An LDA-based relative hysteresis classifier with application to segmentation of retinal vessels

Alexandru Paul Condurache, Florian Müller and Alfred Mertins
Institute for Signal Processing, University of Luebeck, Ratzeburger Allee 160, D-23538 Luebeck
alexandru.condurache@isip.uni-luebeck.de

Abstract

In a pattern classification setup, image segmentation is achieved by assigning each pixel to one of two classes: object or background. The special case of vessel segmentation is characterized by a strong disproportion between the number of representatives of each class (i.e. class skew) and also by a strong overlap between classes. These difficulties can be solved using problem-specific knowledge. The proposed hysteresis classification makes use of such knowledge in an efficient way. We describe a novel, supervised, hysteresis-based classification method that we apply to the segmentation of retina photographies. This procedure is fast and achieves results that comparable or even superior to other hysteresis methods and, for the problem of retina vessel segmentation, to known dedicated methods on similar data sets.

1. Introduction

Photographies of the retina showing the vasculature are used, for example, to support medical diagnosis and for intervention planning. To this end, the retinal vessels need to be segmented to compute measures like vessel area and length, vessel width, abnormal branching, and also to provide a localization of vascular structures.

The contrast of vessels in the analyzed images is related to the quantity of blood found therein. Hence, small vessels have a weak contrast. Also differences between vessels and background pixels are localized in the vicinity of the vessel. The background is usually inhomogeneous and can be locally similar to the vessels. Such characteristics make the problem of vessel segmentation difficult, and many solutions have been proposed for it [8]. They can be divided into supervised and unsupervised methods. In many applications, a set of labeled examples is difficult to obtain, and therefore, unsupervised methods are used. However, there are some important applications in which supervised methods are well suited and automatic methods are needed, like e.g., screening for diabetic rethinopathy [9].

The hysteresis classification paradigm [3], [1] makes use of prior knowledge to yield solutions to binary classification problems. For vessel segmentation, it uses the connectivity of vessels. Two classifiers are defined: the first one, called the pessimist, works with a practically zero false positives rate, which with overlapping classes implies a high false negatives rate; the second one, called the optimist, works with a practically zero false negatives rate and a high false positives rate. Then, using the connectivity property of vessels, the pessimist results can be used to select true vessels from among the optimist results.

The hysteresis paradigm can yield both supervised [3], [5] and unsupervised [2], [10] classifiers, for scalar and vectorial inputs, which are all accurate and very fast. In this contribution, we describe a new type of supervised relative hysteresis classifier that we use for the segmentation of retina images with application to screening for diabetic rethinopathy. We obtain results superior to both previous supervised hysteresis methods described in [3], [5] and state-of-the-art methods [6], [7], [11], [12].

2. A novel type of relative hysteresis classifier

Hysteresis classification If the supports of two classes in a binary classification problem overlap strongly but not completely in the original feature space A, then error-free classification is impossible there. If the components of one class do exhibit some type of connectivity in a different feature space B, where there is also no overlap, then the hysteresis paradigm is used to design methods that may achieve error-free classification. Two classifiers working in feature space A (i.e. the pessimist and the optimist), coupled over the connectivity constraint in B, build a *hysteresis classifier*. The pessimist and the optimist are called base classifiers. For image segmentation, A is given by the gray levels of the analyzed image. B is then the 2D space of image coordinates. In this paper the connectivity in B is defined by the eight neighborhood. True objects are considered all optimist object points linked to a pessimist object point by a chain of neighbors.

The absolute and the relative hysteresis classifier In [3] a supervised hysteresis classifier is described, where the pessimist and the optimist are two Fisher's classifiers with parameters $\langle \vec{w}, T_p \rangle$ and $\langle \vec{w}, T_o \rangle$ respectively. Feature vectors of all pixels from all training images built a *pixel-feature vector space*, that was then used to compute the parameters of the hysteresis classifier (i.e. $\langle \vec{w}, T_p, T_o \rangle$). These parameters remain constant for all analyzed images. We call this an absolute hysteresis classifier.

In [5], we proposed a relative hysteresis classifier, where the parameters change from image to image, thus better adapting to the analyzed data and providing better results. The pessimist and the optimist are in this case defined relative to each image. For scalar inputs this can be done by means of percentiles, for vectorial inputs we have introduced the linear-classifier percentile. A linear-classifier percentile is defined by a linear separating surface that selects percentages of the total number of realizations in the sample, and by its position on an axis perpendicular to it.

In this contribution we introduce a new type of relative hysteresis classifier where the input data is first transformed to a line by means of LDA (which will always yield 1D outputs for binary classification problems) and then the percentiles corresponding to the base classifiers are defined in the resulting feature space.

Hysteresis classification and percentiles Next, if not otherwise specified, we assume that vessels are darker than background. The image investigated can be either the original image or a *vessel map*, which represents the result of different vessel enhancement methods applied to the original image.

In [2] it was shown how to compute the base classifiers by hypothesis testing. Then, they are chosen such that the probability of a certain event is very small, i.e., at most equal to the significance. This can also be expressed in terms of quantiles.

For the pessimist, the null hypothesis is that the pixel under investigation belongs to the background class, hence we impose $P(x_b < t_p) = \alpha$, with x_b being a pixel gray level in the background class, t_p a threshold and α the significance. We have then $P(x_b < t_p) = \sum_{i=v_{bmin}}^{t_p} \frac{n_{bi}}{N_b} = \alpha$ with v_{bmin} denoting the minimum gray level on the histogram of the background gray levels, n_{bi} being the number of background pixels with gray level i, and N_b being the total number of background pixels in the image. The value t_p is then the α'th quantile of the histogram of the background.

The histogram of the image is the discrete approximation of the mixture of vessel and background class-conditional probability density functions (*pdfs*). Therefore, t_p is also a quantile of the histogram of the image and can be found via

$$P(x < t_p) = \sum_{i=v_{min}}^{t_p} \frac{n_i}{N} = \alpha_{im} \tag{1}$$

where x is a pixel gray level in the image, v_{min} is the minimum gray level on the histogram, n_i is the number of pixels with gray level i and N denotes the total number of pixels in the image. The threshold t_p is then the α_{im}'th quantile of the histogram of the image, and it should be chosen such that it selects practically *only* vessel pixels.

Similarly, the optimist is computed using the object class-conditional *pdf*. This time we hypothesize that the pixel under consideration is an object pixel. To compute the threshold, we impose again a small significance level β, $P(x_o > t_o) = \sum_{i=t_o}^{v_{omax}} \frac{n_{oi}}{N_o} = \beta$, where x_o is a pixel gray level in the object class, v_{omax} is the maximum gray level on the histogram of the object gray levels, n_{oi} is the number of object pixels with gray level i, and N_o is the total number of object pixels in the image. t_o is some quantile of the histogram of the object gray levels and it is also a quantile of the histogram of the image. It can be found from

$$P(x < t_o) = \sum_{i=v_{min}}^{t_o} \frac{n_i}{N} = \beta_{im} \tag{2}$$

The threshold t_o is then the β_{im}'th quantile of the histogram of the image, and it should be chosen such that it selects practically *all* vessel pixels. For the purpose of hysteresis classification we use percentiles – i.e. 100'th quantiles.

2.1. LDA-based relative hysteresis classifier

The relative hysteresis classifier of [5] uses two linear-classifier percentiles $h_p = \mathbf{b}^T\mathbf{x} + c_p$ and $h_o = \mathbf{b}^T\mathbf{x} + c_o$ as pessimist and optimist respectively. The parameters c_p and c_o represent the positions of the corresponding linear separation surfaces along the axis defined by $\mathbf{b}$ such that they separate certain percentages of the available sample from the rest.

The scalar product $\mathbf{b}^T\mathbf{x}$ can also be seen as applying a transformation $\mathbf{b}$ to the data vector $\mathbf{x}$ that maps it to a scalar value. To compute $\mathbf{b}$ we have made the rather restrictive assumptions that the class-conditional *pdfs* of the object and the background are Gaussian with equal covariance matrices and thus obtained $\mathbf{b} = 2\mathbf{\Sigma}(\mathbf{m}_b - \mathbf{m}_o)$ with $\mathbf{m}_b$ the mean of the background class and $\mathbf{m}_o$ the mean of the object class.

We propose here to use the LDA to compute a mapping from the initial multidimensional feature space to a scalar feature space, where, for the purpose of hysteresis classification we then compute the percentiles corresponding to the base classifiers. As we have a binary classification problem, the LDA will always yield a transformation that maps a multidimensional input to a scalar value.

During LDA, one looks for a transformation such that in the transformed space, the separability criterion: $F = \frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}$, where $\mu_{1,2}$ and $\sigma_{1,2}$ are the means and variances in the transformed space, is optimized. Using a labeled input feature space, the transformation weights are:

$$\mathbf{w} = (\mathbf{m}_1 - \mathbf{m}_2)^T \left(\frac{n_1}{n}\mathbf{\Sigma}_1 + \frac{n_2}{n}\mathbf{\Sigma}_2\right)^{-1} \tag{3}$$

where $\mathbf{m}_{1,2}$ and $\mathbf{\Sigma}_{1,2}$ are the class-conditional means and covariance matrices respectively, $n_{1,2}$ is the number of components in each class and n the total number of components.

The base classifiers are now Fisher classifiers, but in contrast to [3], they are defined in a relative manner as the corresponding thresholds are computed from percentile values set during training. The vector $\mathbf{w}_i$ is computed for each image i in the training set and then $\mathbf{w}$ is computed as the mean over all $\mathbf{w}_i$. Alternatively, $\mathbf{w}$ can be computed in the pixel-feature vector space (see Section 2)

The training To find the two percentiles for the base classifier in the LDA-transformed feature space, we use the *iterative training* procedure from [5]. First the ROC of a percentile-based decision is used to initialize the optimist as that percentile corresponding to the point most distant from the baseline. This ROC is constructed from the fp and correct classification (cc) rates of each percentile from zero to 100. Then a hysteresis ROC is built using the previously established optimist and all possible pessimist classifiers corresponding to the percentiles from zero to 100. The pessimist corresponding to the point that is most distant from the baseline is selected. The procedure is repeated this time for the optimist and so on for a predetermined number of steps or until the base classifiers remain unchanged for two consecutive iterations.

3. Experimental setup and results

We have applied the hysteresis classifier to the segmentation of vessels in images of the retina from two databases that are publicly available: the Utrecht database [12] and the Hoover database [6]. The first contains 40 images, divided into a training and a test set with 20 images each. For the test images there are two different sets of hand-labeled ground-truth images, marked as first and second observer. For the training images, there is just one set of ground-truth images, marked first observer. The Hoover database contains 20 images that we have divided into a test set containing the first ten images and a training set with the rest. The Hoover database contains two sets of hand-labeled ground-truth images, again marked as first and second observer respectively. For both databases, we have used the first-observer set as ground truth in our experiments. Before processing, the images of the Hoover database were cropped to a size of 512×512 pixels.

Feature extraction To generate feature vectors, so-called vessel maps [4] are computed from the green channel of each original image by applying a set of processing steps aimed at improving the separability of vessels and background. Each vessel map, which is an image of the same size as the original vessel image, makes use of other vessel properties to improve the separability. All maps are used jointly during segmentation. For this purpose, feature vectors were formed by stacking the values of the vessel maps for each pixel position to a vector. Originally, five vessel maps were computed, and after feature selection [3] three-dimensional feature vectors are obtained.

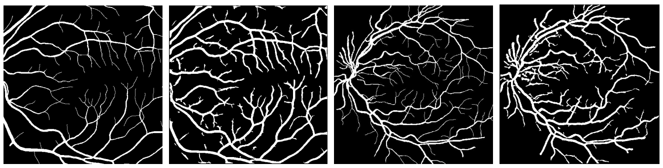
Results We have compared the LDA-based relative hysteresis classifier with the linear-percentile-based relative hysteresis classifier from [5]. Their performance was measured by the area under the ROC (AROC). For a perfect system AROC=100. The corresponding ROC is computed by holding the pessimist fixed and modifying the optimist such that it assigns to the vessel class between 0% and 100% of the available test samples. We have also computed accuracy (A_c), sensitivity (S_e) and specificity (S_p).The parameters of the LDA transform were computed using the pixel-feature space. Table 1 contains the results for the two databases. All results are average values over all test images in the respective database. Some classification examples are shown in Figure 1.

4. Discussion and conclusions

For applications where supervised methods can be used, we have introduced a novel relative hysteresis

		AROC	A_c	S_e	S_p
Utrecht	LDA	97.26	95.16	90.94	95.91
	lin.-prc.	97.13	95.09	90.86	95.8
Hoover	LDA	95.6	89.46	88.37	89.71
	lin.-prc.	95.56	89.44	89.76	89.45

Table 1. Results.



(a) (b) (c) (d)

Figure 1. Ground truth from the Hoover (a) and Utrecht (c) database, and segmentation result (b) and (e) respectively.

classifier for vectorial inputs. Rather than using linear-classifier percentiles as base-classifiers for vectorial inputs, we now map the input to a line by means of LDA and use there the usual percentiles as base-classifiers. We leave thus the restrictive Gaussian assumption made while computing the linear-classifier percentiles and resort instead to improving the separability as computed by F (see Section 2.1).

For comparison, in [11] the following results are obtained by the method proposed there, on the Utrecht database: AROC=96.14 with an accuracy of 94.66%. In [12] their own primitive-based method yields AROC=95.20 with an accuracy of 94.41%, and the verification-based multithreshold probing [7] yields AROC=93.27 and an accuracy of 89.11%. On a dual core Pentium E6700 processor under Matlab, the training time for the new classifier was about one and a half hours on the Utrecht data set, half an hour less than the linear-percentile based relative hysteresis classifier. The time needed to reach a result in the test-phase is 0.8 seconds for an image from the Utrecht database. The time needed to compute the pixel-feature vector set for the same image is six seconds. Therefore a new image is segmented every 6.8 seconds. In comparison, segmentation of one image by the primitive-based methods takes more than 10 minutes.

The relative LDA-based hysteresis classifier yields results that are comparable or slightly better (for the Utrecht database) than the relative classifier introduced in [5], but it does so under less restrictive conditions and it can be trained faster. With respect to some other state-of-the-art vessel segmentation algorithms, it is both better and faster. Hysteresis segmentation can successfully segment objects of inhomogeneous gray-level representation found on an inhomogeneous background, as long as there is a slight difference between object and background at a local level around the object's borders and the supports of the two classes in the pixel feature space do not overlap completely. These conditions should be enforced during feature extraction also.

References

[1] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[2] A. P. Condurache and T. Aach. Vessel segmentation in angiograms using hysteresis thresholding. In *Proceedings of MVA-2005*, pages 269–272, Tsukuba, Japan, May 16–18 2005.

[3] A. P. Condurache and T. Aach. Vessel segmentation in 2D-projection images using a supervised linear hysteresis classifier. In *Proceedings of ICPR-2006*, volume 1, pages 239–243, Hong Kong, China, August 20–24 2006. IEEE.

[4] A. P. Condurache, T. Aach, S. Grzybowski, and H.-G. Machens. Vessel segmentation and analysis in laboratory skin transplant micro-angiogram. In *Proceedings of the 18'th IEEE Symposium on CBMS-2005*, pages 21–26, Dublin, Ireland, June 23–24 2005. IEEE.

[5] A. P. Condurache, A. Mertins, and T. Aach. Supervised, hysteresis-based segmentation of retinal images using the linear-classifier percentile. In *Medizinische Bildverarbeitung und Mustererkennung*, volume P154, Luebeck, Germany, October 2009. GI.

[6] A. Hoover, V. Kouznetzova, and M. Goldbaum. Locating blood vessels in retinal images by picewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19(3):203–210, 2000.

[7] X. Jiang and D. Mojon. Adaptive local thresholding by verification based multithreshold probing with application to vessel detection in retinal images. *IEEE Transactions on Pattern Anaysis and Machine Intelligence*, 25(1):131–137, 2003.

[8] C. Kirbas and F. K. H. Quek. A review of vessel extraction techniques and algorithms. *ACM Comput. Surv.*, 36(2):81–121, 2004.

[9] D. C. Klonoff and D. M. Schwartz. An economic analysis of interventions for diabetes. *Diabetes Care*, 23(3):390–404, 2000.

[10] A. Niemistö, V. Dunmire, O. Yli-Harja, W. Zahng, and I. Shmulevich. Robust quantification of in vitro angiogenesis through image analysis. *IEEE Transactions on Medical Imaging*, 24(4):549–553, 2005.

[11] J. V. B. Soares, J. J. G. Leandro, R. M. C. Jr., H. F. Jelinek, and M. J. Cree. Retinal vessel segmentation using the 2-D Gabor wavelet and subervised classification. *IEEE Transactions on Medical Imaging*, 25(9):1214–1222, 2006.

[12] J. Staal, M. D. Abramoff, M. Niemeijier, M. A. Viergever, and B. van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004.