

Supervised, hysteresis-based segmentation of retinal images using the linear-classifier percentile

Alexandru Paul Condurache¹, Alfred Mertins¹ and Til Aach²

¹ University of Luebeck, Institute for Signal Processing,
Ratzeburger Allee 160, D-23538 Luebeck, Germany

² RWTH Aachen University, Institute of Imaging and Computer Vision,
Sommerfeldstrasse 24, D-52074, Aachen, Germany

Abstract: Image segmentation can be seen as a pattern classification problem, where each pixel is assigned, on the basis of, e.g., its gray level, either to the object or to the background class. In this setup, vessel segmentation is characterized by large class skew, as there are usually far more background pixels than vessel pixels and by weak separability, as there is a strong overlap between the two classes. The proposed hysteresis classification makes use of specific problem-domain knowledge to overcome such difficulties. We describe here a novel, supervised, hysteresis-based classification algorithm that we apply to the segmentation of retina photographs. This procedure is fast and achieves results that are superior to other vessel segmentation 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.

There are several aspects that make vessel segmentation challenging. To name only a few: the contrast of vessels varies with size, small vessels having a weak contrast, the background is usually inhomogeneous and can be locally similar to the vessels.

The vessel-segmentation methods[KQ04] can be divided into supervised and unsupervised, where the supervised methods need a set of labeled examples and work fully automatically. In many applications, a set of labeled examples is difficult to obtain, and therefore, unsupervised methods are often 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 retinopathy [KS00].

Considering a pixel as a point in a feature space, image segmentation is similar to a binary (i.e., two-class) pattern classification problem. The hysteresis classification paradigm [Con08], [Can86] makes explicit use of the prior knowledge about the connectivity of ves-

sels to provide a solution to this binary classification problem. It uses two classifiers: 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 be used to construct both supervised [CA06] and unsupervised [CA05], [NDYH⁺05] classifiers, for scalar and vectorial inputs, which are all accurate and very fast. In this contribution, we describe a new type of supervised hysteresis classifier that we use for the segmentation of retina images with application to screening for diabetic rethinopathy. We obtain results superior to previous supervised hysteresis methods described in [CA06] and some state-of-the-art methods [HKG00], [JM03], [SAN⁺04], [SLJ⁺06].

2 Relative 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. If the connectivity and disjointness in B can be described by some features, then these should be included in A , where a standard classifier can then be used to achieve error-free classification. For image segmentation, A is given by the gray levels of the analyzed image. B is then the 2D space of image coordinates, where object points are neighbors to one another according to a certain neighborhood.

In [CA06] we have described a supervised hysteresis classifier where the pessimist and the optimist were two Fisher's classifiers with parameters $\langle \vec{w}, T_p \rangle$ and $\langle \vec{w}, T_o \rangle$ respectively. \vec{w} defines a transformation from A to a 1D feature space and $T_{p,o}$ represent thresholds in this 1D space. 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 the present paper, we propose 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 now defined relative to each image. For scalar inputs this can be done by means of percentiles, for vectorial inputs we introduce next the linear-classifier percentile.

2.1 The linear-classifier percentile

In [CA05] it was shown how to compute the base classifiers by hypothesis testing. 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. Next, 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.

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. The variable x_b is discrete, and we can estimate

$$P(x_b < t_p) = \sum_{i=v_{bmin}}^{t_p} \frac{n_{bi}}{N_b} = \alpha \quad (1)$$

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 gray levels.

The histogram of the vessel map is the discrete approximation of the mixture of vessel and background class-conditional *pdfs*. Therefore, t_p is also a quantile of the histogram of the vessel map and can be found via

$$P(x < t_p) = \sum_{i=v_{min}}^{t_p} \frac{n_i}{N} = \alpha_{vm} \quad (2)$$

where x is a pixel gray level in the image, v_{min} is the minimum gray level on the histogram of the vessel map, 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_{vm} 'th quantile of the histogram of the vessel map, 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 ,

$$P(x_o > t_o) = \sum_{i=t_o}^{v_{o max}} \frac{n_{oi}}{N_o} = \beta, \quad (3)$$

where x_o is a pixel gray level in the object class, $v_{o max}$ 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 the β 'th quantile of the histogram of the background gray levels and also is a quantile of the histogram of the vessel map. It can be found from

$$P(x < t_o) = \sum_{i=v_{min}}^{t_o} \frac{n_i}{N} = \beta_{vm} \quad (4)$$

The threshold t_o is then the β_{vm} 'th quantile of the histogram of the vessel map, 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.

The k 'th percentile is defined as that value of a 1D random variable which is larger than k percent of all other realizations in the available sample. As we are on the real axis, it is self-evident where the two margins of the sample are, i.e. the maximal and the minimal value. Therefore, the percentile spans the real axis between these two extreme values, selecting in unit steps percentages of the number of realizations in the sample.

In the following, we extend the percentile to vectorial inputs. In particular, we describe the 2D case, but a generalization to ND is trivial. In the 2D case, one should first establish the margins of the sample. In the end we actually look for the axis of largest separability. One should also define a way to select percentages of the total number of realizations in the sample. For this second purpose, we need a type of separating surface. A “linear-classifier” percentile is obtained when this separating surface is a line. Thus, a linear-classifier percentile is defined by a linear separating surface and by its position on an axis perpendicular to it, i.e., the axis of largest separability. A linear separating surface

$$h_l(\vec{x}) = \vec{b}^T \vec{x} + c = 0 \quad (5)$$

is defined in 2D by the vector of weights $\vec{b} = [b_1, b_2]^T$ and the position by c . By modifying c , the separating surface is moved over certain distances on the axis defined by \vec{b} , such that it selects percentages of the available sample in unit steps. We need to define also a direction on this axis, i.e., an orientation of the separating surface. We define this from the mean of the object class towards the mean of the background class, which is equivalent in the 1D case to considering objects dark (see Figure 1 (a)). This equivalence is ensured during feature extraction. Therefore, the first linear-classifier percentile separates one percent of the data from the rest, and usually most of this data will belong to the object. Then, the pessimist and the optimist are chosen from the set of decision surfaces given by the linear-classifier percentiles from zero to 100 (see also Figure 1 (b)). \vec{b} encodes both the axis of largest separability and the orientation of the separating surface.

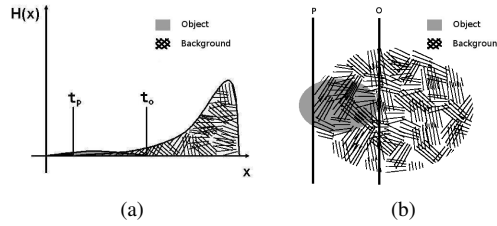


Figure 1: Schematic representation of an object-map histogram with the pessimist and optimist thresholds, t_p and t_o respectively (a) and schematic representation of a 2D pixel feature-space with the optimist and pessimist classifiers, O and P respectively (b).

2.2 Training of the relative classifier

We differentiate between training for scalar and for vectorial inputs. For vectorial inputs we need additionally to compute \vec{b} , which is the same for both classifiers. The two percentiles of the hysteresis classifier are then found using the Receiver Operating Characteristic (ROC), as the ROC-analysis is well suited to deal with class skew [SLT06]. In principle, the *ROC training* of a hysteresis classifier works the following way: we define a set of percentiles/linear-classifier percentiles, such that the corresponding separating surfaces span the entire support of each training-set image, and then we choose two percentiles such that performance of the hysteresis classifier over the entire training set is optimal, as measured by a performance measure.

For each image i in the training set, \vec{b}_i is computed by

$$\vec{b}_i = 2(\mathbf{K}_2^{-1}\vec{m}_2 - \mathbf{K}_1^{-1}\vec{m}_1) \quad (6)$$

assuming the class-conditional *pdfs* are Gaussian, with \mathbf{K}_i being the covariance matrix and \vec{m}_j denoting the mean vector for class ω_j , $j = 1, 2$, where ω_2 is the background class. In our experiments we set $\mathbf{K}_1 = \mathbf{K}_2 = \mathbf{K}$, with $\mathbf{K} = \arg \min_{\mathbf{K}_i} |\mathbf{K}_i|$, $i = 1, 2$, where $|\cdot|$ denotes the determinant of a matrix. The matrices are estimated parametrically from the labeled training data. Then \vec{b} is computed as the mean orientation over all examples in the training set

$$\vec{b} = \frac{1}{N} \sum_{i=1}^N \vec{b}_i \quad (7)$$

To find the two percentiles and hence – in the vectorial case – to set the parameter c relative to each analyzed image, different training methods can be imagined. Here we propose the following *iterative training* procedure: the training data is investigated several times to established together the pessimist and the optimist. The algorithm iterates alternately for the pessimist and the optimist. The performance of the classifier during training is judged by the distance to the baseline linking the ROC points corresponding to a false positives rate (*fp*) of 0% and of 100%. First the ROC of a percentile is used to initialize the optimist as that percentile corresponding to the point most distant from the baseline. The 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. A flow-chart of the training is shown in Figure 2.

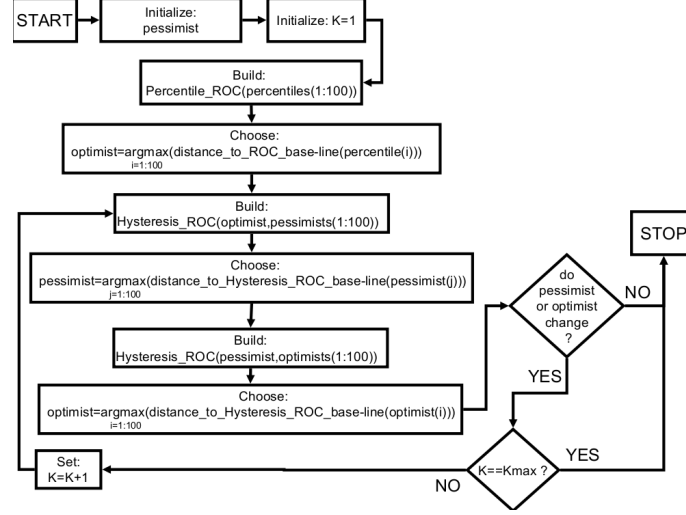


Figure 2: Training of the hysteresis classifier.

3 Experimental setup and results

We have applied the hysteresis classifier to the segmentation of vessels in images of the retina. For this purpose we have used a dataset that is publicly available: the Hoover database [HKG00]. The Hoover database contains 20 images. We have divided these images ourselves into a training and a test set. The test set contains the first ten images and the training set the rest. The Hoover database contains two sets of hand-labeled ground-truth images, again marked as first and second observer respectively. We have used the first-observer set as ground truth in our experiments. The images were cropped to a size of 512×512 pixels from 700×605 pixels.

3.1 Feature extraction

To generate feature vectors, so-called vessel maps, which are images of the same size as the original vessel image, are computed from the original image through applying a set of processing steps aimed at improving the separability of vessels and background [CAGM05c]. Each vessel map makes use of other vessel properties to improve the separability when all maps are used jointly during segmentation. Feature vectors were formed by stacking the values of the vessel maps for each pixel position to a vector [CAGM05a]. Originally, five vessel maps were computed, and after feature selection [CA06] three of them were retained to form three-dimensional feature vectors. The vessel maps that were used to generate this vector were as follows: (1) the first eigenvalue of the Hessian matrix, which is sensitive to elongated structures, (2) the result of the analysis of the eigenvalues of the Hessian matrix in a multiscale approach, (3) the result of a linear filtering procedure,

	AROC	acc.	sens.	spec.
absolute	94.6	89.34	87.15	89.59
relative	95.5	89.48	89.77	89.45

Table 1: Results achieved on Hoover database.

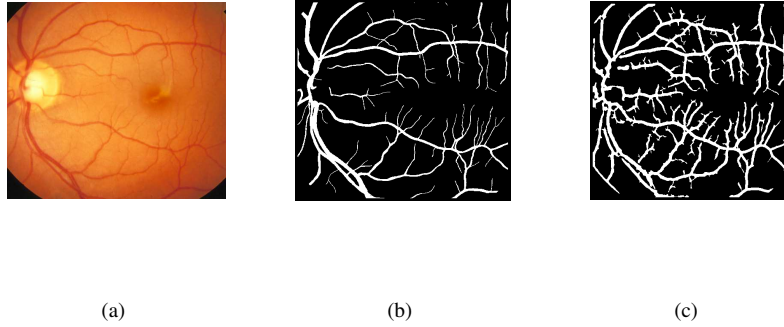


Figure 3: Image of the Hoover database (a), the corresponding manually labeled ground truth (b) and segmentation result achieved by the hysteresis classifier (c).

implemented via a multiresolution analysis based on the Laplacian pyramid [CAGM05b]. To ensure the correct orientation of the separating surface, each vessel map was re-scaled – if needed – such that the vessels are darker than the background.

3.2 Results

We have compared the relative hysteresis classifier introduced here with the absolute hysteresis classifier from [CA06]. Their performance was measured by the area under the ROC (AROC). The corresponding ROC is computed by fixing the pessimist 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 (acc.), sensitivity (sens.) and specificity (spec.). 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 3.

On a dual core Pentium E6700 processor under Matlab, the training time for a relative classifier was about one hour on the Hoover data set. A new image is segmented every 6.8 seconds, including the computation of the feature vector. For comparison, segmentation of one image of the same size by the primitive-based methods [SAN⁺04] takes more than 10 minutes.

4 Discussion and conclusions

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. The relative hysteresis classifier was found to be fast and accurate, being slightly better than other hysteresis methods and a lot faster than some other state of the art methods for the problem of retinal-vessel segmentation.

For applications where supervised methods are applicable, we have introduced a novel relative hysteresis classifier. In contrast to previous absolute supervised hysteresis classifiers, now, the set of parameters (like $\langle \vec{b}, c \rangle$ for vectorial inputs – see equation (5)) changes from image to image. This way the algorithm adapts to the analyzed data. This adaptation is the reason why relative hysteresis methods outperform absolute ones. This was made possible by the usage of percentiles for the base classifiers. We have shown how to extend the notion of percentile to vectorial inputs for hysteresis classification by means of the linear-classifier percentile.

Even though we have used the hysteresis classifier to segment vessel images, we believe that it represents a more general image-segmentation tool that can be used as well for other applications afflicted by large class skew and overlap.

References

- [CA05] 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.
- [CA06] 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.
- [CAGM05a] A. P. Condurache, T. Aach, S. Grzybowski, and H.-G. Machens. Imaging and analysis of angiogenesis for skin transplantation by microangiography. In *Proceedings of ICIP-2005*, pages II/1250 – II/1253 (also on CD-ROM: ISBN 07803 9135–7), Genoa, Italy, September 11–14 2005. IEEE.
- [CAGM05b] 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.
- [CAGM05c] A. P. Condurache, T. Aach, S. Grzybowski, and H.-G. Machens. Vessel Segmentation for Angiographic Enhancement and Analysis. In *Proceedings. BVM*, pages 173–177, 2005.
- [Can86] J. Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.
- [Con08] A. P. Condurache. *Cardiovascular biomedical image analysis: methods and applications*. GCA, Waabs, 2008.

- [HKG00] A. Hoover, V. Kouznetsova, 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.
- [JM03] 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 Analysis and Machine Intelligence*, 25(1):131–137, 2003.
- [KQ04] C. Kirbas and F. K. H. Quek. A review of vessel extraction techniques and algorithms. *ACM Comput. Surv.*, 36(2):81–121, 2004.
- [KS00] D. C. Klonoff and D. M. Schwartz. An economic analysis of interventions for diabetes. *Diabetes Care*, 23(3):390–404, 2000.
- [NDYH⁺05] 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.
- [SAN⁺04] J. Staal, M. D. Abramoff, M. Niemeijer, 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.
- [SLJ⁺06] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar Jr., H. F. Jelinek, and M. J. Cree. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Transactions on Medical Imaging*, 25(9):1214–1222, 2006.
- [SLT06] M. Stäger, P. Lukowicz, and G. Tröster. Dealing with class skew in context recognition. In *Proceedings of IWSAWC-2006*, Lisbon, Portugal, July 4–5 2006.