We formulate the post-processing as a denoising problem: we seek to recover a clean, *piecewise-constant* classification image from a noisy version. We adopt the TV-ℓ_1 model of [11] since it accounts both for our prior (piecewise constant solution) and for the likelihood (impulse noise). This leads to the following variational problem

$$I^* \leftarrow \min_I \|\nabla I\|_1 + \kappa \|I - I_0\|_1, \tag{6}$$

where I_0 is a noisy classification image. We note at this point that we do not impose any binary constraints in (6) but we simply threshold I^* after solving (6).

[2] The *Frobenius*-norm of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is: $\|\mathbf{A}\|_F = \left[\sum_{i=1}^{m} \sum_{j=1}^{n} A[i,j]^2 \right]^{1/2}$

4 Experiments

The previously presented methods are tested on the following datasets:
– The DRIVE (Digital Retinal Images for Vessel Extraction) [12] dataset contains 40 expert annotated color retinal images taken with a fundus camera. It is divided into two sets of 20 images: the training and testing sets. Each image comes with a ground-truth segmentation (two for the test images and one for the training ones) and a mask image delineating the field of view (FOV). The first observer's ground-truth is considered in this paper.
– The STARE (STructured Analysis of the REtina) [13] is another well-known, publicly available database. The dataset is composed of 20 color fundus photographs. Half of the images presents pathological cases and contains abnormalities, which make the segmentation task even harder. Unlike the DRIVE dataset, the mask images are not provided. We construct them with a threshold on the grayscale images followed by a morphological filter with a structuring element of size 10 pixels.

4.1 Data Preparation

Given the contrast variation from one image to another, data preparation aims at normalizing the illumination beforehand. The grayscale versions of the original RGB images are considered throughout this experiment.

Pre-Processing
– Image normalization and patch extraction: the first normalization consists in applying the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm to the grayscale image. Then, all pixels outside the FOV are set to zero. For a given pixel, we extract the centered squared neighborhood patches of size 8×8. On both datasets, about 140 000 pixels are randomly selected to build our training set (*i.e.* around 70 000 patches per class). Patches with standard deviation less than 0.15 are not considered in the training set.
– Patch normalization: the squared neighborhood patches are then flattened into 64-dimensional vectors. Additional contrast normalization consists in normalizing each patch vector to have unit ℓ_2-norm.

Post-Processings
After classifying each pixel of an image, we first multiply the resulting image with an eroded version of the mask image. This procedure aims at removing the pixels on the edges of the FOV. Then, we apply a TV-ℓ_1 regularization.

4.2 Experimental Setup and Measurements

The number of atoms p in the dictionary depends on the method: for SCTC $p = 1000$, for JDCL $p = 1000$, and finally $p = 500$ for DPC for each subdictionary. The online dictionary learning [14], available in the sparse modeling

software [3] (SPAMS), is used in all our experiments as a dictionary learning algorithm. The primal-dual algorithm of Chambolle and Pock [15] is used for solving the TV-ℓ_1 problem (6). The parameters λ (Eq. (2)) and κ (Eq. (6)) are set, respectively, to 0.5 and 0.9. Note that, all these values are obtained using a grid-search and cross-validating on the training sets.

Let TP, TN, FN, and FP respectively denote the number of true positive, true negative, false negative, and false positive. We use the sensitivity $\textbf{Sens} = \frac{TP}{TP+FN}$, the specificity $\textbf{Spec} = \frac{TN}{TN+FP}$ and the accuracy $\textbf{Acc} = \frac{TP+TN}{TP+FN+TN+FP}$ to quantify the performance of the RBVS methods.

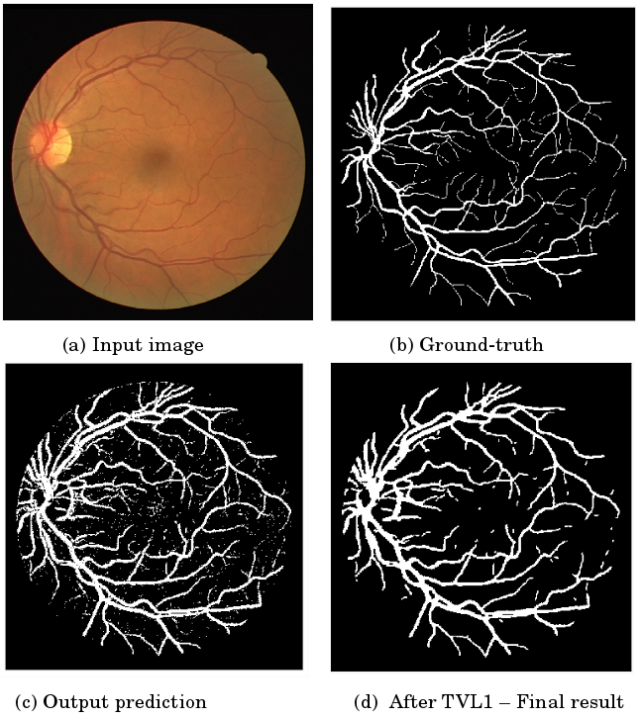

Fig. 1. Segmentation example using SCTC

4.3 Results and Discussions

Our results are depicted on Table 1 along with state-of-the-art results. Fig. 1 illustrates a segmentation example.

[3] http://spams-devel.gforge.inria.fr/

Among the proposed methods, the SCTC approach seems to be the best only in terms of vessel detection (*i.e.* good sensitivity) while the DPC approach outputs good results with respect to all the performance measures.

The SCTC and JDCL methods tend to classify each patch with some line as vessel. This is due to the fact that these patches and the true vessel patches activate the same atoms, thus their sparse vectors are quite close. This problem is reduced when using the DPC model which uses the reconstruction error to classify a patch. Still, all the proposed methods reach the state-of-the-art results on the two datasets while being simple in terms of their architecture and number of parameters.

Other experiments, not reported in this paper, have been carried out with larger patches (*e.g.* 16×16). It turns out that the sensitivity can be improved but we are loosing on the specificity. This is due to the fact that, when using larger patches, more pixels near a blood vessel (but not belonging to it) tend to be classified as vessels.

	DRIVE			STARE		
Methods	**Spec**	**Sens**	**Acc**	**Spec**	**Sens**	**Acc**
This paper - SCTC	95.55	83.49	94.48	94.46	85.11	93.81
This paper - JDCL	96.32	80.60	94.93	95.78	77.24	94.45
This paper - DPC	97.05	77.88	95.36	96.75	75.58	95.23
Javidi et al.[4]	97.02	72.01	94.50	96.53	77.80	95.17
Singh et al.[16]	-	75.94	95.22	-	79.39	92.70
Orlando et al.[17]	96.84	78.97	-	97.38	76.80	-
Vega et al.[3]	96.00	74.44	94.12	96.71	70.19	94.83
Liskowski et al.[6]*	96.73	84.60	95.07	97.10	92.89	96.67
Wang et al.[5]*	97.33	81.73	97.67	97.91	81.04	98.13
Dasgupta et al. [18]*	98.01	76.91	95.33	-	-	-

*deep learning methods

Table 1. Our results on DRIVE and STARE versus the state-of-the-art.

5 Summary & Perspectives

In this paper, we presented three RBVS methods based on sparse representations in learned dictionaries. We showed that these methods can reach state-of-the-art results on the DRIVE and STARE datasets while remaining conceptually simple and computationally tractable. Future work will concentrate on taking patch correlations into account when learning the dictionary and on using more discriminative features.

References

1. Fraz, M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A., Owen, C.,Barman, S.: Blood vessel segmentation methodologies in retinal images - A survey. Computer Methods and Programs in Biomedicine, 108 (1), 407–433 (2012).
2. Mookiah, M.R.K., Acharya, U.R., Chua, C.K., Lim, C.M., Ng, E.Y.K., Laude, A.: Computer-aided diagnosis of diabetic retinopathy: A review. Comp. in Bio. and Med., 43 (12), 2136–2155 (2013).
3. Vega, R., Sánchez-Ante, G., Falcón-Morales, L., Sossa, H., Guevara, E.: Retinal vessel extraction using Lattice Neural Networks with dendritic processing. Comp. in Bio. and Med. 58, 20–30 (2015).
4. Javidi, M., Pourreza, H.R., Harati, A.: Vessel segmentation and microaneurysm detection using discriminative dictionary learning and sparse representation. Computer Methods and Programs in Biomedicine 139, 93–108 (2017).
5. Wang, S., Yin, Y., Cao, G., Wei, B., Zheng, Y., Yang, G.: Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. Neurocomputing 149, 708–717 (2015).
6. Liskowski, P., Krawiec, K.: Segmenting Retinal Blood Vessels With Deep Neural Networks. IEEE Trans. Med. Imaging 35 (11), 2369–2380 (2016).
7. Mairal, J., Bach, F., Ponce, J.: Sparse Modeling for Image and Vision Processing. Foundations and Trends in Computer Graphics and Vision 8(2-3), 85–283 (2014).
8. Elad, M.: Sparse and Redundant Representation. Springer New York Dordrecht Heidelberg London (2010).
9. Zhang, Q., Li, B.: Discriminative K-SVD for dictionary learning in face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2691–2698 (2010).
10. Yang, A.Y., Wright, J., Ma, Y., Sastry, S.S.: Feature selection in face recognition: A sparse representation perspective. Technical Report UCB/EECS-2007-99, EECS Department, University of California, Berkeley (2007).
11. Nikolova, M.: A Variational Approach to Remove Outliers and Impulse Noise. Journal of Mathematical Imaging and Vision 20(1-2), 99–120 (2004).
12. Staal, J., Abrmoff, M.D., Niemeijer, M., Viergever, M.A., Ginneken, B.V.: Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging, 501–509 (2004).
13. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Transactions on Medical Imaging 19, 203–210 (2000).
14. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online Learning for Matrix Factorization and Sparse Coding. Journal of Machine Learning Research 11, 19–60 (2010).
15. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision 40(1), 120–145 (2011).
16. Singh, N.P., Srivastava, R.: Retinal blood vessels segmentation by using Gumbel probability distribution function based matched filter. Computer Methods and Programs in Biomedicine 129, 40–50 (2016).
17. Orlando, J.I., Prokofyeva, E., Blaschko, M.B.: A Discriminatively Trained Fully Connected Conditional Random Field Model for Blood Vessel Segmentation in Fundus Images. IEEE Trans. Biomed. Engineering 64(1), 16–27 (2017).
18. Dasgupta, A., Singh, S.: A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. In: IEEE International Symposium on Biomedical Imaging (ISBI), 248–251 (2017).