A Comparative Study on Feature Selection for Retinal Vessel Segmentation Using FABC

Carmen Alina Lupaşcu[1], Domenico Tegolo[1], and Emanuele Trucco[2]

[1] Dipartimento di Matematica e Applicazioni, Università degli Studi di Palermo, Palermo, Italy
[2] School of Computing, University of Dundee, Dundee, Scotland

Abstract. This paper presents a comparative study on five feature selection heuristics applied to a retinal image database called DRIVE. Features are chosen from a feature vector (encoding local information, but as well information from structures and shapes available in the image) constructed for each pixel in the field of view (FOV) of the image. After selecting the most discriminatory features, an AdaBoost classifier is applied for training. The results of classifications are used to compare the effectiveness of the five feature selection methods.

Keywords: Retinal images, vessel segmentation, AdaBoost classifier, feature selection.

1 Introduction

Automatic vessel segmentation in retinal images is very important, as the retinal vasculature may reveal vascular and nonvascular pathologies.

In the literature supervised methods have been used for vessel segmentation. Pixel classification based on supervised methods require hand-labeled ground truth images for training. Sinthanayothin et al. in [13] classify pixels using a multilayer perceptron neural net, for which the inputs were derived from a principal component analysis (PCA) of the image and edge detection of the first component of PCA. In [12] a simple feature vector is extracted for each pixel from the green plane and then a K-nearest neighbor (kNN) is used to evince the probability of being a vessel pixel. Another supervised method, called primitive-based method, was proposed in [16]. This algorithm is based on the extraction of image ridges (expected to coincide with vessel centerlines) used as primitives for describing linear segments, named line elements. Consequently, each pixel is assigned to the nearest line element to form image patches and then classified using a set of features from the corresponding line and image patch. The feature vectors are classified using a kNN-classifier. The method presented by Soares et al. in [14] produces also a segmentation after a supervised classification. Each image pixel is classified as vessel or non-vessel based on the pixel feature vector, which is composed of the pixel intensity and two-dimensional Gabor wavelet transform responses taken at multiple scales. A Gaussian mixture model (a Bayesian classifier in which each class-conditional probability density function is described as a linear combination of Gaussian functions) classifier is then applied to obtain a final segmentation. Feature selection is applied only by Staal et al. in [16]. The scheme used is the sequential forward selection method. The selection method starts with a null feature set and at each

X. Jiang and N. Petkov (Eds.): CAIP 2009, LNCS 5702, pp. 655–662, 2009.

step, the best feature that satisfies the criterion function (the area under the receiver operating characteristic curve) is added to the feature set. The set with the best performance is chosen, after all features have been included.

Feature selection is used as a preprocessing step to machine learning, because is effective in reducing dimensionality, removing irrelevant data and increasing learning accuracy. Algorithms that perform feature selection can be divided into two categories: the filter model and the wrapper model. The filter model is computationally efficient especially when the number of features is very large, because it doesn't involve any learning algorithm when selecting features. It relies only on general characteristics of the training data, while the wrapper model needs one predetermined learning algorithm in feature selection and uses its performance in order to evaluate and determine which features are selected.

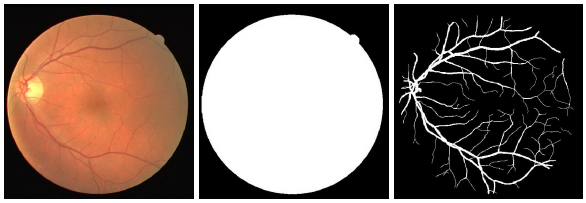
We have developed a new supervised method for retinal vessel segmentation called FABC. The method is based on computing feature vectors for every pixel in the image and training an AdaBoost classifier with manually labeled images. The feature vector is a collection of measurements at different scales taken from the output of filters (the Gaussian and its derivatives up to the 2 order, matched filters and two-dimensional Gabor wavelet transform responses), from the identification of edge and ridge pixels and from other local information which are extracted after computing the Hessian of the image for each pixel. The basic idea is to encode in the feature vector local information (pixel's intensity, Hessian-based measures), spatial properties (the gray-level profile of the cross-section of a vessel can be approximated by a Gaussian curve) and structural information (vessels are geometrical structures which can be seen as tubular). We used an AdaBoost classifier to divide pixels into two classes, i.e., vessels and non-vessel pixels.

In order to analyze the improvement of the computational time, as well as of the accuracy and the performance of the classification, we decided to perform a comparative study on feature selection methods applied as a preprocessing step to the AdaBoost classification of the vessel and non-vessel pixels. In this paper we present five feature selection heuristics, which are evaluating the goodness of features through feature subsets. All five methods are subset search algorithms based on the filter model (as the filter model is computationally less expensive than the wrapper model).

2 Database and Features

2.1 Database

The database we use for testing and evaluating the methods is the public database named DRIVE (Digital Retinal Images for Vessel Extraction). The photographs for the DRIVE database were obtained from a diabetic retinopathy screening program in The Netherlands. Each image has been JPG compressed. The images were acquired using a Canon CR5 non-mydriatic 3CCD camera with a 45 degree field of view (FOV). Each image was captured using 8 bits per color plane at 768 by 584 pixels. The FOV of each image is circular with a diameter of approximately 540 pixels. For this database, the images have been cropped around the FOV. For each image, a mask image is provided that delineates the FOV and also a ground truth segmentation of the vessels (Figure 1). The



(a) (b) (c)

Fig. 1. Retinal image (a), mask image (b) and ground truth segmentation of the vessels (c)

data set includes 40 584×565 fundus images, divided into a training and test set, each containing 20 images. All images are available for download at the web site of the Image Sciences Institute of the University Medical Center Utrecht `http://www.isi.uu.nl/Research/Databases/DRIVE/download.php`.

2.2 Pixel Features

Features are extracted from the green plane of the retinal images, because in the green plane the contrast between vessel and background is higher than in the blue or red plane.

The feature vector consists of the output of filters (items 1 and 2 in the list below), vesselness and ridgeness measures based on eigen-decomposition of the Hessian computed at each image pixel (items 3, 4 and 5), and the output of a two-dimensional Gabor wavelet transform taken at multiple scales (item 6). Moreover the feature vector includes the principal curvatures, the mean curvature and the values of principal directions of the intensity surface computed at each pixel of the green plane image. The value of the root mean square gradient and the intensity within the green plane at each pixel are also included in the feature vector (items 7 and 8). The total number of features composing the feature vector at each image pixel is 41. Four scales are used to detect vessels of different width: $\sqrt{2}$, 2, $2\sqrt{2}$ and 4.

We give below the list of components of the feature vector.

1. Gaussian and its derivatives up to order 2. (features $1^{st} - 24^{th}$ in Table 3)
2. Green channel intensity of each pixel. (feature 25^{th} in Table 3)
3. Multiscale matched filter for vessels using a Gaussian vessel profile. [15] (feature 26^{th} in Table 3)
4. Frangi's vesselness measure. [3] (features $27^{th} - 28^{th}$ in Table 3)
5. Lindeberg's ridge strengths. [9] (features $29^{th} - 31^{st}$ in Table 3)
6. Staal's ridges. [16] (feature 32^{nd} in Table 3)
7. Two-dimensional Gabor wavelet transform response taken at multiple scales. [14] (feature 33^{rd} in Table 3)
8. Values of the principal curvatures (features $34^{th} - 35^{th}$ in Table 3), of the mean curvature (feature 36^{th} in Table 3), of the principal directions (features $37^{th} - 40^{th}$ in Table 3) and of the root mean square gradient (feature 41^{st} in Table 3) of the image.

3 Sample Selection

The training set within the DRIVE database consists of 20 images of size 584×565 pixels, hence for training we had to choose from $6,599,200$ pixels (but we considered only the pixels inside the FOV which are $4,449,836$). Due to the large number of pixels, only $789,914$ pixel samples where randomly chosen to train the classifier, i.e., $789,914/20$ pixels from each image, keeping the same proportion as in the ground-truth image between vessel and non-vessel pixels. The sample size was computed with a Z-test, considering a confidence level of 95% and a margin of error of 10%.

4 Feature Selection Heuristics

Feature selection is applied prior to classification, with the purpose to find a subset of features that optimizes the classification process, in terms of accuracy, performance and computational time.

As described in [6], Correlation-based Feature Selection (CFS) first calculates a matrix of feature-class and feature-feature correlations from training data. Based on these matrices a heuristic for evaluating the merit of a subset of features is computed. The heuristic takes into account the usefulness of individual features and in the same time the level of intercorrelation among them. The hypothesis on which the heuristic is based is: *Good feature subsets contain features highly correlated with the class, yet uncorrelated with each other.*

The following equation (as in [4]) formalizes the heuristic:

$$Merit_S = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}}$$

where $Merit_S$ is the heuristic merit of a feature subset S containing k features, $\overline{r_{cf}}$ is the average feature-class correlation, and $\overline{r_{ff}}$ is the average feature-feature intercorrelation. The numerator gives an indication of how predictive a group of features are; the denominator of how much redundancy there is among them.

4.1 Correlation-Based Feature Selection with Hill-Climbing Search Strategy (Heuristic H1)

In order to find the best subset that has the higher merit without trying all possible subsets, hill-climbing search strategy may be used. The algorithm starts with an empty set of features and generates all possible single feature expansions. The subset with the highest merit is chosen and expanded in the same way by adding single features until all features are added. The subset with the highest merit found during the search will be selected.

4.2 Correlation-Based Feature Selection with Best-First Search Strategy (Heuristic H2)

Usually CFS uses a best-first search strategy. Best-first search is similar to hill-climbing search, with the difference that if expanding a subset the merit doesn't maximize, the search drops back to the next best unexpanded subset and continues from there. CFS uses a stopping criterion of five consecutive fully expanded non-improving subsets. The subset with the highest merit, found in this way, is returned when the search terminates.

4.3 Consistency-Based Feature Selection with CFS (Heuristic H3)

An inconsistency is defined in [2] as two instances having the same feature values but different class labels. As the consistency measure is applied to each feature and not to a feature subset, in order to choose a minimum number of features that separate classes as consistently as the full set of features does, we sort the set of features by the number of inconsistencies in ascending order and apply a modified CFS to this sorted set. We apply the CFS in order to find the best correlated and consistent subset. The algorithm starts with the set containing only the feature with less inconsistencies. For this subset the merit is computed. The subset is expanded by adding one by one the features from the sorted set described above. The subset with the highest merit is returned as the best subset which is consistent and highly correlated.

4.4 Entropy-Based Feature Selection with CFS (Heuristic H4)

The entropy of each feature is defined as $-sum(plog_2(p))$, where p is the histogram counts of all the feature values. The smaller a feature's entropy, the more discriminatory the feature; hence we sort the set of features by the entropy values in ascending order and apply the modified CFS described above. The idea behind the modified CFS is not to decide a priori how many features to choose having low entropy, but to choose those highly correlated.

4.5 MIT Correlation with CFS (Heuristic H5)

The MIT correlation is based on the t-statistics and it is also known as signal-to-noise statistic. As described in [10] and [5], the method starts with a data set S containing m feature vectors: $X^i = (x_1^i, ..., x_n^i)$, where $1 <= i <= m$, m is the number of samples and n is the number of features. Each sample is labeled with $Y \in \{+1, -1\}$ (for classes, such as vessel pixels vs. non-vessel pixels). For each feature x_j, the mean μ_j^+ (resp. μ_j^-) and the standard deviation σ_j^+ (resp. σ_j^-) using only the samples labeled $+1$ (resp. -1) are calculated. A score $T(x_j)$ is then computed as

$$T(x_j) = \frac{|\mu_j^+ - \mu_j^-|}{\sqrt{\frac{(\sigma_j^+)^2}{n_+} + \frac{(\sigma_j^-)^2}{n_-}}},$$

where n_+ (resp. n_-) is the number of samples labeled as $+1$ (resp. -1). Features are then ordered by the score in descending order, as the features with the highest scores will be the most discriminatory features. In the same way as described above, we choose the best subset as the subset with the features highly correlated among, hence we compute the merit with the modified CFS.

5 Experimental Evaluation

The performances of the classification are measured using receiver operating characteristic (ROC) curves. ROC curves are represented by plotting true positive fractions

versus false positive fractions as the discriminating threshold of the AdaBoost algorithm is varied. The true positive fraction (TPF) is determined by dividing the number of pixels correctly classified as vessel pixels (TP) by the total number of vessel pixels in the ground truth segmentation, while the false positive fraction (FPF) is the number of pixels incorrectly classified as vessel pixels (FP) divided by the total number of non-vessel pixels in the ground truth. The axes of the plot are rescaled so the true positives and false positives vary between 0 and 1. The area under the ROC curve (Az) measures discrimination, in our case, is the ability of the classifier to correctly distinguish between vessel and non-vessel pixels. An area of 1 indicates a perfect classification. We compute also the accuracy (ACC), which is degree of veracity of the classification, being the fraction of pixels correctly classified.

6 Experimental Results

As it can be seen in Table 2, after applying the heuristics described in Section 4, the number of features recommended by the first heuristic is 7, while the ones recommended by the second one are 4. The third heuristic selects 21 features, the fourth one 20, while for the fifth a set containing 16 features is doing the best separation between classes.

Table 1 shows the performance of different methods with or w/o feature selection that have been tested on DRIVE database. The performance is measured by the ROC index, defined as the area under the ROC curve and by the accuracy, defined as the fraction of pixels correctly classified. The performances shown here are those reported on www.isi.uu.nl/Research/Databases/DRIVE/.

Table 1. Overview of the performance of different methods. Az indicates the area under the ROC curve and ACC indicates the accuracy.

Segmentation method	Drive database		Segmentation method	Drive database	
	A_z	ACC		A_z	ACC
FABC	0.9560	0.9584	Niemeijer et al.	0.9294	0.9416
(w/o feature selection)			Zana et al.	0.8984	0.9377
Soares et al.	0.9614	0.9466	Jiang et al.	0.9114	0.9212
Human observer	-	0.9473	Martinez et al.	-	0.9181
Staal et al.	0.9520	0.9442	Chaudhuri et al.	0.7878	0.8773

Table 2. Performance of FABC with feature selection (H1, H2, H3, H4, H5) or w/o feature selection (All features). COLUMN 2: the area under the ROC curve A_z. COLUMN 3: the accuracy ACC. COLUMN 4: the computational time of constructing the model used by the AdaBoost classifier.

Heuristic	A_z	ACC	Computational time	Heuristic	A_z	ACC	Computational time
All features	0.9560	0.9584	$\approx 230\ min.$	H3 (21 features)	0.8970	0.9536	$\approx 24\ min.$
H1 (7 features)	0.9482	0.9554	$\approx 46\ min.$	H4 (20 features)	0.8953	0.9532	$\approx 21\ min.$
H2 (4 features)	0.9393	0.9551	$\approx 20\ min.$	H5 (16 features)	0.9544	0.9572	$\approx 85\ min.$

Our method was tested on an Intel(R) Core(TM)2 Duo CPU (3.16 GHz) with 3326 Mb memory. Feature generation for an image from the DRIVE database takes less than 2 minutes, while the classification of its pixels takes less than 5 seconds. The process of learning the AdaBoost model is computationally more expensive. It takes almost 4 hours when using the full set of features, while after feature selection it speeds up a lot as it can be seen in Table 2.

7 Discussion and Conclusion

When selecting features with the consistency-based feature selection, as well as with the entropy-based and MIT correlation feature selection, we don't use an arbitrary number to select top ranked features. In all cases we apply the CFS to strengthen the correlation among features and as well to determine automatically the number of discriminatory features.

We notice from Table 3, that the features that are playing an important discrimination role (the ones that were selected by all heuristics) are: feature 18 (the 2 order derivative of the Gaussian in the y direction at scale $2\sqrt{2}$), feature 26 (the maximum response of a multiscale matched filter using a Gaussian vessel profile) and feature 32 (the one containing information about Staal's ridges).

Even if after feature selection with the proposed heuristics the classification performance doesn't improve, the accuracies achieved outperforms the accuracies of the state-of-the-art approaches (even the one of the second human observer). As it can be seen in Table 2, of the five proposed selection heuristic the one using MIT correlation appeared to be the best one (it approached the performance obtained by the algorithm without feature selection). The performance of the fifth heuristic, as the one of FABC without feature selection, is still competitive to the best performance (the one achieved in [14] by Soares et al.).

Table 3. Features recommended by each heuristic

Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
H1	×	×	×	×	✓	✓	×	×	×	×	×	×	×	×	×	×	×	✓	×	×	×
H2	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	✓	×	×	×
H3	×	×	×	✓	✓	✓	×	×	×	✓	✓	✓	×	×	×	✓	✓	✓	×	×	×
H4	×	×	×	✓	✓	✓	×	×	×	✓	✓	✓	×	×	×	✓	✓	✓	×	×	×
H5	×	×	×	×	✓	✓	×	×	×	×	✓	✓	×	×	×	×	✓	✓	×	×	×

Feature	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	
H1	×	×	×	×	✓	×	×	×	✓	×	✓	✓	×	×	×	×	×	×	×	×	
H2	×	×	×	×	✓	×	×	×	×	×	✓	✓	×	×	×	×	×	×	×	×	
H3	✓	✓	✓	×	✓	×	✓	×	×	×	✓	×	✓	×	✓	✓	✓	✓	✓	×	
H4	✓	✓	✓	×	✓	×	✓	×	×	×	✓	×	✓	×	×	✓	✓	✓	✓	×	
H5	×	×	✓	×	✓	✓	×	×	✓	✓	✓	✓	×	✓	✓	×	×	×	×	✓	

References

1. Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M., Goldbaum, M.: Detection of blood vessels in retinal images using two-dimensional matched filters. IEEE Transactions on Medical Imaging 8(3), 263–269 (1989)
2. Dash, M., Liu, H.: Consistency-based search in feature selection. Artificial Intelligence 151, 155–176 (2003)

3. 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)
4. Ghiselli, E.E.: Theory of psychological measurement. McGraw-Hill, New York (1964)
5. Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., Caligiuri, M., Bloomfield, C.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286, 531–537 (1999)
6. Hall, M.A.: Correlation-based feature selection for discrete and numeric class machine learning. In: Proc. 17th International Conference on Machine Learning, pp. 359–366 (2000)
7. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piece-wise threshold probing of a matched filter response. IEEE Transactions on Medical Imaging 19(3), 203–210 (2000)
8. Jiang, X., Mojon, D.: Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(1), 131–137 (2003)
9. Lindeberg, T.: Edge detection and ridge detection with automatic scale seletion. Int. J. Comp. Vis. 30, 117–156 (1998)
10. Liu, H., Li, J., Wong, L.: A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns. Genome Informatics 13, 51–60 (2002)
11. Martínez-Pérez, M., Hughes, A., Stanton, A., Thom, S., Bharath, A., Parker, K.: Scale-space analysis for the characterisation of retinal blood vessels. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI 1999, pp. 90–97 (1999)
12. Niemeijer, M., Staal, J., van Ginneken, B., Loog, M., Abràmoff, M.: Comparative study of retinal vessel segmentation methods on a new publicly available database. In: SPIE Medical Imaging, vol. 5370, pp. 648–656 (2004)
13. Sinthanayothin, C., Boyce, F., Cook, L., Williamson, H.: Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. Br. J. Ophthalmol. 83, 902–910 (1999)
14. Soares, V.J., Leandro, J.J., Cesar, R.M.J., Jelinek, F.H., Cree, M.J.: Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. IEEE Transactions on Medical Imaging 25(9), 1214–1222 (2006)
15. Sofka, M., Stewart, C.V.: Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures. IEEE Transactions on Medical Imaging 25(12), 1531–1546 (2006)
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 Transactions on Medical Imaging 23(4), 501–509 (2004)
17. Zana, F., Klein, J.: Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. IEEE Transactions on Image Processing 10(7), 1010–1019 (2001)