Refining Coarse Manual Segmentations with Stable Probability Regions

Lauri Laaksonen[1], Joni Herttuainen[1], Hannu Uusitalo[23], and Lasse Lensu[1]

[1]Machine Vision and Pattern Recognition Laboratory
Department of Mathematics and Physics
LUT School of Technology
Lappeenranta University of Technology
PO Box 20, FI-53851 Lappeenranta, Finland
{firstname.lastname}@lut.fi
[2]Department of Ophthalmology, SILK
University of Tampere, Finland
[3]Tauh Eye Center
Tampere University Hospital, Finland

Abstract. Most feature-based lesion detection and computer-aided diagnosis methods for medical images require representative data of each region of interest (ROI) for parameter selection. Furthermore, the spatial accuracy of the segmentation of the ROIs from the background can significantly affect certain image features extracted from the ROIs. However, requiring spatially accurate manual segmentations of the ROIs to be used as the ground truth is infeasible for large image sets due to the amount of manual work involved. To relax the requirement of spatial accuracy and to enable spatial refinement of coarse manual segmentations to have more representative feature data, a method based on color information and maximally stable extremal regions of lesion likelihoods is presented. The proposed method is quantitatively compared to several segmentation approaches by using a challenging set of retinal images with spatially accurate ground truth of exudates. The experiments show that the proposed method produces good results measured as Dice coefficients between the refined segmentation and ground truth.

1 Introduction

Modern medical imaging modalities enable efficient diagnosis, monitoring and documentation of various diseases. In the context of eye health, the diagnosis and monitoring of eye diseases such as age-related macular degeneration (AMD), diabetic retinopathy (DR) and glaucoma based on retinal imaging is standard practice. Widening the screening programs for diagnosing these diseases produces increasing amounts of data, making the manual diagnosis work a bottleneck. As a remedy, computer aided diagnostic and segmentation tools become increasingly important in reducing the work load of medical experts, and in decreasing the variation in diagnostics.

A number of different approaches have been proposed for the automatic and semi-automatic image segmentation and detection of lesions in the retinal images, see, e.g. [14]. To properly set the method parameters, training data consisting of a set of images and ground truth information on the ROIs such as segmented lesions is common practice. To get appropriate ground truth for developing and evaluating image segmentation and lesion detection methods, a common solution is to ask a group of experts to perform manual or semi-automatic segmentation of the ROIs, that is, the physiological details or lesions.

However, manual segmentation of a large set of images is difficult due to the time-consuming and tedious nature of producing spatially accurate markings on the images. Consequently, some retinal image databases contain approximate manual segmentations or delineations of the lesions. Especially in the case where multiple small lesions of the same type, such as exudates, are spread over a region, the delineated area may consist predominantly of the background (an example is shown in Figure 1).
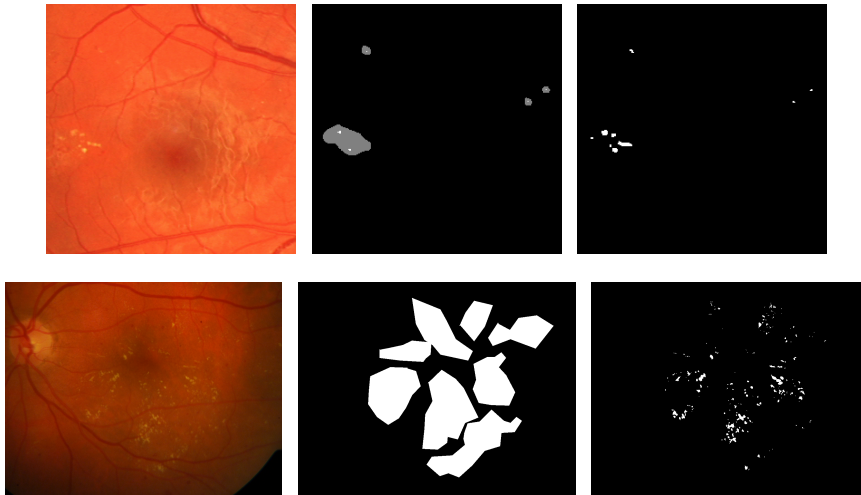

Fig. 1: Examples of retinal images, manual segmentations and the refinement task. Top row from left to right: RGB image, coarse segmentation of exudates (white pixels mark the representative regions) and spatially accurate ground truth. Bottom row (refinement example by using DiaRetDB1 [7]): RGB image, coarse segmentation of exudates and the refined segmentation result.

The parameter tuning of methods making use of image features relies on the representativeness of the ground truth information, which is affected by the accuracy of the ground truth. Requiring spatially accurate manual segmentations of the ROIs is infeasible for large image sets as stated above. On the other hand, spatially inaccurate ground truth can be used for method development

and image-wise performance evaluation of segmentation or lesion detection, but it affects certain image features, can have a significant effect on segmentation results, and makes pixel-wise evaluation of the segmentation or detection results infeasible. This motivates automatic spatial refinement of manual segmentations to improve the representativeness of the coarse manual segmentation data either at the collection stage (semi-automatic segmentation), or as a post-processing step.

In this paper, a method utilizing the color feature and maximally stable extremal regions of lesion likelihoods is presented for the spatial refinement of manual segmentations by experts. The proposed method is compared to a set of methods by using the non-public Bristol retinal image dataset (BristolDB) containing spatially accurate segmentations of exudates. Exudates being yellowish lesions typically with a low contrast to the background and without clear borders make the segmentation task a challenging one.

2   Related work

Various approaches have been proposed for segmenting regions indicative of pathologies and other ROIs in retinal images (e.g., [14]). This section presents a short review of different approaches for automatic and semi-automatic segmentation methods, with the emphasis on retinal image segmentation.

Jafaar et al. [5] segment images with logical intersection of coarse segmentation based on local variation, and the results of adaptive thresholding based on combination of pure splitting and histogram-based thresholding. The coarse segmentation assumes that hard exudates have clear boundaries, and is based on the automatic thresholding of a standard deviation image using Otsu's method, followed by classification of features such as major axis length, minor axis length, area and solidity.

Reza et al. [12] applied watershed transform for bright lesion detection. They used average filtering, that blends the small objects with low intensity variations into the background, and contrast stretching transformation as preprocessing procedures. Afterwards, bright parts of the image were isolated by thresholding. Morphological opening of resulting binary image and the result of extended maxima operator applied to the initial image gave regions with high intensity values, that were used for obtaining markers in subsequent watershed transform.

Osareh et al. [10] segment candidate regions from fuzzy c-means clustering (FCM) using a bank of Gabor filters and identify the best feature subset with genetic optimization. The extracted regions are classified by a multilayer perceptron neural network (NN). Xu and Luo [16] propose a combination of stationary wavelet transform and gray-level co-occurrence matrix as a feature for pixel-level classification by Support vector machine (SVM).

In addition to medical image processing, various segmentation approaches have been applied in numerous other fields. Kakumanu et al. [6] present a review of method for modelling and detection of skin-color. The review includes the use

of different color spaces, modelling with Gaussians and Gaussian mixtures, and classification with naïve Bayes classifier.

Donoser and Bischof [13] use color likelihood maps with modified maximally stable extremal regions (MSER) tracking for robustly tracking a hand through a sequence of images. The likelihood map is based on Kullback-Leibler distance between Gaussian distributions fit to the neighbourhood of a pixel and the background distribution.

Graph-cuts (Graph) based segmentation methods [2] view labeled image areas as nodes of a connected graph, and rearrange the labels corresponding to a minimum cut on the graph.

Ning et al. [9] present an interactive image segmentation method. Unlabelled regions are merged with initial regions from mean shift segmentation based on similarity (color histogram distance) until all regions have been labelled.

GrowCut algorithm (GrowCut) by Vezhnevets and Konouchine [15] is a cellular automaton based semi-automatic segmentation method. The image is labeled by interaction of cellular automata containing different labels, mimicking the growth and competition of bacteria.

Many of the proposed semi-automatic segmentation methods can be considered as suitable for the refinement of manual segmentations. A selection of approaches are evaluated in Section 4.

3   Region refinement through stable probability regions

The proposed method for the refinement of manual segmentations is based on stable regions of a naïve Bayesian posterior probability map of lesion presence following the approach in [13]. Empirical probability density functions (PDFs) are fit to the color values of the background and lesions by kernel smoothing density estimate using a Gaussian kernel. The data samples representing the lesions arise from manually made representative regions.

A probability map of each pixel belonging to the lesion is calculated as $p_{lesion}(1 - p_{bg})$ where p is the Bayesian posterior probability of the pixel color value in the lesion and the background model, respectively. The lesion mask is produced based on the MSER [8] of the resulting probability map. MSER, can be seen as a systematic search for regions where local binarization stays stable over a wide range of thresholds. An algorithmic representation of the method is shown in Algorithm 1.

4   Experiments and results

The performance of the segmentation methods was evaluated on the BristolDB, a non-public dataset collected by Bristol Eye Hospital, consisting of 107 RGB-images of subjects with diabetic retinopathy and healthy control subjects. Exudates, when present, were marked with pixel-level accuracy by an expert. The test set with coarse segmentations was produced by collecting coarse manual

Input: RGB image I, coarse segment set s, representative region set R
Output: Refined segmentation binary mask M_{ref}
Initialize $M_{ref} \leftarrow 0$;
foreach *segment* s_i **do**
  Fit KDE G_{bg} to boundary color values of s_i;
  Fit KDE G_{lesion} to color values of region R_i;
  foreach *pixel color value* v_j *of* s_i **do**
   | $pmap_j \leftarrow P(v_j|G_{lesion})(1 - P(v_j|G_{bg}))$;
  end
  $M_{ref}(MSER(pmap)) \leftarrow 1$;
end

Algorithm 1: Stable probability regions based region refinement: kernel density estimate (KDE), maximally stable extremal regions (MSER).

segmentations of exudate regions and representative regions (foreground for the segmentation methods) for each segment. Possible representative region pixels outside the spatially accurate ground truth were removed.

To implement the method comparison, the following methods were used: binary decision tree (BDT) based classification (with Gini's diversity index optimization criterion) [4], FCM clustering [1,3], segmentation based on Graph-cuts [2], maximal similarity region merging (MSRM) [9], GrowCut [15], and the automatic thresholding by Otsu [11].

The performance of the methods was evaluated by using the Dice similarity criterion between the ground truth and the exudate mask refined by a given method, together with the standard detection performance metrics.

The method evaluation results are shown in Table 1. Otsu thresholding performed well when there was a clear and uniform intensity difference between the lesion and the background. However, its very large variation over the whole image set made the method unreliable for segmentation refinement. While Otsu achieved the highest true-positive rate (TPR), it produced a very high number of false positive detections. Poor performance of MSRM was mostly due to the inability of mean shift segmentation to provide reasonable regions in the initial segmentation. BDT segmentation, GrowCut and the proposed method outperformed the other evaluated methods by a large margin.

A boxplot of the Dice coefficients achieved by the methods is shown in Fig. 2. It is clear that the best methods are BDT and Bayesian probability regions (NB), and their result is statistically different from the others. The latter method has a slightly larger variation in the results and some problematic images cause outliers for both methods. The differences of the best methods are elaborated more in Fig. 3 where the resulting Dice coefficients are sorted in increasing order. The differences emphasize that BDT produces fewer bad segmentations than NB, but after the set of weakest segmentations (one third of the image set in size), NB consistently produces higher Dice coefficient values.

The proposed method performed well with most images of the set. The majority of the differences between the ground truth and refined mask were false

Table 1: Performance metrics of the methods binary decision tree (BDT), fuzzy c-means clustering (FCM), GrowCut algorithm (GrowCut), maximal similarity region merging (MSRM), Bayesian probability regions (NB) and Otsu thresholding (Otsu). The metrics are as follows: Dice similarity coefficient, true-positive rate (TPR), positive predictive value (PPV).

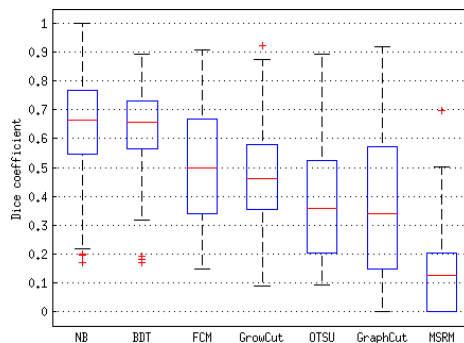|	Dice		TPR		PPV	
	Median	Std	Median	Std	Median	Std
NB	0.6634	0.1753	0.6213	0.2318	0.7821	0.1357
BDT	0.6577	0.1444	0.6617	0.1915	0.7188	0.1738
FCM	0.4975	0.1897	0.6674	0.1582	0.4413	0.2823
GrowCut	0.4622	0.1868	0.6101	0.1961	0.4251	0.2184
OTSU	0.3576	0.2027	0.861	0.1648	0.2301	0.2090
GraphCut	0.3409	0.2425	0.5957	0.2442	0.2287	0.2782
MSRM	0.1242	0.1411	0.1636	0.2533	0.1030	0.2206

Fig. 2: Similarity of refined regions and ground truth. The methods are as follows: binary decision tree (BDT), fuzzy c-means clustering (FCM), graph-cuts (Graph), GrowCut algorithm (GrowCut), maximal similarity region merging (MSRM), Bayesian probability regions (NB), Otsu thresholding (Otsu). The red lines mark the medians, the boxes represent the 25th and 75th percentiles, the whiskers extend to inlier extrema, and the individual markers represent the outliers.

negatives. Therefore, the refined segments can be considered to provide more robust training data also with problematic images. Examples of refinement results using the proposed method are shown in Figure 1f.

5   Discussion and conclusion

A method based on computing the Bayesian probability of a pixel color value in the foreground distribution and extracting MSER from the likelihood maps to refine expert manual segmentations was presented. The proposed method was
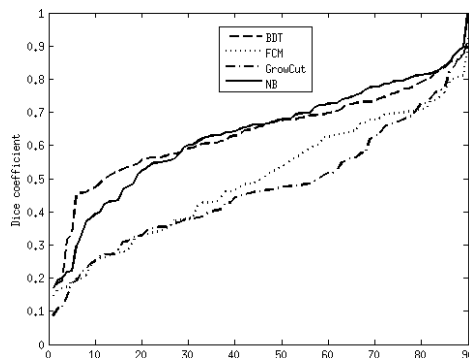
Fig. 3: Dice coefficients of the best performing methods, binary decision tree (BDT), fuzzy c-means clustering (FCM), GrowCut algorithm (GrowCut) and Bayesian probability regions (NB), sorted in ascending order.

quantitively compared with other segmentation methods using BristolDB where accurate ground truth for exudates is available.

While the median results of BDT and the proposed method were very close, NB repeatedly provided higher Dice coeffiecients. However, the larger variation in the results decreased the median Dice value for the method. A plausible reason for larger Dice variation with NB arises from the fact that for some larger segments, the data contains representative regions of only a few of many small lesions, resulting in Gaussian mixtures too limited to represent all the exudates within the segment. Nevertheless, the results show that the proposed method has good overall performance in refining the synthetic inaccurate annotations for the BristolDB retinal image database.

In preliminary tests on annotations in DiaRetDB1 diabetic retinopathy database (DiaRetDB1) [7], the performance of the proposed method was promising as can be seen in Figure 1. The initial results suggest that, for lesions for which color is a good discriminating feature such as haemorrhages and exudates, the proposed method is able to considerably improve the coarse manual data.

When the performance is validated by further tests, the proposed method can be applied as a part of a semi-automatic segmentation tool to facilitate the annotation work of medical experts and, under expert supervision, refinement of existing annotations in medical image databases.

Acknowledgements

References

1. Bezdek, J.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press (1981)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(12), 1222–1239 (2001)
3. Chiu, S.: Fuzzy model identification based on cluster estimation. Journal of Intelligent & Fuzzy Systems 2(3), 267–278 (1994)
4. Duda, R., Hart, P., Stork, D.: Pattern Classification. John Wiley & Sons (2001)
5. Jafaar, H., Nandi, A., Al-Nuaimy, W.: Detection of exudates in retinal images using pure splitting technique. In: IEEE Conference on Engineering in Medicine and Biology Society. pp. 6745–6748. Buenos Aires, Argentina (August 2010)
6. Kakumanu, P., Makrogiannis, S., Bourbakis, N.: A survey of skin-color modeling and detection methods. Pattern Recognition 40(3), 1106–1122 (2007)
7. Kauppi, T., Kämäräinen, J.K., Lensu, L., Kalesnykiene, V., Sorri, I., Uusitalo, H., Kälviäinen, H.: Constructing benchmark databases and protocols for medical image analysis: Diabetic retinopathy. Computational and Mathematical Methods in Medicine 2013, 1–15 (2013)
8. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing 22(10), 761–767 (2002)
9. Ning, J., Zhang, L., Zhang, D., Wu, C.: Interactive image segmentation by maximal similarity based region merging. Pattern Recognition 43(2), 445–456 (2010)
10. Osareh, A., Shadgar, B., Markham, R.: A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images. IEEE Transactions on Information Technology in Biomedicine 13(4), 535–545 (2009)
11. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9(1), 62–66 (1979)
12. Reza, A., Eswaran, C., Hati, S.: Automatic tracing of optic disc and exudates from color fundus images using fixed and variable thresholds. Journal of Medical Systems 33(1), 73–80 (2009)
13. Tracking, R.T.A.B.H.: M. donoser and h. bischof. In: International Conference on Pattern Recognition. pp. 1–4. Tampa, Florida (December 2008)
14. Trucco, E., Ruggeri, A., Karnowski, T., Giancarlo, L., Chaum, E., Hubschman, J.P., Al-Diri, B., Cheung, C.Y., Wong, D., Abramoff, M., Lim, G., Kumar, D., Burlina, P., Bressler, N.M., Jelinek, H., Maiaudeau, F., Quellec, G., MacGillivray, T., Dhillon, B.: Validating retinal fundus image analysis algorithms: Issues and a proposal. Investigative Ophthalmology & Visual Science pp. 3546–3559 (2013)
15. Vezhenevets, V., Konouchine, V.: Growcut: Interactive multi-label nd image segmentation by cellular automata. In: Proceedings of Graphicon. pp. 150–156. Novosibirsk, Russia (June 2005)
16. Xu, L., Luo, S.: Support vector machine based method for identifying hard exudates in retinal images. In: IEEE Youth Conference on Information, Computing and Images. pp. 138–141. Beijing, China (September 2009)