Single and Multiple Object Detection Problems in Medical Image Analyses

Doktori (Ph.D.) értekezés

HARANGI BALÁZS
TÉMAVEZETŐ: DR. HAJDU ANDRÁS

DEBRECENI EGYETEM
TERMÉSZETTUDOMÁNYI DOKTORI TANÁCS
INFORMATIKAI TUDOMÁNYOK DOKTORI ISKOLA
DEBRECEN, 2015.

Single and Multiple Object Detection Problems in Medical Image Analyses

Doktori (Ph.D.) értekezés

HARANGI BALÁZS
TÉMAVEZETŐ: DR. HAJDU ANDRÁS

DEBRECENI EGYETEM
TERMÉSZETTUDOMÁNYI DOKTORI TANÁCS
INFORMATIKAI TUDOMÁNYOK DOKTORI ISKOLA
DEBRECEN, 2015.

Ezen értekezést a Debreceni Egyetem Természettudományi Doktori Tanács Informatikai Tudományok Doktori Iskola Diszkrét matematika, képfeldolgozás és komputergeometria programja keretében készítettem a Debreceni Egyetem természettudományi doktori (Ph.D.) fokozatának elnyerése céljából.
Debrecen, 2015.

.....
a jelölt aláírása

Tanúsítom, hogy Harangi Balázs doktorjelölt 2010 - 2013 között a fent megnevezett Doktori Iskola Diszkrét matematika, képfeldolgozás és komputergeometria programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Az értekezés elfogadását javasolom.
Debrecen, 2015.

.....
a témavezető aláírása

Single and Multiple Object Detection Problems in Medical Image Analyses

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében
az informatikai tudományágban
Írta: Harangi Balázs okleveles programtervező matematikus
Készült a Debreceni Egyetem Informatikai Tudományok Doktori Iskolája, Diszkrét matematika, képfeldolgozás és komputergeometria programja keretében

Témavezető: Dr. Hajdu András

A doktori szigorlati bizottság:

elnök: Dr. Kruppa András .....
tagok: Dr. Palágyi Kálmán .....
Dr. Fazekas Gábor .....

A doktori szigorlat időpontja: 2013. október 25.

Az értekezés bírálói:

....
....

A bírálóbizottság:

elnök: ....
tagok: ....
....
....
....

Az értekezés védésének időpontja:

Contents

List of Figures

List of Tables

Abstract

In our days, diabetic retinopathy (DR) is the most common cause of blindness in the developed countries. In this PhD thesis, which is meant to be an answer to the question how we can improve the accuracy of screening this disease which affects a huge segment of the society, two basic components of an automated DR screening system are introduced. The first one is capable of locating the optic disc (OD) and the second one can detect exudates in the fundus images.

Exudates are one of the first signs of DR, which appear on the fundus of the patient's eyes. Thus, accurate and reliable detection of the exudates is an important part of a state-of-the-art computer aided diagnostic (CAD) system. We developed a novel algorithm which can detect this type of lesions with high accuracy. The basic idea is the following: a grayscale morphology-based approach extracts each of the possible exudate regions as a candidate extractor, then an active contour-based method determines the exact contours of these candidates. Finally, an optimally adjusted classifier selects the true exudates using region-wise features.

As exudates appear in retinal images as bright patches, it is highly recommended that the OD should be localized and masked out before the detection of exudates would be started, as exudates are similar to the OD concerning color and shape characteristics. Thus, our aim is to determine the location of the OD with high accuracy. For this purpose, we propose an ensemble of individual OD detectors to improve their precision and show how the accuracy increases by using more information in order to localize the OD.

1 Introduction

The retina is a delicate light sensitive lining located at the back of our eyeball (see Figure 1) that "takes pictures" and sends the images to the brain. Long-term diabetes affects the eyes, resulting in a disease called diabetic retinopathy (DR). If DR remains undiagnosed or is treated inappropriately, it can lead to loss of vision. Approximately 387 million people were registered worldwide living with diabetes in 2014 [1]. This huge number is the reason why DR is the most common cause of blindness.

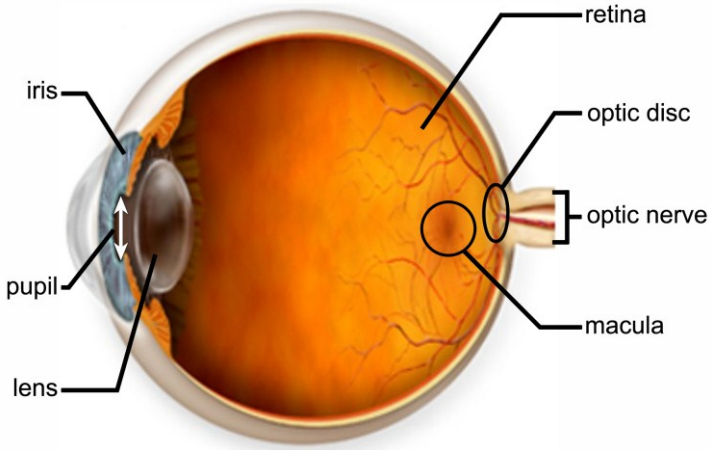

Figure 1: Anatomical parts of the human eye.
Source: http://www.mastereyeassociates.com/presbyopia

According to the estimations [2] as the number of the patients with diabetes has grown rapidly in the last few years and as their number is expected to rise in the future, the estimated increase will be 205 million people till 2035. However, there are appropriate ways of treatment to slow down the deterioration of the eye-sight if the signs of this disease are recognized at an early stage. The fluorescein angiography and retinal

photography are the most commonly used methods to take images of the eyes, which provide ophthalmologists the possibility to monitor the progression of the disease and to make a decision for an appropriate treatment. Unfortunately, 77% of the people with diabetes live in low- and middle-income countries where 53% of the cases remain undiagnosed due to the fact that the medical staff is overloaded [1]. Thus, an automatic DR screening system would have a great importance mainly in the developing countries.

The blood vessels that provide nourishment to the retina, in case of a person with diabetes, may weaken and leak, forming small, dot-like hemorrhages (see Figure 2). These leaking vessels often cause swelling or edema in the retina and so they cause eyesight deterioration. Exudates come into being when fluid exudes from tissue because of its injured capillaries. Since the fluid contains protein, cellular debris and white blood cells, exudates appear as yellowish, bright patches in the retinal background (see also Figure 2). A CAD system is useful if it is capable of detecting these first signs of DR.

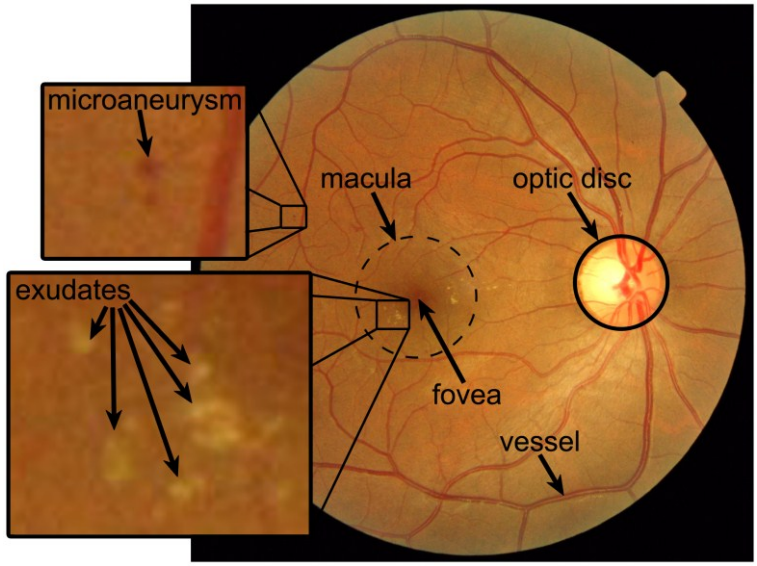

Figure 2: Anatomical parts of the retina with signs of DR.

The automated process of analyzing fundus image can be performed through the following steps: preprocessing, localization and segmentation of the optic disc (OD), segmentation of the retinal vasculature, localization of the macula and fovea and detection the signs of DR. So, an important prerequisite for automation is the accurate method of localizing the OD. The OD is a circular shaped anatomical structure with a bright appearance. It is the location where the optic nerve enters the eye. If the position and the radius of the OD are detected correctly, then they can be used as references for approximating other anatomical parts e.g., the macula and the fovea, as it is proposed in [3]. Furthermore, masking out the OD region in the image is highly recommended before the detection of the lesions, as lesions appear in the retinal images as bright patches which are similar to the OD concerning their color and shape characteristics.

In the corresponding literature, several OD detectors have been published in [4], [5], [6], [7], [8], [9], [10], [11], [12] and [13]. Most of these try to perform the extraction of the OD based on the color, size, shape and direction of vessels, etc. They use various methods ranging from filtering [4] and threshold methods [5] to *k*NN regression [6]. There is, in fact, no reason to assume that any single algorithm would be optimal for the detection of various anatomical parts of the retina. It is difficult to determine which the best approach is, because good results were reported for healthy retinas but weaker ones for more challenging datasets containing diseased retinas with variable appearance of ODs in term of intensity, color, contour definition and so on.

In this thesis, to overcome the imperfectness of the individual detectors, we propose an ensemble of the individual algorithms. We propose three different approaches which combine individual algorithms in order to benefit from their strengths while overcoming their weaknesses. In the first approach, we tested a majority voting scheme with a circular template to detect the correct position of the OD center, where the individual algorithms have just a single candidate. In the second one, we extract more than one candidate for each algorithm to increase the chance of getting the OD location among them. We assign weight to each candidate to replace simple majority voting by a weighted one and treat them as weighted nodes of a weighted graph. In order to localize the OD center, we select that completed subgraph from the constructed weighted graph which meets the OD geometry constraint and has a maximum total weight. For this selection, we borrow a graph theoretical approach supplying the optimal solution in terms of a maximum-weighted clique. Finally, to utilize all available information corresponding to the possible location of the OD provided by the algorithms, we let them assign confidence values to each pixel of the input image. In this way, probability maps are composed by the member

algorithms and a suitable combination of these maps can be applied to locate the correct OD region.

Taking advantage of more information is supposed to lead to improvement, which is a natural expectation validated by our experiment studies. Namely, in our tests, the proposed method outperformed both the simpler ensemble-based systems and the state-of-the-art individual member algorithms on publicly available datasets.

After the OD has been localized properly, it can be used to improve the detection of the macula and the fovea as it is proposed in [3]. Next to OD and macula, CAD system also should segment the blood vessels in ophthalmological images because of the high similarities between the small vessels and the microaneurysms. Much work has been carried out in the field of detection of retinal vasculature [14], [15], [16]. Most of these approaches can detect the main and thick vessel segments with high precision but the detection of tiny ones is a difficult task. Since microaneurysms (first signs of DR) appear as small circular dark dots near thin vessels (see Figure 2.), the tiny non-segmented vessels cause many false positives at their detection [17], [18], [19]. So the reliability of an automatic screening system for DR can be much higher if it is also suitable to detecting signs of DR as exudates.

Exudates can be distinguished more efficiently from the background than microaneurysms from blood vessel segments. On the other hand, the fluid can flow without restrain, so the exudates have various size and irregular shape (see Figure 3), which makes the automatic detection of exudates a challenging task as well. In the corresponding literature, a large number of exudate detection algorithms have been published. In general, we can divide these approaches into two main groups. The first group contains algorithms based on grayscale morphology [20], [21], [22], while the second one consists of methods considering pixel/region-

wise classification [23], [24], [25], [26], [27]. Furthermore, we can find some special approaches (e.g., [28], [29]) beyond these groups.

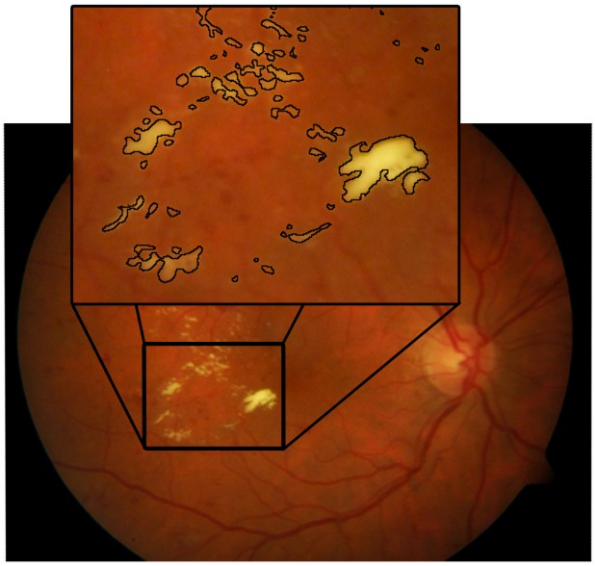

Figure 3: Exudates appear as yellowish, bright patches of various sizes and irregular shapes.

A method for exudate detection, which combines the mainstream approaches (morphology and classification) within a single framework, is proposed as the second main focus of this thesis. Our aim is to take advantage of several image enhancement methods for recognizing the precise boundaries of candidates extracted by a morphology-based candidate extractor. The motivation behind this objective is that the features extracted from the precisely segmented regions are more suitable to differentiate the true exudates from the false ones than the features extracted from the coarsely segmented ones. So, we recommend the use of several image enhancement algorithms to extract contour of candidates by an active contour method for each enhanced image. The final

candidate contour is found by a combination of these extracted contours. Finally, a region-wise classifier is applied to decide whether the candidates should be considered as exudates or not. The proposed fusion of grayscale morphology and active contour-based segmentation with region-wise classification can outperform several state-of-the-art approaches according to our empirical results.

The main contribution of this thesis applicable in medical science (including the corresponding publications as well) can be summarized as follows:

I. In connection with OD localization:

    1. Elaboration and application of a simple majority voting model for the localization of the OD [3], [30], [31], [32].

    2. Elaboration of a framework for extracting several candidates from member algorithms and the graph-based process of object detection for the localization of the OD [33].

    3. Elaboration of the technique of object detection based on the fusion of 2D probability maps for the localization of the OD [34].

II. In connection with exudate detection:

    4. Elaboration of an exudate-detector developed on the basis of grayscale morphology and pixel-classification [35].

    5. Elaboration of an exudate-detector algorithm based on region-wise classification [36], [37], [38].

6. Elaborating of a system applying some different image enhancement algorithms for achieving the most accurate contour detection. [39], [40].

Notwithstanding the thesis is written in plural form, the author has a principal contribution to the presented results. The models and methods which will be introduced in this thesis are the outcome of the collaboration of the author and his supervisor. The author engaged in the elaboration of the theoretical models, he implemented and tuned up all the introduced methods and finally evaluated them.

The rest of the thesis is organized in the following way: section 2 provides a brief summary on the basic concepts and methodology used in this thesis. Section 3 and 4 contain the main body of the thesis by describing the ensemble-based OD detection methods and the proposed exudate detection methods. Finally, conclusions are drawn in section 5.

2 Basic concepts and methodology

In this section, we introduce the basic notations for the easier and precise reference in the latter parts of the thesis. As we propose image processing algorithms, which can detect single and multiple objects in digital images, first we give the commonly used definition of the 8-bit intensity digital images I of size $w \times h$ as:

$$I: \{1, \dots, w\} \times \{1, \dots, h\} \to [0, \dots, 255], \tag{1}$$

where w and h denote the width and height of the image, respectively. The domain $\{1, \dots, w\} \times \{1, \dots, h\}$ of I will be denoted by $Dom(I)$ in the latter formulas. Thus, I is a 2D function which assigns for an intensity value to each point p of the discrete digital plane, where the coordinates (x_p, y_p) of p $1 \leq x_p \leq w$ and $1 \leq y_p \leq h$ hold. When the digital image is a 24-bit color image as:

$$I: \{1, \dots, w\} \times \{1, \dots, h\} \to [0, \dots, 255] \times [0, \dots, 255] \times [0, \dots, 255] \tag{2}$$

(e.g. RGB, HSI) which has three different 8-bit intensity channels, we denote it in subscript as I_{RGB}, I_{HSI}. Moreover, if we use a specific channel of a 24-bit color image we also denote the letter of the selected channel in the subscript of I as I_R, I_G, I_B and I_I.

A fundus photo is a digital image about the interior surface of the eye, including the retina, OD, macula, and posterior pole [41]. Fundus photography is used by ophthalmologists for monitoring progression of a disease in screening programs and epidemiology [42]. The fundus image is taken by a fundus camera, which basically consists of a special low power microscope with an attached camera providing an RGB color

image. So, I_{RGB} is the most common output of the fundus camera generally in use.

A large number of papers dealing with digital fundus image processing propose the green channel I_G for image analysis [43], since it provides the highest contrast between the anatomical parts, lesions and the retinal background, as it can be seen in Figure 4. Furthermore, we can also consider the intensity channel I_I from I_{HSI}, where I_I is defined as the average of the I_R, I_G and I_B, and we can keep some relevant information from the red and blue channels also (see Figure 4).

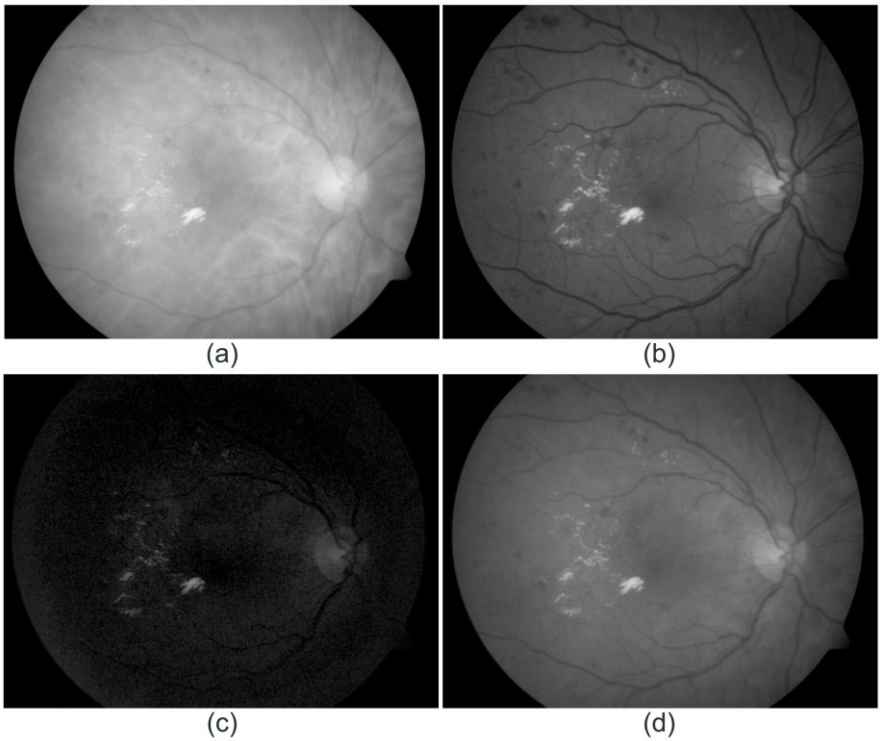

Figure 4: Channels of a sample image of Figure 3; (a) I_R, (b) I_G, (c) I_B, (d) I_I.

The value of w and h in case of I_{RGB} mostly depends on the resolution of the applied CCD sensor in the fundus camera. Nowadays, the commonly used fundus cameras in practice provide at least 5 megapixel resolution for images, so in the images the smallest signs of DR are composed of sufficient number of pixels, which enables their manual or automated evaluation. For example, the microaneurysms usually have a diameter less than 125 μm [44], so one of them consists of at least 100 pixels in image with resolution of 2500×2000.

Fundus cameras have another essential property, namely, the angle of acceptance of the lens (field of view or FOV). A typical camera views 30˚ to 50˚ of retinal area, with a magnification of 2.5× which means that the anatomical parts are 2.5 times larger than in real life [45]. See Figure 5 for visual explanation of the meaning of different FOV size.

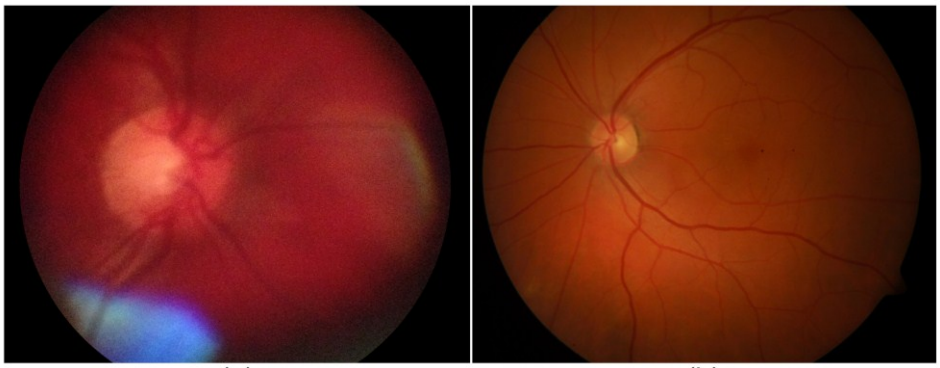

Figure 5: Sample fundus images with different FOV sizes; (a) 25˚ FOV, (b) 50˚ FOV.

As it can be seen, the ratio between the area of OD and the area of region of interest (ROI) is the most significant difference between these images, where the ROI is the region which contains the relevant inner areas of the

retina and it can be extracted by standard algorithm (e.g. [46]), as it can be seen in Figure 6.

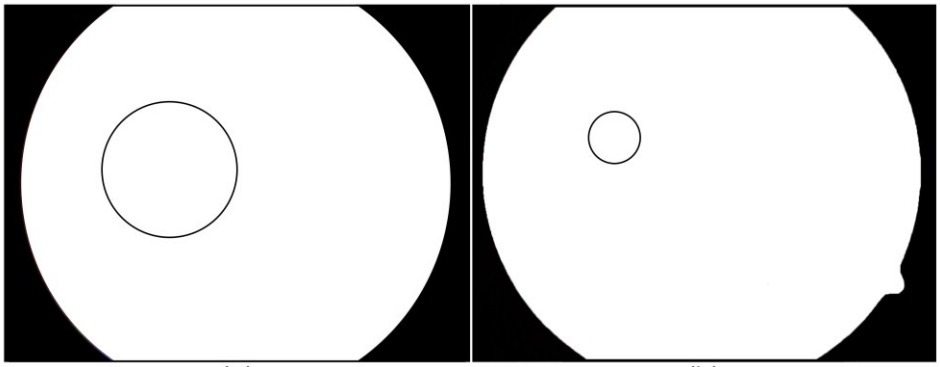

Figure 6: ROI with the boundary of OD (black circle) for the sample images of Figure 5.

The greater is the FOV, the more parts of the retina can be seen in the image. Thus, an automatic screening system can evaluate a fundus image based on much more information if it gets an input image taken by a camera with higher megapixel and FOV. In this thesis the proposed algorithms are tested on datasets which contain images with 45˚ or 50˚ FOV. Based on the evaluation of approximately 2000 images with manually segmented optic disc region (ODR), the radius of an average OD is 6.5% of the width of the ROI of the fundus image at the investigated scale of FOV.

2.1 Fundus image datasets

As we mentioned in section 1, basically we introduce two methods in this thesis. The first one is developed to detect a single object in the images which can be used for accurate OD detection in retinal images. The

second one combines grayscale morphology, precise boundary detection and region-wise classification approaches to decide whether the image contains exudates or not. In this section, we introduce some publicly available datasets which are supplied with manual annotations of ophthalmologists which data can be used for the evaluation of the automatic detection and segmentation methods regarding anatomical parts and lesions.

2.1.1 DIARETDB0 dataset

DIARETDB0 – Standard Diabetic Retinopathy Database [47] includes 130 color fundus images. 20 of these are healthy and 110 of these contain signs of DR (exudates, microaneurysms, hemorrhages and neovascularization). The images were taken with a 50˚ FOV, at a resolution of 1500×1152 pixels and the average diameter of the ROI is 1410 pixels.

2.1.2 DIARETDB1 dataset

The dataset DIARETDB1– Standard Diabetic Retinopathy Database [48] includes 89 fundus images. 84 of these contain microaneurysms or more serious signs of DR, while 5 are considered as healthy with containing no signs of DR according to clinical experts. The images were taken with a 50˚ FOV, at a resolution of 1500×1152 pixels and the average diameter of the ROI is 1410 pixels.

2.1.3 DRIVE dataset

DRIVE [49] images were acquired using a Canon CR5 non-mydriatic 3CCD camera with a 45˚ FOV. Each image is 24-bits RGB of resolution

768×584 pixels. The ROI of the images is circular with an average diameter 540 pixels.

2.1.4 MESSIDOR dataset

The image set MESSIDOR [50] includes 1200 fundus images which were acquired by 3 ophthalmologic departments using a color video 3CCD camera on a Topcon TRC NW6 non-mydriatic retinograph with a 45˚ FOV. The images were captured using 24-bits RGB at the resolution of 1440×960, 2240×1488 or 2304×1536 pixels.

2.1.5 HEI-MED

The dataset HEI-MED – Hamilton Eye Institute Macular Edema Dataset [51] consists of 169 images of resolution 2196×1958 pixels with a 45˚ FOV, among which 54 images are classified manually by an ophthalmologist as containing exudates.

2.1.6 Manual annotations

The sets DIARETDB0, DIARETDB1, DRIVE and MESSIDOR have been considered for performance analysis of OD detection method. For the measurability of the algorithm accuracies, a manually selected ODR is also supplied for the images of these datasets. A sample manual drawn ODR can be seen in Figure 7.

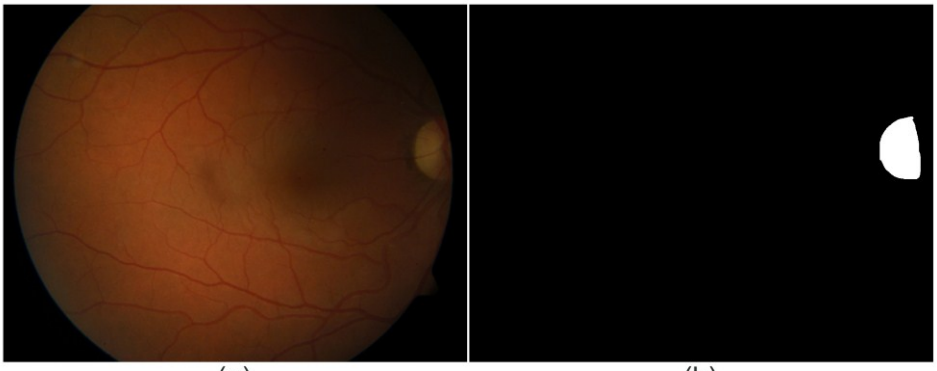

Figure 7: Sample binary image for manual annotated ODR; (a) RGB fundus image, (b) manual drawn ODR.

We evaluate our proposed method for exudate detection on the dataset DIARETDB1 and HEI-MED also. 53 images of DIARETDB1 contain exudates based on the labeling of four clinical experts. Each expert marked manually the most representative points of the exudates in these images, and the coordinates of these points are stored in text files. Based on these manually-marked anchor points, a local ophthalmologist precisely segmented the exudates in these images. Thus, we have gained 53 binary masks containing the precise exudate boundaries as it can be seen in Figure 8. In the case of HEI-MED set, we know the labels (containing exudates or not) for all images but binary masks containing the manually segmented exudates are not available for them.

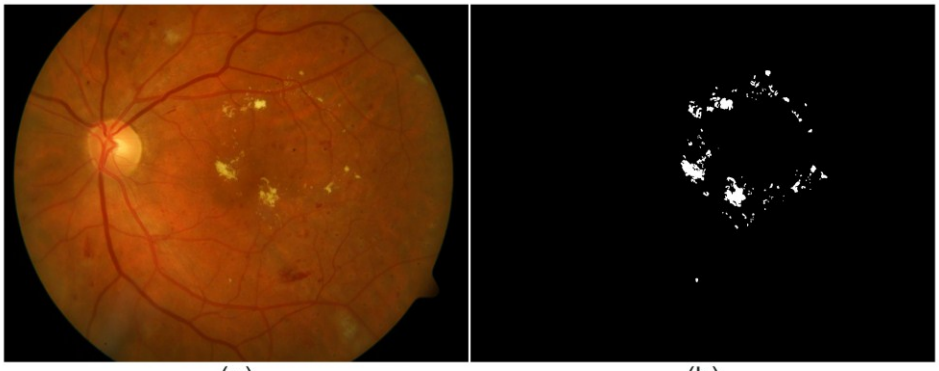

Figure 8: Manual annotation containing precise exudate regions; (a) RGB fundus image, (b) manual drawn binary mask.

2.1.7 Training sets

The proposed approaches for detection single object and multiply objects require a training phase when training images are used to construct the applied machine learning-based models. Therefore, our private dataset (TR_1) is used only during the training phase of the single object detection method. To avoid the problem of information leakage TR_1 is not used for the evaluation of the proposed methods. The images from our dataset were captured with a 50˚ FOV at a resolution of 3072×2048 and 1360×1024 pixels with the average diameter of the ROI is 2287 and 1340 pixels, respectively. This set includes 327 images, for which manually drawn regions of the OD are also available. This dataset was provided by Moorfields Eye Hospital (London, United Kingdom) and the manual annotations were made by Adrienne Csutak and Tünde Pető.

On the other hand, for the training phase of the exudate detection method, the dataset DIARETDB1 is divided into training (TR_2) and test

part, because manually segmented precise exudates exist only in the case of these images. Separating the dataset is performed by taking the distribution of normal and abnormal images also into account, as proposed by Kauppi *et al.* [48]. The TR_2 consists of 24 images containing exudates and 4 images with no such lesions.

Table 1 contains the summary of datasets used for the training and the evaluation of the proposed methods.

	Images	Normal	Abnormal	FOV	ROI (pixels)	Resolution (pixels)
DIARETDB0	130	20	110	50˚	1410	1500×1152
DIARETDB1	89	5	84	50˚	1410	1500×1152
DRIVE	40	33	7	45˚	540	768×584
MESSIDOR	1200	546	654	45˚	908 1374 1450	1440×960 2240×1488 2304×1536
HEI-MED	169	3	166	45˚	2193	2196×1958
TR_1	327	241	86	50˚	2288 1340	3072×2048 1360×1024

Table 1: Summary of the datasets

3 Optic disc detection by fusing the outputs of individual detectors

The automatic detection of the location of the OD is an important issue in the automatic analysis of the retinal images. The OD is a circular shaped anatomical structure with a bright appearance as it has been already mentioned. The location is to be found where the optic nerves enter the eye. If the position and the radius of the OD are detected correctly, then they can be used as references for approximating other anatomical parts. Moreover, information about the OD can be used to examine the severity of some diseases, such as glaucoma. It is often needed to mask the OD out before bright lesion detection (e.g. exudates, cotton-wool spots) because of its similarity in brightness and color in a diseased retina.

Several approaches have been published how the OD can be localized automatically. For example, Mendels *et al.* [4] used morphological filtering and active contours to find the boundary of the OD, while Sekhar *et al.* [5] applied morphological operations and Hough-transformation to localize it. Here, the proposed method consists of two steps: in the first step, a circular region of interest is found by isolating the brightest area in the image by means of morphological processing, and in the second step, the Hough-transformation is used to detect the main circular feature (corresponding to the OD) in the positive horizontal gradient image within this region of interest. Abramoff *et al.* [6] proposed a *k*NN regression approach to find relationship between the dependent variable which represents the distance from the OD center, and a feature vector extracted around a circular template. Shijian [7] used circular transformation which is capable of detecting both the OD center and its boundary from images of pathological retinas, as well. Lalonde *et al.* [8] proposed an algorithm which generates a pyramid with a simple Haar-

based discrete wavelet transformation. The input image has low resolution after its decomposition is repeated four or five times and the pixel of this low-resolution image with highest intensity value is considered as the center of the OD. Another approach in [8] uses edge detection applying the Rayleigh-based CFAR threshold. This solution guarantees the strong edges which are associated with the border of anatomical structures. In the next step, the Hausdorff distance is calculated between the set of edge points and a circular template like the average OD. The lowest distance value corresponds to the center of the OD. Sopharak *et al.* [9] proposed a method which applies a median and a contrast-limited adaptive histogram equalization filter on the green intensity channel. The entropy of the intensity values in the local region around each pixel is calculated and the location of the highest entropy values is considered to belong to the OD. Niemeijer *et al.* [10] proposed feature extraction and classification steps to determine the area of the OD. They propose the following features: number, width, orientation and density of vessels segments. In the next step, a *k*NN classifier made a decision about each pixel whether they belong to the OD, or not. Hoover *et al.* [11] thinned the vessel system and each line-shape segment is modeled by a fuzzy segment. This model creates a voting map and the pixel obtaining the most votes is considered to be the center of the OD. Ravishankar *et al.* [12] proposed an algorithm which uses Hough-transformation on the thinned vessel system. Lines which have slope less than 45˚ are eliminated. The intersection points of the rest of the lines give a voting map. Number of votes is weighted with the original intensity values of the intersection points in the image, and the highest corresponding value indicates the center of the OD. Finally, Zhu *et al.* [13] locate the border of the OD in terms of a circle with a given diameter using circular Hough-transformation. For this aim, edge detection is applied and the circle containing the most edge points is selected.

The majority of these algorithms are able to locate the OD center on healthy images. However, their accuracies decrease if the image quality is not sufficiently high or severe lesions appear on the image. To overcome the imperfectness of the individual algorithms, we study and adapt some of the state-of-the-art OD detectors and finally organize them into an ensemble framework in order to combine their strengths and maximize the accuracy of the localization of the OD.

In the literature of financial, medical and social, many popular ensemble based methods [93] can be find as bagging, boosting, stacked generalization, and hierarchical mixture of experts, rules, including algebraic combination of outputs, voting based techniques, behavior knowledge space, and decision templates. These methods can be used when we have a second opinion before making a decision, sometimes a third one, and sometimes many more and when our goal is to choose one option from a previously defined set of options. However, in the case of OD localization we have to search the center point of the OD in a 2D discrete space where the final output is not a class label from a set of options. This means that these fusion methods trivially could not be applied for this problem.

In this section, we introduce our three different fusion strategies, which are based on the majority voting scheme, finding maximum-weighted clique and combining probability maps, respectively. We study and adapt some of the state-of-the-art OD detectors and finally organize them into a complex framework in order to maximize the accuracy of the localization of OD. Besides providing detailed description of these frameworks, we show the positive effect of exploiting more and more information for this issue.

The rest of this section is organized as follows. In section 3.1, we introduce some concepts regarding the single object detection and we give a brief overview about the OD detectors involved in the ensembles in

section 3.2. In section 3.3, we give an overview about the proposed majority voting-based ensemble system. In section 3.4, we describe how we can extract more candidates from the member algorithms and determine the OD center by finding maximum-weighted clique in weighted graph. In section 3.5, we introduce our methodology for the detection of the center and region of the OD based on the combination of probability maps provided by the individual algorithms. Section 3.6 is dedicated to our experimental results also in comparison with some other state-of-the-art OD detectors. Finally, further technical details are discussed in section 3.7.

3.1 Concepts in the fields of single object detection

In this section, we provide some basic concepts for single object detection in digital image for the easier and precise reference in the latter parts of the section 3. As it has already been mentioned, we propose three different ways to combine automatically the outputs of different detectors to benefit from their strengths while overcoming their weaknesses.

We can consider the problem of the localization of a single object knowing that the presence of the object has high confidence in the image. Namely, we do not have to determine whether the single object is there in the image or not because the input image surely contains the object. For making an ensemble system to localize the center of the object, we should have several member algorithms to combine their outputs.

Definition 1: *Let* $N \in \mathbb{N}$*, and* A_i *(*i=1,2,…,*N) be a member algorithms as* $A_i: I \to C_i \subseteq Dom(I)$*, which can detect the center point of the derived single object and provide same output to the same input image.*

N denotes the number of the involved member algorithms, I means the input image and C_i is the output of A_i.

Definition 2: *Let* $C_i \subseteq Dom(I)$ *be the set of the outputs of* A_i *(*i*=1,2,…,N) for image I.* $|C_i|$ *denotes the number of candidates provided by* A_i *(*i*=1,2,…,N).*

Definition 3: *Let* $c_{i,j} \in C_i$ *be the j-th candidate of the i-th member algorithm* A_i *(*i*=1,2,…,N), (*j*=1,2,...,*$|C_i|$*) for the center of the single object.*

We propose a majority voting-based ensemble of A_i if $|C_i| = 1$ (i=1,2,…,N) which means that A_i (i=1,2,…,N) provides only one pixel as a candidate of the center of the object. In case $|C_i| > 1$, A_i (i=1,2,…,N) suggests more than one candidate and we apply maximum-weighted clique-based ensemble. Finally, when $|C_i| = |I|$, we can determine the center and the region of the object by combining the probability maps.

For finding maximum-weighted clique in weighted graph, we have to compose a weighted graph using $c_{i,j} \in C_i$ of A_i with letting A_i (i=1,2,…,N) assign a confidence level to each of its candidates. Let the ground-truth center coordinates (x_{man}, y_{man}) of the single object be denoted by p_{man}. Since most of the object detection algorithms use various features of p and its neighborhood for localization, $H_i(p)$ *(*i*=1,2,…,N)* denotes the set of features based on which A_i assign a confidence value to $p \in I$. Here, we omit any details on the feature sets H_i, as they are completely algorithm dependent.

Definition 4: *Let* $Conf_i$ *is a function generated by* A_i *(*i*=1,2,…,N) as follows:*

$$Conf_i: \{1, \dots, w\} \times \{1, \dots, h\} \to \mathcal{R}^+. \quad (3)$$

Consequently, if $p \in I$ then $Conf_i(p) = P\big(p = p_{man} | H_i(p)\big)$, where $P\big(p = p_{man} | H_i(p)\big)$ is the conditional probability of that $p = p_{man}$ given $H_i(p)$ features which shows the confidence of A_i that p is the center of the object.

If we consider $Conf_i(c_{i,j}) \epsilon \mathcal{R}^+$ assigned to $c_{i,j} \in C_i$ by A_i *(i*=1,2,…,*N), (j*=1,2,…, $|C_i|$*)* then we can sort $c_{i,j} \in C_i$ in a priority order so that $Conf_i(c_{i,k}) \geq Conf_i(c_{i,l})$ for any 1≤*k*<*l*≤$|C_i|$, *(i*=1,2,…,*N)* and assign weights to the candidates.

Definition 5: *Let* $\omega_{i,j}$ *be a weight assigned to* $c_{i,j} \in C_i$ as:

$$\omega_{i,j} = \frac{Conf_i(c_{i,j})}{Conf_i(c_{i,1})} \quad (4)$$

for i=1,2,…,*N and j*=1,2,…,$|C_i|$.

Note that, this normalization will give $\omega_{i,1} = 1$ (weight 1 for every first candidate) for *i*=1,2,…,*N*. After having defined $c_{i,j} \in C_i$ and $\omega_{i,j}$, we can construct a weighted graph as in [52].

Definition 6: *Let* $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{W} \rangle$ *be an undirected weighted graph, where the set of vertices*

$$\mathcal{V} = \bigcup_{i=1}^{N} C_i \quad (5)$$

$\mathcal{E}$ *is a set of edges and* $\mathcal{W}$ *is a set of weights* $\omega_{i,j} \in \mathcal{R}^+$ *for each vertex (i*=1,2,…,*N), (j*=1,2,…,$|C_i|$*). Edge connects two adjacent vertices*

$(v_{i,j}, v_{k,l} \in \mathcal{V})$ *if* $\|v_{i,j}, v_{k,l}\|_2 < d$, *where the distance* $d \in \mathcal{R}^+$ *depends on the size of single object and* $\|.\|_2$ *stands for Euclidean norm.*

Definition 7: *For* $S \subseteq V$, $\mathcal{G}(S) = \langle S, \mathcal{E} \cap S \times S, \mathcal{W}(S) \rangle$ *is the subgraph induced by S. A set of vertices* S^* *is called a clique if the subgraph* $\mathcal{G}(S^*)$ *is completed, namely, there is an edge between any two adjacency vertices in* $\mathcal{G}(S^*)$.

Definition 8: *Let* $S_1^*, S_2^*, \ldots, S_n^*$ *denote the cliques of* $\mathcal{G}$. S_i^* *is the maximum-weighted clique (MWC) if the total weight of the vertices of* S_i^* *is maximal.*

For the combination of probability maps we expect the member algorithms to assign a confidence value to each $p \in I$ indicating their confidence that p is the center of the object.

Definition 9: *Let* PM_i *denote the probability map generated by* A_i *for I, where*

$$PM_i = \bigcup_{j=1}^{|C_i|} Conf_i(c_{i,j}), \tag{6}$$

$c_{i,j} \in C_i$ *and* $|C_i| = |I|$.

So, our aim is to localize the center of the OD with high accuracy in fundus image. Namely, the detected center point falls within the true OD region and the distance between the detected and the manually drawn center point should be as smallest as possible. For the measurability of the accuracies of the proposed ensembles, we consider three ways. First, as a primary error, we expect that a properly detected center falls within the manually drawn OD region (ODR). As a secondary measurement, we

check the distance of the combined output from the centroid of the ODR. Finally, we measure the similarity between the segmented region and the ODR. Let the finally localized center coordinates (x_{res}, y_{res}) of the single object be denoted by p_{res}.

Definition 10: *Let $ODR \subseteq Dom(I)$ be a manually segmented region of OD and $p_{man} \epsilon ODR$. For the p_{res} on the same image regarding ODR we apply the following:*

- *p_{res} is a true positive (TP_1), if $p_{res} \epsilon ODR$.*
- *p_{res} is a false positive (FP_1), if $p_{res} \notin ODR$.*

We note that, in the case of localization of a single object which appears in the input image with high confidence, the definition of false negative and true negative are meaningless.

Definition 11: *To measure the primary error within a given dataset we can calculate the precision or positive predicted value (PPV) to measure the primary error function as:*

$$PPV = \frac{\#\ of\ TP_1}{\#\ of\ TP_1 + \#\ of\ FP_1}, \tag{7}$$

where $\#\ of\ TP_1/FP_1$ means the number of TP_1/FP_1 in the whole test dataset.

Definition 12: *As for the secondary error function (Err_2), for a fair comparison we exclude those images, where the detected OD center falls beyond the OD region:*

$$Err_2 = \frac{1}{\#\ of\ TP_1} \sum_{p_{res} \in ODR} \|p_{res}, p_{man}\|_2. \quad (8)$$

Definition 13: *Let* $ODR_{res} \subseteq Dom(I)$ *denotes a region of OD determined by the ensemble of probability maps generated by* A_i *for I.* Err_3 *function measure the overlapping between the ODR and* ODR_{res} *regarding the same I within a given dataset as a follows:*

$$Err_3 = \frac{1}{\#\ of\ TP_1} \sum \frac{ODR_{res} \cap ODR}{ODR_{res} \cup ODR}. \quad (9)$$

3.2 Member algorithms for OD detection

In this section, we give a brief overview about the OD detectors which are involved in the proposed ensemble systems for single object detection. As we show in section 3, we can find huge number of OD detectors in the literature [4], [5], [6], [7], [8], [9], [10], [11], [12], [13]. We reimplemented the next seven detectors based on different principles in Matlab environment by following the respective instructions given in [8], [9], [10], [11], [12], [13].

3.2.1 OD detection based on pyramidal decomposition (OD_{decomp})

This algorithm relies on three assumptions. First, the image is centered on the macula or OD, second, the OD represents a bright region (not necessarily the brightest) and finally, the form of the OD is approximately circular. Based on the hypothesis that the OD is roughly a circular patch of bright pixels surrounded by darker pixels Lalonde *et al*. [8] propose to

locate the candidate OD regions in the I_G by mean of pyramidal decomposition (Haar-based discrete wavelet transform [53]). In the low resolution image pixels, which have the highest intensity values compared with the mean pixel intensity over the search area, were selected as possible candidates. Next, smoothing is done within each of these regions and the brightest pixel is selected as a possible OD center point and its confidence value is computed as the ratio of average pixel intensity inside a circular region centered at the brightest pixel and the average intensity in its neighborhood.

3.2.2 OD detection based on edge detection (OD_{edge})

In this method, Lalonde *et al*. [8] search the area identified by the pyramidal decomposition (see section 3.2.2) for a circular shape. To reduce the number of regions of interest, contiguous regions were aggregated into a single zone. A binary edge map is obtained by performing Canny-edge detection [54] in the region of interest first, and then a thresholded image is obtained with a special threshold value computed from noisy edge map. The search for the OD contour is performed by using an algorithm-based on Hausdorff-distance [55]. The Hausdorff-distance provides a degree of mismatch between two sets of points.

Several circular templates of variable sized diameters were used to compute the Hausdorff distance between the templates and thresholded image containing edges. Hence, a percentage of matches are computed, and if the certain proportion of the pixels template is found to overlap edge pixels in thresholded image then the location is retained as the center point of a potential OD candidate.

3.2.3 OD detection based on entropy measurement ($OD_{entropy}$)

Sopharak *et al*. [9] presented the idea of detecting the OD by entropy filtering. The original I_{RGB} image is transformed into I_{HSI} color space, median filtering is applied to remove possible noise and for contrast enhancement Contrast Limited Adaptive Histogram Equalization (CLAHE) [56] is done to the I_I. After preprocessing, OD detection is performed by probability filtering based on computed entropy value in a local region around each $p \in I_I$. The local pixel intensity entropy measure is high when the region around a pixel is complex and low when it is smooth. Binarization is done with Otsu's algorithm [57] to separate the complex regions from the smooth ones, and the largest connected region with an approximately circular shape is marked as a candidate for the OD.

3.2.4 OD detection based on *k*NN classification ($OD_{classify}$)

Niemeijer *et al.* [10] defines a set of features based on vessel map and image intensity, like number of vessels, average width of vessels, standard deviation, orientation, maximum width, density, average image intensity, measured under and around a circular template to determine the location of the OD. After pre-processing, each image is scaled so that the width of its ROI is 630 pixels. The binary vessel map obtained by [59] is thinned until only the centerlines of the vessels remain and all the centerline pixels that have two or more neighbors are removed. Next, the orientation of the vessels is measured by applying principal component analysis on each centerline pixel with its neighboring pixel on both sides. A two-step sampling process is launched to get the training database. First, by using the circular template having manually selected the OD center within the radius, all features are extracted for each sample location (a uniform grid spaced 8 pixels apart) of the template including

distance *d* to the true center. In the second step, 500 randomly selected locations (i.e., not on a grid) in the training image were sampled in a similar fashion. To locate the OD, a sample grid (grid points spaced 10 pixels apart) is overlaid on top of the complete ROI and the feature vector are extracted and the rough location of the OD is found containing pixels having lowest value of *d*. The process is repeated with a 5×5 pixel grid centered on the rough OD location to get the more accurate OD center.

3.2.5 OD detection based on fuzzy convergence of vessels (OD_{fuzzy})

Hoover *et al*. [11] described a method based on a fuzzy voting mechanism to find the location of the OD. The method uses the convergence of the blood vessel network as the primary feature for detection. In the absence of a unique and strongly identifiable convergence, the brightness of the nerve is used as a secondary feature. The input to the algorithm is a binary segmentation of the blood vessels. The segmentation is achieved by thinning the vessel image and relabeling the branch points as background, thus, breaking up the foreground (vessel network) into segments that contain two end points each. Each line segment (vessel) is extended at both ends by Fuzzy element (15 pixels). The area of this fuzzy segment contributes a vote to its constituent pixels. The summation of votes at each pixel produces an image map where each pixel contains a value proportional to its strength of convergence. The map is then blurred and thresholded to produce one point of strongest convergence.

3.2.6 OD detection based on Hough-transformation (OD_{hough})

Ravishankar *et al.* [12] tried to track the OD by combining the convergence of the only thicker blood vessel initiating from it and high disk intensity properties in a cost function. On initially resized image to

standard resolution (768×576), a grayscale closing operation is performed on I_G. This step is followed by thresholding and median filtering to obtain the binary image of the blood vessels. The segments of the thicker blood vessels skeleton are modeled as lines found by the Hough-transformation [58]. The dataset of lines generated is reduced by removing those lines with slopes $\theta < 45^{\circ}$. This reduced dataset of lines is intersected pairwise to generate an intersection map. The map is dilated to make the region of convergence more apparent. A weighted image is produced by combining this dilated intersection map and preprocessed I_G. A cost function is defined to obtain the optimal location of the OD that is a point which maximizes the cost function.

3.2.7 OD detection based on circle detection (OD_{cHough})

Zhu *et al.* [13] locate the border of the OD in terms of a circle with a given diameter using the Hough-transformation. The proposed method locates automatically the OD in fundus images based on its properties. The method includes edge detection using the Sobel method [56], and detection of circles using the extended Hough-transformation [60] which can identify circles and other parameterized geometrical shapes. The extended Hough-transformation assists in the detection of the center and radius of a circle that approximates the margin of the OD. Based on the feature that the OD is one of the bright areas in a fundus image, potential circles detected by the Hough-transformation are analyzed using intensity.

3.3 Majority voting-based ensemble

In this section, we propose an ensemble-based single object detection system based on simple majority voting which outperforms the involved

seven (N=7) member detectors ($A_1 = OD_{decomp}$, $A_2 = OD_{edge}$, $A_3 = OD_{entropy}$, $A_4 = OD_{classify}$, $A_5 = OD_{fuzzy}$, $A_6 = OD_{hough}$, $A_7 = OD_{cHough}$).

3.3.1 One member algorithm – one candidate

In this ensemble we leave the member algorithms unaltered, so they function exactly as their authors proposed in [8], [9], [10], [11], [12] and [13]. Namely, they localize the center of the OD and finally they propose only one candidate ($|C_i| = 1$, i=1,2,…,N) per image. These single outputs for the object center are merged and the majority voting scheme is applied using a template of the shape of the object to detect its correct position. For an impression, see Figure 9 for the output of A_i (i=1,2,…,N) together with the manually selected center (ground-truth) p_{man} of the OD.

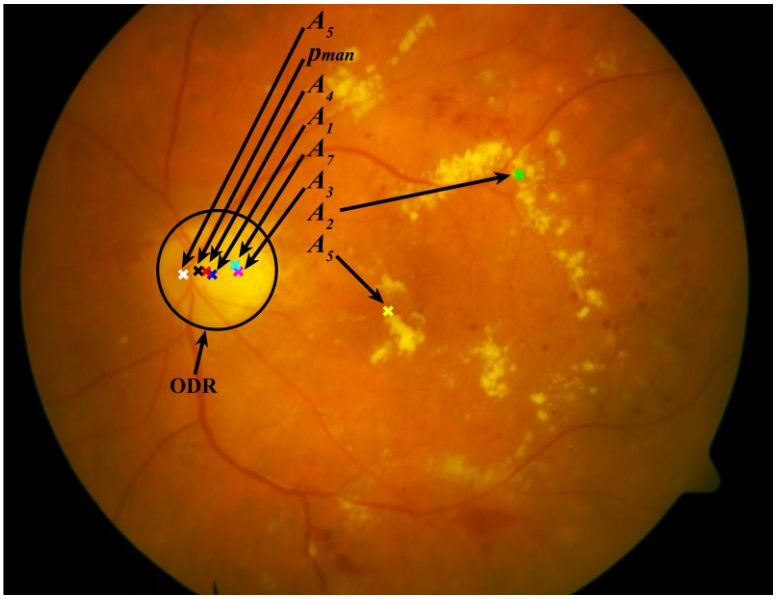

Figure 9: Results of different OD detector algorithms and the manually annotated OD center.

3.3.2 Methodology of the majority voting-based ensemble

As it can be seen we cannot localize the right position of the OD by using only the result of the A_2 or A_5. To overcome the imperfectness of the member algorithms, we devise a circular template voting scheme to determine the hotspot region, namely, an area in the image where majority of the outputs lies. A circular template D_R of radius $R > 0$ is fit on each $p \in I$ and outputs of candidate algorithms that fall within D_R are counted. The D_R covering the maximum number of OD detector outputs in its radius is considered to be a hotspot. There can be more hotspots covering the maximum number of detector outputs in their radius; hence they together define a hotspot region, the patch with highest probability. As we have mentioned in section 2, the value of R is set to 6.5% of the width of the ROI of the fundus image.

Following the principal of the majority voting, the center of D_R covering the maximum number of OD detectors output in its radius is considered to be the OD hotspot region. If there is a tie, such conflicts are handled by using an additional post processing step. Namely, a Gaussian filter is applied on I_G with a large variance parameter (σ = 300). The smoothed image is subtracted from I_G to make the OD appear as a brighter patch compared with the background. The average intensities around the output of the detectors are computed in the hotspot regions and the region with the highest average intensity is selected as the final OD hotspot region.

To find the final center of the single object p_{res}, the centroids of the outputs within the finally selected hotspot region (HPR) are computed. So, if we consider the single candidates $c_{i,1} \in HPR$ of A_i (i=1,2,…,N) and their coordinates as $(x_{c_{i,1}}, y_{c_{i,1}})$, then the coordinates of the final OD center $(x_{p_{res}}, y_{p_{res}})$ is computed as:

$$\left(x_{p_{res}} = \frac{1}{|HPR|} \sum_{c_{i,1} \in HPR} x_{c_{i,1}}, y_{p_{res}} = \frac{1}{|HPR|} \sum_{c_{i,1} \in HPR} y_{c_{i,1}} \right), \tag{10}$$

where $|HPR|$ denotes the number of $c_{i,1} \in HPR$.

We note that the center of the final hotspot regions could be found by averaging algorithms outputs based on (10), however, for a more accurate estimation, weights can be associated with detector outputs to determine the final location. The principal difficulty is how to choose the 'right' weights. We investigated this problem in [3] but the results are not part of this thesis.

3.3.3 Localization of the OD by majority voting

In Figure 10, we can observe the hotspot region composed by the hotspots containing the maximum number of outputs. As for the specific example in Figure 10, with ignoring the candidates of A_2 and A_5, the true OD location has been found based on the ensemble of the result of A_1, A_3, A_4, A_6 and A_7.

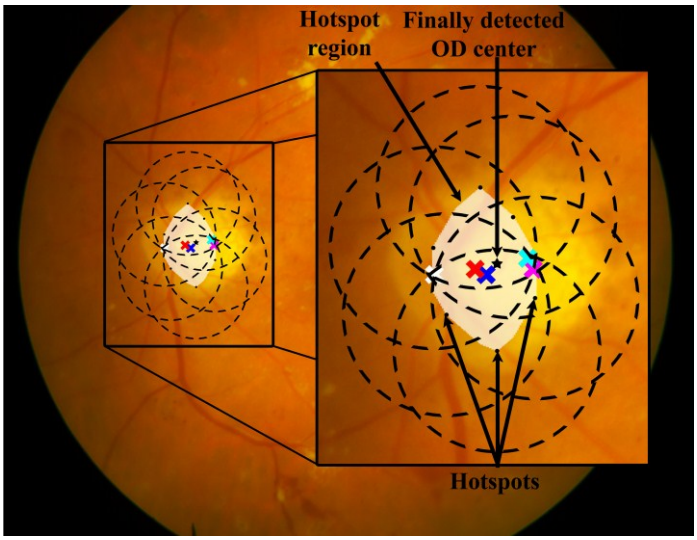

Figure 10: Locating the OD, when member algorithms have single candidates. Black dots indicate hotspot centers, while the bright patch is the hotspot region.

However, as it can be seen in Figure 11, majority voting may fail to detect the correct OD center, especially, in case of a diseased retina. Therefore, we propose a more sophisticated method to avoid these shortcomings, explained in the following section.

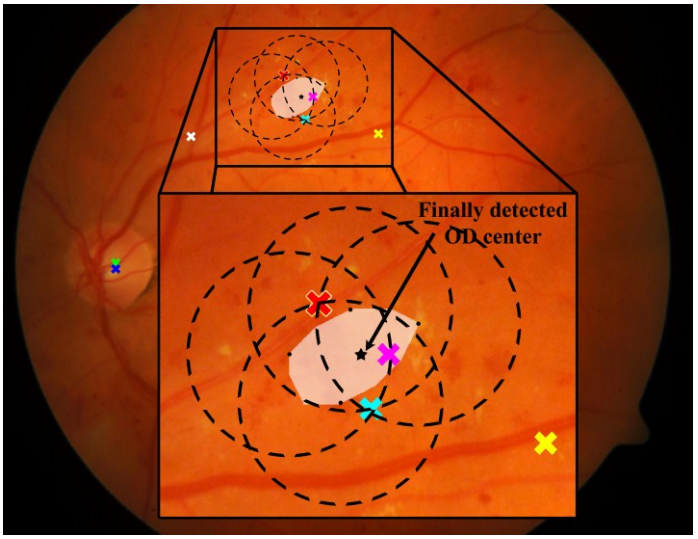

Figure 11: Incorrect OD localization based on simple majority voting.

3.4 Maximum-weighted clique-based ensemble

In this section, we examine how to improve the performance of the ensemble-based system in that case when the candidates of the member algorithms do not fall within the true single object region. We let the members suggest more candidates to increase the accuracy of that one of them falls within the true region of the object. So, our objective becomes to extract more candidates from the algorithms to have the appropriate location of the object among them. We assign weights to each candidate based on their priority proposed by the members. Then, these weighted candidates are considered as vertices of a graph, where we are looking for a subgraph with a maximum sum of weights constrained also by the geometry of the single object.

To see the performance of the framework considering more candidates of the members regarding the problem of OD detection, we

have included the same seven member (N=7) algorithms ($A_1, A_2, A_3, A_4, A_5, A_6$ and A_7) to make the results comparable with the simple majority voting scheme.

3.4.1 One member algorithm – several candidates

In this approach, we also preserve the basic principles of the member algorithms, but we have to extract more than one candidate per image. Since, most of the involved member algorithms are capable of assigning a value $Conf_i(p) \epsilon \mathcal{R}^+$ (i=1,2,…,N) to each $p \in I$ to show the probability of being the OD center as given in definition 4. These values together compose the PM_i (i=1,2,…,N) for the image as given in definition 7. See Figure 12 for the probability maps of the algorithms for the sample image of Figure 11.

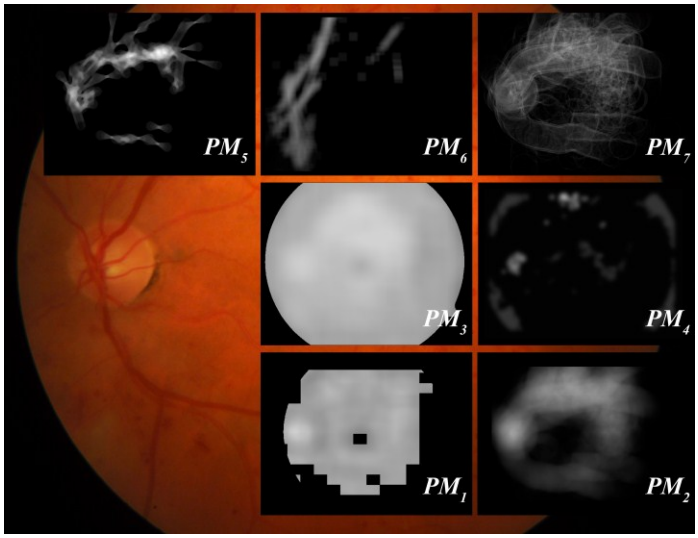

Figure 12: PM_i of A_i (i=1,2,…,N) showing their confidence whether an image pixel corresponds to the OD center.

To have multiple candidates from an algorithm, we can simply select more pixels as OD center candidates which have large probability value. As a geometric constraint, the candidates of the same algorithm are not allowed to be closer than the average OD diameter, thus:

$$\left\| c_{i,j}, c_{i,k} \right\|_2 > 2R, \tag{11}$$

where $c_{i,j}, c_{i,k} \in C_i$ of A_i (i=1,2,…,N), (j, k=1,2,…, $|C_i|$) and $j \neq k$. In this way we can assure that a possible OD region is suggested by only one candidate of each algorithm as it can be seen in Figure 13.

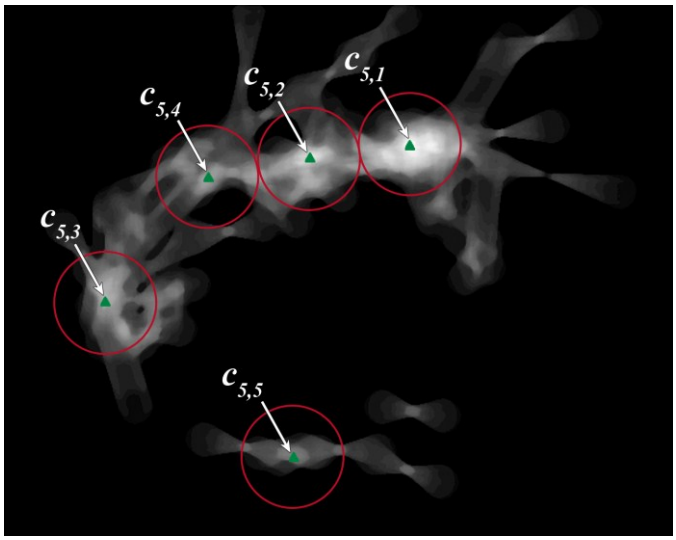

Figure 13: PM_5 of A_5 with candidates meeting the OD geometry constraint.

In the case of algorithms which are not capable of assigning a $Conf_i(p) \epsilon \mathcal{R}^+$ (i=1,2,…,N) to each $p \in I$, we can also extract more than one candidate per image in a natural way. For this purpose, we should mask out the first candidate of the member algorithm by D_{2R}, following

that the algorithm should locate the next candidate within the remaining image region and so on. For visual explanation of this alternative candidate extraction see Figure 14. In this way, we can also guarantee that only one candidate of each member suggests one possible OD region.

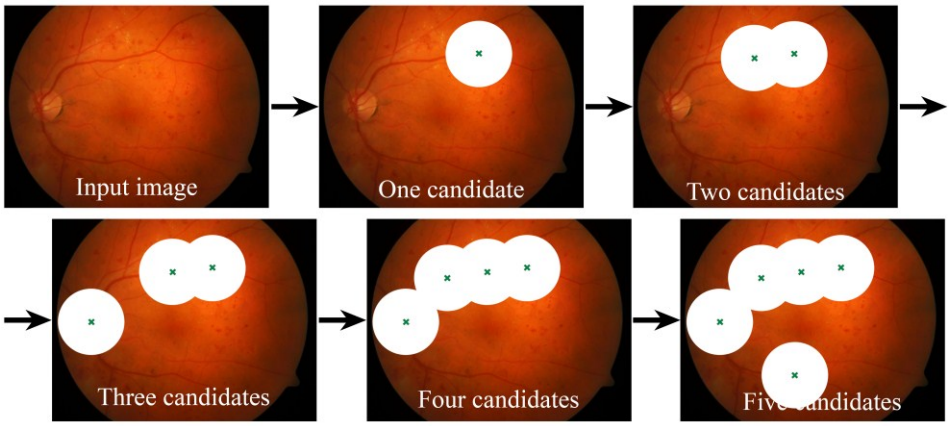

Figure 14: Visual explanation of an alternative way of the extraction of multiple candidates.

3.4.2 Right number of candidates

In this ensemble-based approach it is a key issue to determine how many candidates for a member we should consider. Let c_i^* mean that candidate of A_i which falls within the ODR. Then, on the one hand we should increase $|C_i|$ till $c_i^* \in C_i$. On the other hand we have to keep $|C_i|$ relatively small to avoid too many false positive candidates confusing the final combination. To solve this problem, for each algorithm, we used a training dataset (TR_1) and checked the ratio of the images with $c_i^* \in C_i$ and selected the smallest $|C_i|$ as the number of candidates above which the increase of $P(c_i^* \in C_i)$ slows down (see Figure 15 for the proper

behavior). Note that, this selection is rather close to the elbow method in kNN clustering.

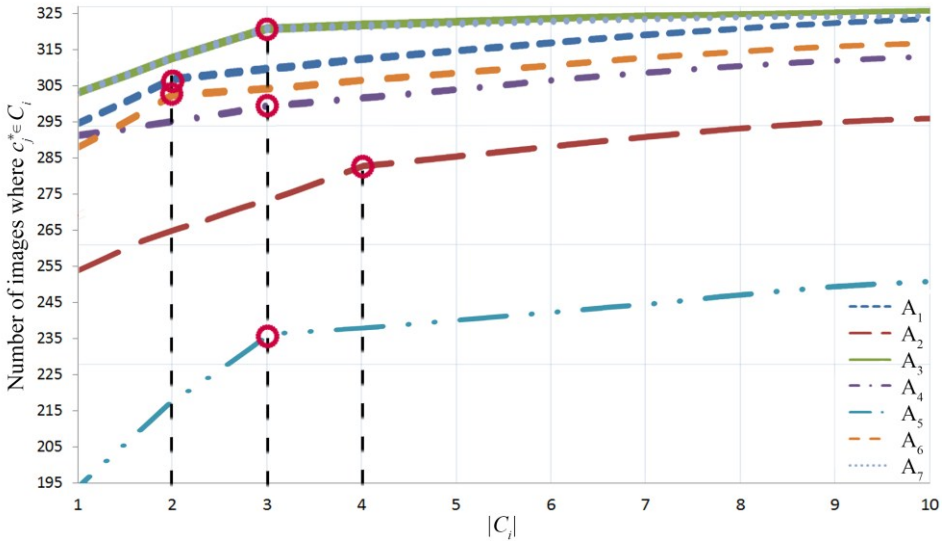

Figure 15: Number of candidates of member algorithms for the OD center. Red circles correspond to the chosen values ($|C_1| = |C_6| = 2, |C_3| = |C_7| = |C_4| = |C_5| = 3, |C_2| = 4$).

3.4.3 Finding clique with maximum total weight

After determining $|C_i|$ for A_i (i=1,2,…,N), we assign weights to candidates as given in definition 5. As a generalization of simple majority voting, we locate the OD as the region containing candidates having largest weight in total. To solve this problem we build up a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{W} \rangle$ as given in definition 6. In this graph, edge connects two adjacent vertices $(v_{i,j}, v_{k,l} \in \mathcal{V})$ if $\|v_{i,j}, v_{k,l}\|_2 < d$, where d=2R. We

note that, this OD geometry constraint ensures that exclusively candidates belonging to different algorithms may fall within a D_R.

So, after having connected those vertices of the graph with edges which fall within a possible region of the object. In this graph, we look for the complete sub-graph (definition 7) with maximum sum of weights of nodes. Namely, we have to solve the problem of looking for the clique $\mathcal{G}(S_i^*)$ of $\mathcal{G}$ with maximum weighted vertices or in other words the problem of looking for the maximum-weighted clique (*MWC*) defined in definition 8. The corresponding graph theoretical problem in this representation, namely *MWC* finding problem is known to be NP-hard. For the solution of this task, we borrow an algorithm [61] which is based on heuristic vertex-coloring and backtrack search. This algorithm provides the *MWC* and works quickly also on dense graphs.

The obtained maximum-weighted clique is considered as the solution of the proposed ensemble-based OD detection system. The final object center p_{res} is determined as the centroid of the maximum-weighted clique, as:

$$\left(x_{p_{res}} = \frac{1}{|MWC|} \sum_{v \in MWC} x_v \, , y_{p_{res}} = \frac{1}{|MWC|} \sum_{v \in MWC} y_v \right), \tag{12}$$

where $|MWC|$ denotes the number of vertices of *MWC*.

3.4.4 Localization of the OD by finding *MWC*

For such a graph and the final result of the *MWC*-based combination of the members see Figure 16. Note that, the input fundus images are the same in Figure 11 and in Figure 16 for the better comparability of the ensemble-based systems.

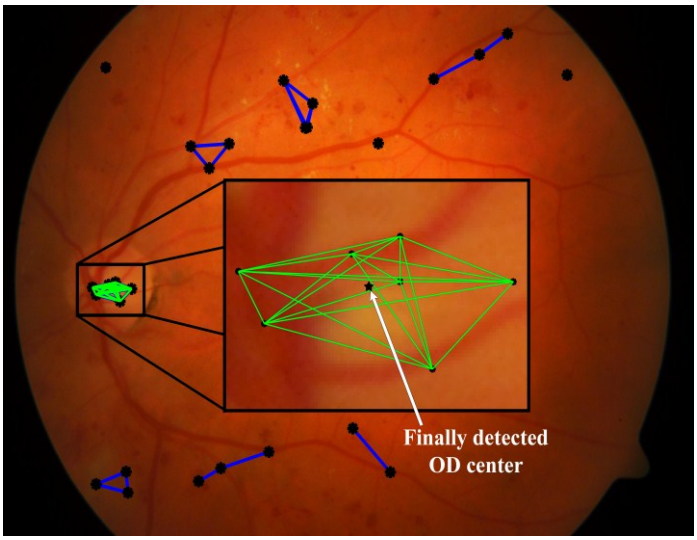

Figure 16: Localization of the OD by finding *MWC*. Candidates (black dots) are connected if their distances are less than 2*R*; the maximum weighted clique is indicated by green lines

Unfortunately, in some cases the maximum-weighted clique-based ensemble of the member algorithms still could not find the correct location e.g. in Figure 17. Moreover, these extracted candidates cannot be used for the detection of the precise region of the object. These are the reasons why we examine further the possibilities to take advantage of all the information provided by the member algorithms about the location of the single object.

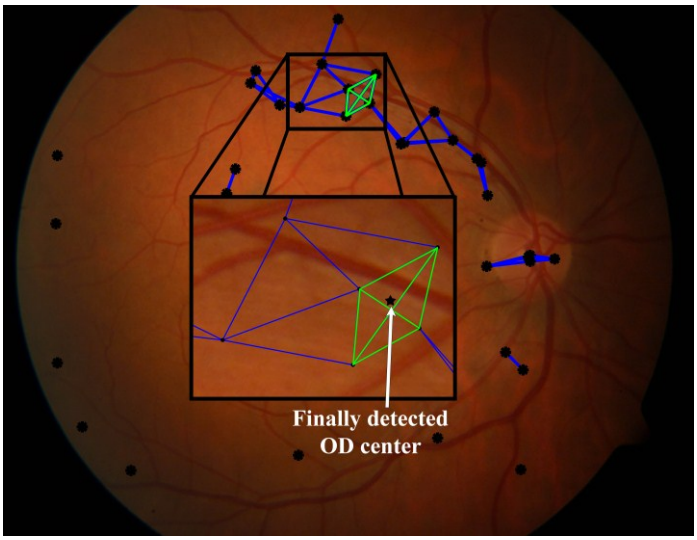

Figure 17: The finally localized OD center lays out the true ODR.

3.5 Probability maps-based ensemble

The basic idea is to utilize as much information as possible about the location of single object. Namely, we expect the member algorithms to assign $Conf_i(p)$ (i=1,2,…,N) to each $p \in I$ indicating their confidence that p is the center of the object. These confidence values together define probability maps for the input image. Now, we introduce some possible approaches to fuse these maps in order to increase the accuracy of single object detection.

The fields of decision making and risk analysis, where information derived from several experts and aggregated by a decision maker, have a well-established literature [62], [63], [64]. In general, the aggregation of information increases the precision of the forecast. In our scenario, we can consider the assigned $Conf_i(p)$ (i=1,2,…,N) to each $p \in$ I as the

opinion of the member algorithms on how probable it is that the given pixel is the center point of the object. Based on the fact of the positive effect of the ensemble, if we consider the algorithms as experts with voting their confidence value and apply aggregation accordingly, the accuracy of the single object detection should improve.

As a short summary concerning the combination of information deriving from experts, basically two approaches are known in the corresponding literature [62]. One of them is based on clearly established mathematical rules, whereas the other one is entrusted to the interaction of an expert, also known as a behavior-based method. In a behavior-based model, the experts contact with the decision maker directly or indirectly to make him/her take their arguments and statements into consideration to reach consensus. In this approach, the quality of the individual experts and the dependencies among them are considered implicitly rather than explicitly. So, we examine only the applicability of strict theoretical approaches which are widely available in the literature from the simple axiomatic methods to the processes requiring different information aggregation models. In the case of single object detection, axiomatic approaches can be applied easily to each $p \in I$ to aggregate the probability values assigned by A_i (i=1,2,…,N) to p. Considering the more complex approaches, we should apply a training set to determine all the necessary parameters to set up the model for the ensemble.

3.5.1 One member algorithm – one probability map

The most of the involved algorithms ($A_1 = OD_{decomp}$, $A_2 = OD_{edge}$, $A_3 = OD_{entropy}$, $A_4 = OD_{classify}$, $A_5 = OD_{fuzzy}$, $A_6 = OD_{hough}$, $A_7 = OD_{cHough}$) basically assign $Conf_i(p)$ (i=1,2,…,N) to each $p \in I$ (as it is introduced in section 3.4.1) but they apply a threshold to select only one location corresponding to the highest value. Thus, we can easily modify them by

omitting their final thresholding step and can consider each $p \in I$ as a $c_{i,j} \in C_i$ equipped with $Conf_i(c_{i,j})$ by A_i (i=1,2,…,N), where $|C_i| = |I|$. In this way A_i (i=1,2,…,N) provides a probability map (PM_i as given in definition 9) about the possible location of the single object.

The PM_i (i=1,2,…,N) can also be considered as probability distribution function, if the following conditions hold:

$$PM_i(p) > 0, for\ all\ p \in I, \text{ for all p} \in I, \tag{13}$$

$$\sum_{p \in I} PM_i(p) = 1 \tag{14}$$

Condition (13) can be fulfilled by assigning a very small probability value $\varepsilon > 0$ to each position, which originally has zero confidence:

$$PM\varepsilon_i(p) = max(PM_i(p), \varepsilon). \tag{15}$$

Finally, to meet condition (14), we perform the following normalization step:

$$PDF_i(p) = \frac{PM\varepsilon_i(p)}{\sum_{p \in I} PM\varepsilon_i(p)}. \tag{16}$$

In this way, the probability maps PM_i (i=1,2,…,N) are transformed to the probability distribution function PDF_i. After we have these *PDF*s, we can fuse them by applying standard axiomatic approaches or more complex aggregation models.

3.5.2 Aggregation based on axiomatic approaches

The product, sum, minimum, maximum of the probability distribution functions are the simplest approaches of the aggregation in the corresponding literature [62], [63]. These techniques are realized by simple arithmetic operations performed between two or more *PDF*s produced by the experts. One of the most commonly used axiomatic approaches is the linear opinion pool published by Stone in [65]. This method calculates the weighted sum of the probability distribution functions PDF_i rendered by A_i (i=1,2,…,N):

$$PDF_{LINOP} = \sum_{i=1}^{N} w_i \, PDF_i \,, \tag{17}$$

where PDF_{LINOP} represents the combined probability distribution and w_i the weights assigned to the experts provided that we have information on their reliability. As a natural condition, $\sum_i w_i = 1$ (i=1,2,…,N) must hold. If $w_i = 1/N$, we have a simple linear combination, otherwise a weighted linear one.

Multiplicative averaging (also known as logarithmic opinion pool) is another commonly used fusion approach [62]. In this case, probability distribution functions are combined as:

$$PDF_{LOGOP} = k \prod_{i=1}^{N} PDF_i^{w_i}, \tag{18}$$

where k is a normalizing constant and w_i represents the same weights as above. If $w_i = 1/N$, (18) returns the geometric mean of the individual distributions. These axiomatic approaches combine the *PDF*s in a simple way with ignoring the quality of the members and the dependencies

among them. Now, we start discussing the Bayesian models of the information aggregation process, which require input regarding bias and dependencies of the experts.

3.5.3 Aggregation based on Bayesian models

In [66] and [67], Morris formally laid the foundation of the Bayesian paradigm to aggregate the information collected from different experts. The Bayesian models operate on the individual probability distribution functions to aggregate them. In the case of single object detection, according to these models the pixels can be considered as the center point (L_1) or not (L_0). Thus, using the Bayes' theorem we assign the ensemble-based probability to each pixel $p \in I$ on whether it is the object center through the following way based on the probability distribution functions PDF_i (i=1,2,…,N):

$$p = p_{man} \text{ if } P(L_1|PDF_1(p), \dots, PDF_N(p)) > P(L_0|PDF_1(p), \dots, PDF_N(p)). \quad (19)$$

As only two cases are possible, we have $P(L_0) = 1 - P(L_1)$ for each pixel. Thus, in our case it is sufficient to determine the probability of L_1 for p with the help of the Bayes' theorem. To each $p \in I$ we calculate the posterior probability in (19) by the help of the Bayes' rule in the following way:

$$P(\mathcal{L}|PDF_1, \dots, PDF_N) = P(PDF_1, \dots, PDF_N|\mathcal{L})P(\mathcal{L}), \quad (20)$$

where $\mathcal{L} \in \Lambda_\alpha$. Note that, $\mathcal{L}$ does not appear in the denominator $P(PDF_1, \dots, PDF_N)$, so this term is applied only for normalization. Thus, it can be omitted by following the general recommendations [64].

The *a priori* probability $P(\mathcal{L})$ in the numerator of (20) can be easily estimated from the training database (TR_1). The calculation of the joint probability distribution function $P(PDF_1, \ldots, PDF_N|\mathcal{L})$ depends on the fact whether the model takes the dependencies of the member algorithms into account or not. In this respect, there are two basic approaches in the relevant literature as discussed next.

3.5.3.1 Naïve Bayes combination

In the first Bayesian approach, let us suppose that the experts do not influence each other, there is no connection between them, so they give their opinion or forecast completely independently. That is, according to this naïve hypothesis, the decision maker manages the information collected from the experts independently. This type of aggregation is known as the Naïve Bayes model and the joint distribution function in (20) can be separated according to the conditionally independent assumption based on the following formula:

$$P(PDF_1, \ldots, PDF_N|\mathcal{L}) = \prod_{i=1}^{N} P(PDF_i|\mathcal{L}). \tag{21}$$

Consequently, the aggregation of the probability distribution functions PDF_i (i=1,2,…,N) can be derived based on (20) and (21) as:

$$PDF_{NB}(p) = P(L_1) \prod_{i=1}^{N} P(PDF_i(p)|L_1), \tag{22}$$

where $P(PDF_i(p)|L_1)$ (i=1,2,…,N) are estimated on the basis of the probability values assigned by the algorithms to the p within the manually segmented object in the images of the training set. Since all the terms of

(22) can be estimated from the training examples, the Naïve Bayes model can be easily constructed and adopted, as well. However, this model ignores the dependencies among the members although, the assumption on the conditional independence of the experts is fulfilled very rarely in practice.

To measure up the dependencies of the member algorithms A_i and A_j (i, j=1,2,…,N), we can calculate the Pearson's correlation coefficient $\rho_{i,j}$ [68]. $\rho_{i,j}$ is calculated pairwise for the member algorithms through comparing all the pairs $\left(PDF_i(p), PDF_j(p)\right), (i, j = 1,2, \dots, N)$ as:

$$\rho_{i,j} = \frac{E\left[\left(PDF_i - \mu(PDF_i)\right)\left(PDF_j - \mu(PDF_j)\right)\right]}{\sigma(PDF_i)\sigma(PDF_j)}, \tag{23}$$

where $\mu(PDF_i)$ stands for the mean, while $\sigma(PDF_i)$ for the standard deviation of the probability map PDF_i, respectively. The coefficients $\rho_{i,j}$ describe the dependencies between A_i and A_j. Non-zero coefficients show dependencies suggesting that the model can be improved further as presented in the next section.

3.5.3.2 Augmented Naïve Bayes combination

In the corresponding literature [69] the problem is well-known that experts do not provide their opinions or forecasts entirely independently from each other. So, combining their input in a way that the decision maker considers the experts independent, have a negative effect on the result. In this case, a Bayesian model is required, which is able to take all the dependencies between the experts into account. To address this issue, the optimal Augmented Naïve Bayes (ANB) model has been suggested [70], where during the learning phase the dependencies of the members

are also incorporated. However, creating such an ANB model is an NP-hard problem [71], so it is recommended to choose an alternative approach which takes the dependencies into consideration, however, does not try to disclose them entirely. One of these models is the Tree Augmented Naïve Bayes (TAN) [70] one, which has the disadvantage that only the most dependent pairs are preserved and the effect of the less dependent experts are omitted. As a trade-off, the complexity of the creation of the TAN model is significantly reduced. Contrarily, the Hidden Naïve Bayes (HNB) model developed by Zhang *et al.* [72] is capable of taking all the dependent experts into account collectively. Thus, the HNB model approximates the precision of the optimal ANB model better, while its time complexity for training is only polynomial.

The basic idea of the HNB model is that a hidden expert (*HE*) is created for each expert which can affect it. Thus, the *i*-th expert depends only on the *i*-th *HE* (HE_i), where HE_i contains all the dependency relations between the *i*-th and the other experts. That is, the joint probability in the numerator of (20) can be calculated by the HNB model considering the dependencies among the experts as:

$$P(PDF_1, \dots, PDF_N|\mathcal{L}) = \prod_{i=1}^{N} P(PDF_i|HE_i, \mathcal{L}), \tag{24}$$

where

$$P(PDF_i|HE_i, \mathcal{L}) = \sum_{j=1, j\neq i}^{N} W_{ij}\, P\left(PDF_i\middle|PDF_j, \mathcal{L}\right) \tag{25}$$

with $\sum_{j=1,j\neq i}^{N} W_{ij}$=1. As it can be seen, HE_i is the hidden expert of PDF_i (i=1,2,…,N) and is basically a mixture of the weighted dependencies with other experts. The weights W_{ij} (i, j=1,2,…,N, $i \neq j$) are determined using

the training set based on the conditional mutual information (*CMI*) of PDF_i and PDF_j:

$$W_{ij} = \frac{CMI\left(PDF_i; PDF_j \middle| \mathcal{L}\right)}{\sum_{j=1, j\neq i}^{N} CMI\left(PDF_i; PDF_j \middle| \mathcal{L}\right)}, \tag{26}$$

where *CMI* is generally calculated as:

$$CMI(\mathrm{A}; \mathrm{B}|\Gamma) = \sum_{\alpha,\beta,\gamma} P(\alpha, \beta, \gamma) log \frac{P(\alpha, \beta|\gamma)}{P(\alpha|\gamma)(\beta|\gamma)}. \tag{27}$$

Using the weights W_{ij} (*i, j*=1,2,…,*N*, $i \neq j$), the hidden experts HE_i (*i*=1,2,…,*N*) can be determined. Thus, the HNB model incorporates all the dependencies among experts similarly to the optimal ANB one. However, the time complexity of the training phase of HNB is $\mathcal{O}(tN^2 + kN^2v^2)$, where *t* is the number of training pixels of the training images, *N* is the number of algorithms, *k* is the number of classes and *v* is the average number of values for an attribute.

After defining the weights W_{ij} (*i, j*=1,2,…,*N*, $i \neq j$), the aggregation of the probability maps PDF_i (*i*=1,2,…,*N*) can be executed via the HNB model on the basis of the following formula:

$$PDF_{HNB}(p) = P(L_1) \prod_{i=1}^{N} P(PDF_i(p)|HE_i(p), L_1). \tag{28}$$

3.5.4 Localization of the OD by the fusion of probability maps

Till this point of this section, we introduced general, ensemble-based frameworks for single object detection, when we have more than one

member algorithms which can generate probability maps to locate the object. Now we show how to apply these approaches for OD detection and observe their performance considering the accuracy of localization. For visual and precise comparison of the fusion of probability maps-based approaches, we aggregate PM_i of the $A_1, A_2, A_3, A_4, A_5, A_6$ and A_7 for the sample image of Figure 17 (as it can be seen in Figure 18).

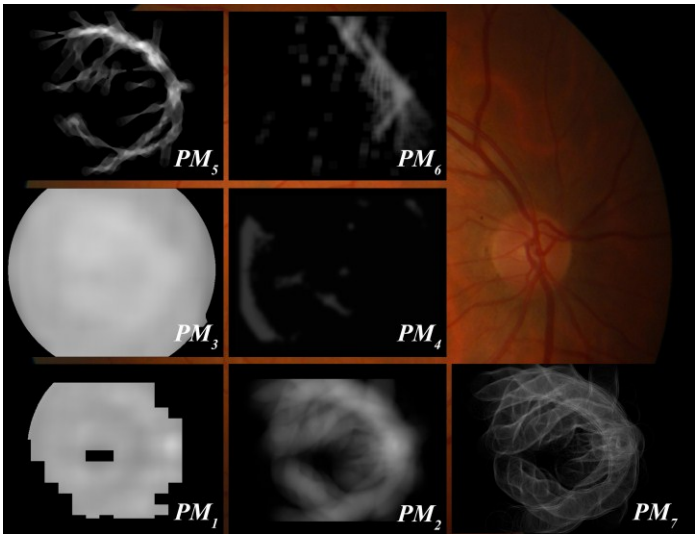

Figure 18: PM_i of A_i (i=1,2,…,N) for the sample image of Figure 17.

As we have discussed in section 3.5.1, the proposed ensemble approaches can be applied if the *PM*s fulfill conditions (13) and (14). For this aim, the *PM*s are transformed to probability distribution functions (*PDF*s) by formulas (15) and (16). In Figure 19, we can see a visual representation of the *PDF*s derived from the *PM*s of Figure 18.

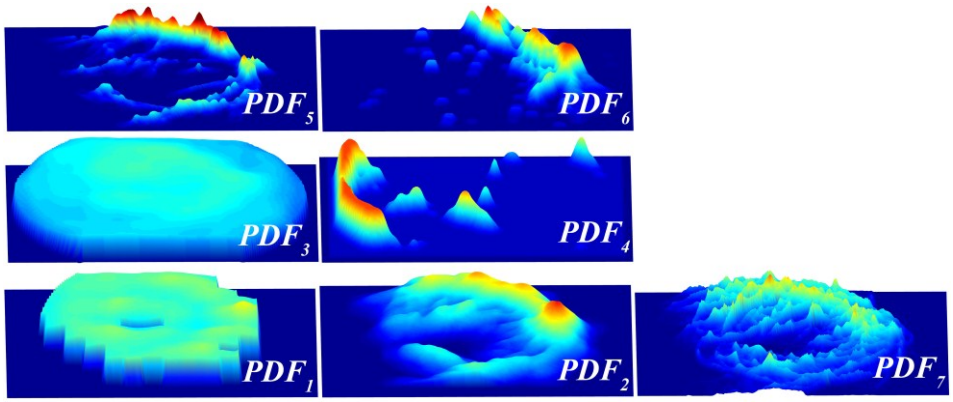

Figure 19: The probability distribution functions of A_i (i=1,2,…,N).

After constructing the *PDF*s, they can be fused by applying the standard axiomatic or Bayesian model-based approaches. For axiomatic ensemble of *PDF*s the weights w_i (see (17) and (18)) are calculated from the individual accuracies as follows:

$$w_i = \frac{Acc_i}{\sum_{i=1}^{N} Acc_i}, \tag{29}$$

where Acc_i denotes the individual accuracy of A_i on the TR_1 training set. After making tests on a training set, we adjust the following weights: w_1=0.16, w_2=0.18, w_3=0.17, w_4=0.16, w_5=0.04, w_6=0.16, w_7=0.13. The result of the combination of PDF_i (i=1,2,…,N) by weighted linear opinion pool and weighted logarithmic opinion pool can be seen in Figure 21(a) and 21(b), respectively.

Now we turn to the Bayesian model for OD detection. During the training stage, we determine the probability of OD center pixels among all the pixels of the training images of TR_1. However, there is only one OD center point in the image, and considering the number of all the

image pixels, *a priori* $P(L_1)$ is a very small value. Since L_1 would be very under-represented in this way in a training dataset, we interpret L_1 in a wider sense. Namely, we let L_1 represent not only the case when $p = p_{man}$, but also when p falls inside the OD region ($p \in ODR$). In other words, we do not restrict our attention to the center, but we accept any OD pixels. Note that, in this way $P(L_1)$ becomes sufficiently large, and from now we work in this extended context. Besides the *a priori* probability $P(L_1)$, the conditional probabilities $P(PDF_i(p)|L_1)$ (i=1,2,…,N) are also calculated inside and outside the region of the OD.

To collect the training dataset for Bayesian models, we have calculated PDF_i (i=1,2,…,N) for all the 327 images of TR_1. Then, we have randomly selected 400 000 pixels from these images together with their corresponding PDF_i (i=1,2,…,7) values. In this way, 4% of the pixels belonged to OD regions in the training set. Figure 20 shows the distribution of the probability values inside and outside the region of the OD for these selected pixels.

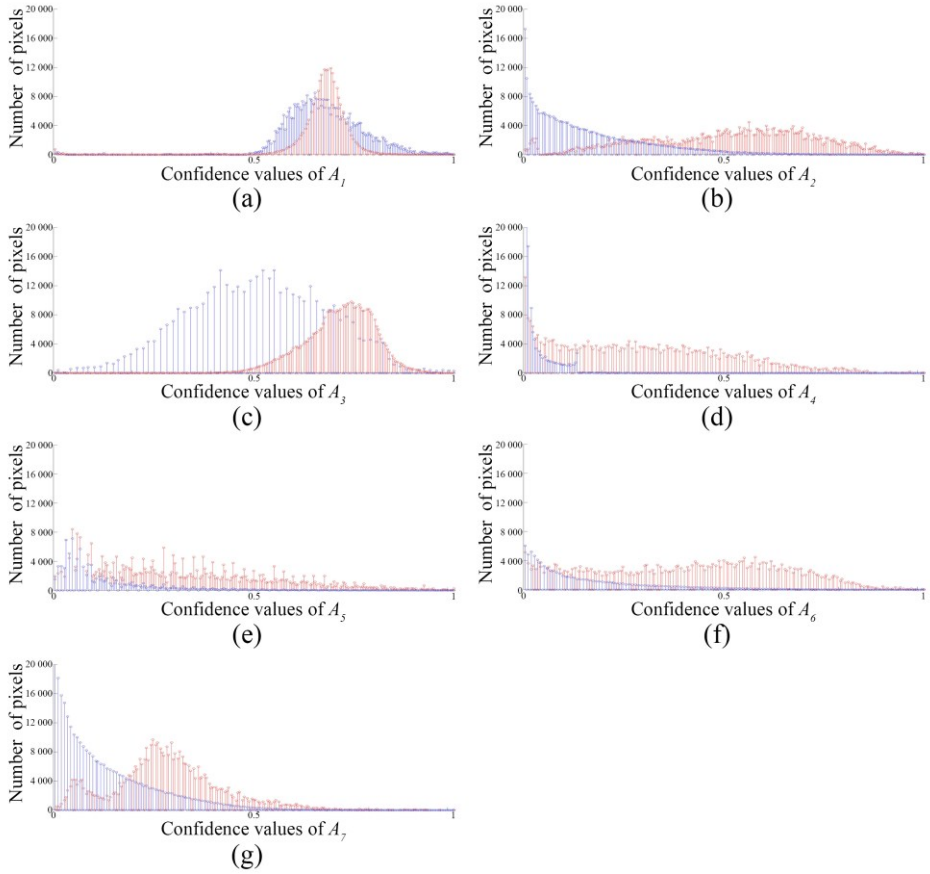

Figure 20: Distribution of probability values provided by (a) A_1, (b) A_2, (c) A_3, (d) A_4, (e) A_5, (f) A_6 and (g) A_7 for images of TR_1. Red/Blue shows the values inside/outside the OD region.

As we have mentioned in section 3.5.3.1, the assumption on the conditional independence of the experts is fulfilled very rarely in practice. This assumption does not hold for the involved OD detectors either. The Pearson's correlation coefficients $\rho_{i,j}$ for all the possible pairs of member algorithms can be observed in Table 2.

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
A_1	1.0000	0.6034	0.6339	0.1122	0.3235	0.2645	0.5634
A_2	0.6034	1.0000	0.6612	0.4070	0.6070	0.6150	0.8172
A_3	0.6339	0.6612	1.0000	0.2094	0.3934	0.3607	0.6437
A_4	0.1122	0.4070	0.2094	1.0000	0.1928	0.3733	0.2427
A_5	0.3235	0.6070	0.3934	0.1928	1.0000	0.5553	0.4733
A_6	0.2645	0.6150	0.3607	0.3733	0.5553	1.0000	0.4793
A_7	0.5634	0.8172	0.6437	0.2427	0.4733	0.4793	1.0000

Table 2: Pearson's correlation coefficients of A_i (i=1,2,…,N).

A smaller (close to 0) correlation value corresponds to smaller dependency of the given algorithms. For instance, A_1 and A_4 seem to be the most diverse algorithms compared to each other regarding this measure. There are no zeros in Table 2 showing the trivial fact that the members cannot be completely independent. Thus, we can apply the HNB model which takes the dependencies among the detectors also into consideration. A sample result for the combination based on the Naïve Bayes model and HNB can be seen in Figure 21(c) and 21(d), respectively.

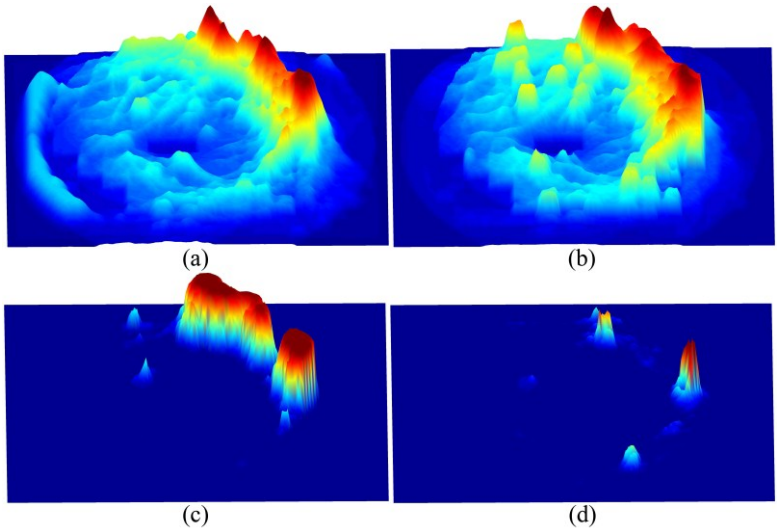

Figure 21: Results of combination of PDFs by (a) linear opinion pool, (b) logarithmic opinion pool, (c) Naïve Bayes model, (d) HNB model.

In all the images of Figure 21, very high probability values can be observed in the region of the OD, so its location can be found with a simple extra step. Namely, the center of D_R is matched on each p of the resulted ensemble *PDF* image and that p is selected as p_{res}, where we find the maximum sum of *PDF* values for the pixels within the matched D_R.

If we compare the results of the axiomatic models (Figures. 21(a) and 21(b)) with the Bayesian ones (Figure 21(c) and 21(d)), we can see significant difference between the areas, where the OD is detected with lower probability. There are high peaks at the possible OD locations, but the rest of Figure 21(c) and 21(d) are flat because of the more complex aggregation models. Furthermore, as we consider L_1 in a wider sense by letting it represent the event $p \in ODR$, the resulted PDFs of the Bayesian models show high probability also at the pixels falling within the OD region not just at its center. Thus, the automated system can determine

OD region (ODR_{res}) if it considers the area connecting to the peak found by the template D_R matching process. If D_R expands this region, their union is considered as the segmentation of the OD region, as it can be seen also in Figure 22.

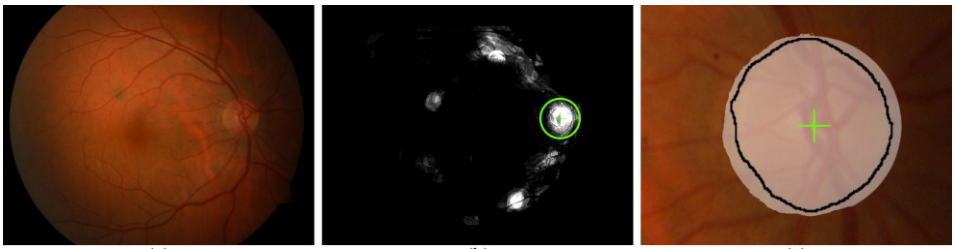

Figure 22: (a) I, (b) result of the combination by HNB model, (c) ODR_{res} (white region) with marking p_{res} (green cross) and manually drawn ODR (black line).

3.6 Experimental results

For the measurability of the accuracies of the introduced ensemble-based single object localization approaches, manually selected ODR are supplied for the images of DIARETDB0, DIARETDB1, DRIVE and MESSIDOR datasets. As we mentioned in section 3.1, we consider three error functions for measuring the accuracy according to the definition 11, 12 and 13.

The main motivation of our contributions is to see how including more and more information from the members can improve the performance of the ensemble. Tables 3, 4, 5, 6 compare the respective performances of the member algorithms ($A_1, A_2, A_3, A_4, A_5, A_6$ and A_7), OD detection based on simple majority voting, finding maximal weighted clique, combining probability maps by the axiomatic approaches and by the Bayesian models on the DIARETDB0, DIARETDB1, DRIVE and

MESSIDOR datasets. We note that, Err_2 depends on the resolution of the input image, which explains why the average Err_2 is lower on the DRIVE dataset than on the others. For simplicity, we have normalized Err_2 for the highest resolution (2304×1536 pixels) regarding the images of the MESSIDOR dataset.

As we can see, the highest accuracy can be achieved, when the aggregation model uses the largest amount of information about the OD location provided by the algorithms and the dependencies between them.

	DIRAETDB0		DIRAETDB1	
	PPV	Err_2	PPV	Err_2
A_1	86.92%	26.45	91.01%	27.64
A_2	90.77%	20.06	91.01%	17.72
A_3	87.69%	36.91	85.39%	36.73
A_4	85.38%	20.35	80.91%	21.79
A_5	30.11%	45.80	40.45%	44.23
A_6	84.62%	42.49	84.27%	47.66
A_7	63.08%	39.04	68.54%	37.49
Majority voting based ensemble	94.62%	15.68	94.38%	17.40
Finding *MWC* based ensemble	95.38%	9.38	95.51%	9.19
Combining *MP*s by **weighted *LinOP***	93.08%	16.76	95.51%	16.14
Combining *MP*s by **weighted *LogOP***	91.54%	24.53	93.26%	24.31
Combining *MP*s by **Naïve Bayes model**	96.15%	22.36	96.63%	21.24
Combining *MP*s by **HNB model**	98.46%	16.86	98.88%	14.49

Table 3: Comparative results on the dataset DIRAETDB0 and DIRAETDB1.

	DRIVE		MESSIDOR	
	PPV	Err_2	PPV	Err_2
A_1	80.00%	17.05	93.49%	27.87
A_2	92.50%	11.48	93.55%	32.73
A_3	85.00%	18.7	91.04%	36.53
A_4	70.00%	9.65	64.78%	25.78
A_5	82.50%	15.41	39.49%	64.14
A_6	72.50%	17.62	90.29%	57.99
A_7	92.50%	14.79	64.57%	41.67
Majority voting based ensemble	100.00%	11.26	95.29%	14.41
Finding *MWC* based ensemble	100.00%	6.08	96.33%	14.78
Combining *MP*s by **weighted *LinOP***	100.00%	12.12	97.33%	39.32
Combining *MP*s by **weighted *LogOP***	100.00%	12.59	97.33%	36.04
Combining *MP*s by **Naïve Bayes model**	100.00%	11.62	97.65%	24.45
Combining *MP*s by **HNB model**	100.00%	9.12	98.33%	21.12

Table 4: Comparative results on the dataset DRIVE and MESSIDOR.

None of the member algorithms involved in the study is suitable for detecting the exact ODR_{res}. Therefore, we did not evaluate the individual OD detectors for Err_3. We merely notice that the segmented OD region based on the combined probability maps by the NB model against the manually drawn OD region achieved the following corresponding precisions: $Err_3 = 0.66$ on DIARETDB0, $Err_3 = 0.71$ on DIARETDB1 and $Err_3 = 0.66$ on DRIVE. When the combination was made by the

HNB model, then we got $Err_3 = 0.84/0.87/0.73$ for DIARETDB0, DIARETDB1, and DRIVE, respectively.

Although, many state-of-the-art algorithms (*e.g.* [7], [73], [74]) are available for the segmentation of the OD boundary which can outperform the proposed ensemble of the algorithms regarding Err_3, we have to take into consideration the fact that none of the member algorithms can detect the region of the OD. However, their combination is suitable for this task as it can be seen in Figure 23.

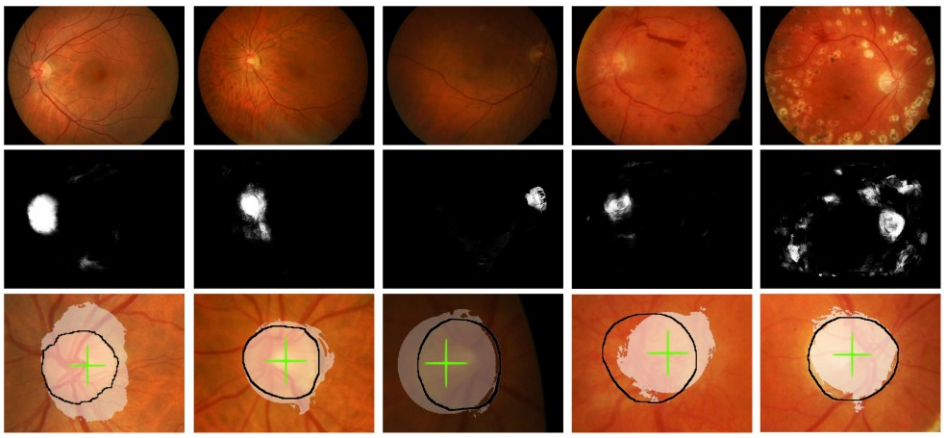

Figure 23: Results of the segmentation of the OD region based on aggregated probability maps using the HNB model.

In this figure, three rows contain images as a follows: sample images; the results of aggregation by HNB and a closer look of the finally segmented ODR_{res}. The black boundary in the third row shows the manually segmented OD regions, the white patch the automatically segmented ones, and the green cross the p_{res}.

3.7 Discussion

As our results suggest, the precision achieved by the member algorithms can be increased, if an ensemble of them is computed based on majority voting or finding maximum-weighted clique in weighted graph. Furthermore, as it can be observed also in Tables 3, and 4, if all the information about the location of the single object is considered, the accuracy of the detection also increases with using the appropriate fusion model. The accuracy can be further increased by disclosing the dependent behavior of the member algorithms. Namely, if we incorporate information about the dependencies of the members into the model, they are less able to influence negatively the final result when they are wrong.

The proposed ensemble-based methods allow the application of any member algorithms, so they can be considered as a general framework for composing ensembles. The only requirement is that any possible single object detector algorithm should be able to assign probability values to all the pixels of the input image for suggesting the center of the object. If the system is extended by new algorithms that are based on different principles, the performance of the ensemble can be expected to improve further.

All the seven involved OD detector algorithms in the demonstration of the single object detection task were reimplemented in Matlab environment by following the respective instructions given in [8], [9], [10], [11], [12], [13] and we used a single core 2.4GHz CPU with 2GB memory for testing. The computational times of the original algorithms are the same with a negligible difference as their variants which are capable of assigning probability values to each pixel with producing the maps PDF_i (i=1,2,…,N) for an input image. The reason for the same demands of computations is that the original algorithms also calculate confidence values for each pixel and select the maximum to return the

center of the OD. The computational times of the member algorithms are OD_{decomp} [8] – 190 s, OD_{edge} [8] – 237 s, $OD_{entropy}$ [9] – 93 s, $OD_{classify}$ [10] – 214 s, OD_{fuzzy} [11] – 5.57 s, OD_{hough} [12] – 1.36 s, OD_{cHough} [13] – 4.76 s per image, respectively. The majority voting-based combination of the outputs of individual members is done in 7.2 ms while the finding of maximum-weighted clique in the weighted graph is done in 57 ms per image. The aggregation of the maps PDF_i (i=1,2,…,N) is done in 40 ms per image for axiomatic approaches as Matlab is efficient in matrix operations. The computational times for the Naïve Bayes model-based aggregation approaches also depend on the size of the input image. Excluding the training phase, we have measured the following times: simple Naïve Bayes model – 220 ms, TAN – 200 ms and HNB – 250 ms per image. For the applied TR_1 training dataset, the number of operation to train the HNB model is 25 972 450.

4 Exudate detection framework

An automatized system developed for DR screening has to be able to detect the symptoms caused by DR at an early stage, beyond being able to localize anatomical parts of the eye such as the OD, the macula and the network of capillaries. Several articles have been published [20], [21], [22], [23], [24], [25], [26], [27], [28] and [29] in connection with the detection of exudates. Some of these are based on the function of the grayscale morphology, while others are based on classification for the segmentation of exudates.

In this section, we introduce a framework created for the detection of objects appearing in image multiple times (e.g. exudates). This framework involves two approaches which are commonly applied in the relevant literature. Namely, it uses grayscale morphology and applies machine learning-based methods for classification. This complex method can be divided into three main stages as it can be seen in Figure 24: candidate extraction, relevant features extraction and the labeling of candidates as true or false exudates.

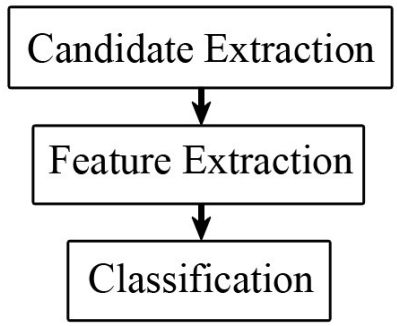

Figure 24: Steps of the proposed exudate detection framework.

For candidate extraction, a grayscale morphology-based method is applied to identify the possible regions which probably contain these

bright lesions. The result of the candidate extraction is a binary image which can be used for determination of the relevant pixels which may be in connection with exudates. Considering these parts of the fundus image, we extract the appropriate pixel/region-wise features and finally a classifier, using these features, decides whether these parts can be considered as exudates or not. The method is tested on publicly available databases to measure the accuracy of the detection of exudate regions at image-level. Regarding the experimental results, our complex framework outperformed the state-of-the-art exudate detectors involved in the study.

In the corresponding literature, a large number of exudate detection algorithms have been proposed. In general, we can divide these approaches into two main groups. The first group contains algorithms based on grayscale morphology [20], [21], [22], while the second one consists of methods considering pixel/region-wise classification [23], [24], [25], [26], [27]. Furthermore, we can find some special approaches (e.g. [28], [29]) falling out of these groups. Walter *et al.* [20] proposed a method using morphological closing to eliminate blood vessels, and then the local standard deviation is calculated and thresholded to find the candidate regions. Finally, morphological reconstruction is applied to find the contours of the exudates. Sopharak *et al.* [21] introduced a technique which is based on optimally adjusted morphological operations. Since the OD is also a bright patch, it is eliminated and Otsu's algorithm is used for thresholding to locate regions with high intensities. Welfer *et al.* [22] applied morphological operations and H-maxima transform after contrast enhancement on the channel L in the color space CIE 1976 L*u*v*. Sopharak *et al.* [23] proposed a method using Fuzzy C-Means Clustering so that it can be determined whether a pixel belongs to an exudate or not. Then, morphological operations are applied to refine the segmentation result. Sopharak *et al.* [24] designed an algorithm for exudate detection, which applies pixel-wise classification. Namely, a Naïve Bayes classifier

sorts out each pixel based on five extracted features. The method proposed in [25] also considers pixel-wise classification, but the training database is defined for each analyzed image separately. That is, the algorithm first detects small isolated exudates and uses those pixels as a positive training set. Then, the rest of the image pixels are classified based on their corresponding properties. Niemeijer *et al.* [26] proposed a multi-level classification approach for segmentation of pixels which belongs to bright lesions with high probability. These pixels are grouped into clusters and the clusters are labeled as exudates, cotton-wool spots or drusens. Jaafar *et al.* [28] proposed exudate detection based on a split-and-merge technique. This algorithm splits the images into disjoint regions first, and merges them based on local variance afterwards. Finally, a histogram-based adaptive thresholding is applied to each merged region. Ali *et al.* [29] proposed an atlas-based method to detect exudates.

The application of extracting candidates to decrease the time of computation, in the case of detection of another lesions caused by DR, is well-known technique as it is used in [17], [25]. Basic concepts of candidate extraction is to gain those pixels/regions by a simple and quick operation which connects the desired object with high probability thus, techniques demanding more complex operations are applied exclusively on these extracted parts of the image. In this manner, the computation time is significantly decreased. Thus, the proposed exudate detector framework use grayscale morphology-based algorithm as candidate extractor. Following this, a classifier algorithm decides whether these extracted candidates can be considered as true exudates or not.

On the one hand, we use the characteristics of the extracted pixels for classification; a classifier makes a decision about each pixel whether those correspond to the pixels of the wanted object. We collected all characteristics of the exudate at pixel-level for the classification and we

selected those which were the most suitable for classification. The computational time of the evaluation at pixel-level is significantly reduced through the application of candidate extractor and the precision of the proposed method outperforms the state-of-the-art algorithms involved in the comparison regarding accuracy. The corresponding result is also published in [35].

On the other hand, the result image contained plenty of false positive pixels due to the fact that the evaluation is done at pixel-level. Thus, as an improvement, the extracted candidates are evaluated at region-level instead of pixel-level by the classifier. Owing to this modification, the algorithm assigns a label to the regions on the basis of descriptors extracted from them. The features used for classification can be derived, amongst others, from the contours of the regions. Because of this, the accuracy of the classifier depends on the precision of the contour of the regions. Thus, we examine some active contour-based algorithms and use them at first occasion in the literature in order to define the exact contour of the region for the detection of exudates. On the basis of the evaluation the descriptors at region-level, calculated from the precise contour, enabled the classifier to make more precise classifications. Furthermore, the number of false positive pixels in the result image is reduced as well. The corresponding result is also published in [36], [37] and [38].

As final improvement, we try to extract the most precise and descriptive features of exudates for classification. For this aim, we fuse the commonly used image enhancement algorithms in order to increase the precision of the segmented contour of the candidate regions. The features describing the regions become more reliable and the classifier can reach higher accuracy when it selects the true exudates from all. The corresponding result is also published in [39] and [40].

The rest of this section is organized as follows. In section 4.1, we introduce some concepts regarding the multiple objects detection and we

give a brief overview about the applied image enhancement methods, grayscale morphology-based method and active contour-based method in section 4.2. In section 4.3, we give an overview about the proposed exudate detection framework and show how we can use active contour-based method for precise boundary detection and which regions-wise descriptors are recommended for classification. Section 4.4 is dedicated to comparative experimental results regarding some other state-of-the-art exudate detection methods. Finally, some discussions are given in section 4.5.

4.1 Concepts in the fields of multiple objects detection

In this section, we provide some basic concepts for multiple objects detection in digital image for the easier and precise reference in the latter parts of the section 4.

We can consider the problem of the exudate detection as a detection of multiple objects where the key question is the presence of the objects in the image. Namely, we have to determine that these objects are in the image or not and this way the proposed method can be integrated into an automatic DR screening system to detect the exudates in fundus images.

As it can be seen in Figure 24, the proposed exudate detection framework starts with a candidate extractor step, which gains those parts of the image which probably contain exudates. We emphasize the similarities between the introduced terms and those used in the literature of image processing [75].

Definition 14: *Let I be a digital image. A candidate extractor algorithm provides a binary image $\mathcal{C}_I$ for I. The pixel values $\mathcal{C}_I(p) = 1$ if the p is a pixel of exudate candidate otherwise $\mathcal{C}_I(p) = 0$.*

Definition 15: *Let p, q be any two pixels of I so that $\mathcal{C}_I(p) = \mathcal{C}_I(q) = 1$. p and q are called 8-neighbors if their coordinates fulfill the following equation:*

$$\max\{|x_p - x_q|, |y_p - y_q|\} = 1. \tag{30}$$

Definition 16: *Let p, q be any two pixels of I so that $\mathcal{C}_I(p) = \mathcal{C}_I(q) = 1$. There are path from pixel p to q if there are a sequence of points $p_0, p_1, \ldots p_n$ where $p_0 = p$ and $p_n = q$ and p_{i+1} is a 8-neighbors of p_i (i=1,…,n-1).*

Definition 17: *A set of pixels can be called as region R if it contains a path between any pair of its pixels, all of whose pixels also belong to the set.*

Definition 18: *The border B (boundary, contour) of a region R is the set of pixels within the region that have one or more neighbors outside R.*

Definition 19: *Let $\mathfrak{R}_I$ be a set which contains all of the regions contained by $\mathcal{C}_I$ and $R_i \in \mathfrak{R}_I$ is a candidate region with boundary B_i (i=1, 2,...,$|\mathfrak{R}_I|$) extracted from input image I by a candidate extractor.*

To improve the descriptive nature of the pixel-wise and region-wise features extracted from R_i (i=1,2,…,$|\mathfrak{R}_I|$), we apply some image enhancement methods which can improve the contrast and the information content of the input image regarding exudates.

Definition 20: *Let I and I' be two digital images. I' is an enhancement version of I if there are exist image enhancement method which can transform I to I'. The set of all generated enhancement version of I is denoted by $\mathcal{J}$. Note that, we also let $I \in \mathcal{J}$.*

Moreover, we propose the application of an active contour method to determine the precise boundary of the candidate region R_i (i=1,2,…,$|\mathfrak{R}_I|$) as a proper classification is necessary, as it will be shown in section 4.3.2.

Definition 21: *Let ACM be an active contour method which can determine a precise contour $B_{i,j}$ of R_i (i=1,2,...,$|\mathfrak{R}_I|$) by minimizing its energy function regarding an enhancement version of $I'_j \in \mathcal{J}$ (j=1,2,...,$|\mathcal{J}|$).*

After having enhanced the I and determined the precise contour of R_i (i=1,2,…,$|\mathfrak{R}_I|$) we collect features into a vector by considering the properties of the pixels within R_i and region-wise attributes of R_i (i=1, 2,…,$|\mathcal{C}|$).

Definition 22: *Let H be a feature extractor which assigns a feature vector for each $p \epsilon R_i$ (i=1,2,...,$|\mathfrak{R}_I|$). When p, q are any two pixels of R_i (i=1,2,...,$|\mathfrak{R}_I|$) and $H(p) = H(q)$ we call the elements of vector as region-wise features and the region-wise feature vector is denotes by $H(R_i)$ (i=1,2,...,$|\mathfrak{R}_I|$) otherwise they are called pixel-wise features.*

Definition 23: *Let Λ be the set of labels. $R_i \epsilon \mathfrak{R}_I$ (i=1,2,…,$|\mathfrak{R}_I|$) is labeled with $\ell_1/\ell_0 \in \Lambda$ by classifier if R_i is a true/false exudate.*

Definition 24: *Let $\mathcal{RES}_I$ be a set of regions $R_i \epsilon \mathfrak{R}_I$ (i=1,2,…,$|\mathfrak{R}_I|$) which are labeled with $\ell_1 \in \Lambda$ by classifier. We note that I is considered as healthy image if $|\mathcal{RES}_I| = 0$, otherwise it is unhealthy according to the proposed exudate detection framework.*

We evaluate our proposed exudate detection method at image-level which means that we consider the following classical measures used in

the relevant literature [20], [21], [22], [23], [24], [25], [26], [27], [28], [29] also.

Definition 25: *Let $\mathcal{REF}_I$ be a set of regions which are in the manually drawn binary image containing the exudate regions found by experts in image I. For the screening of I by the proposed exudate detection framework regarding $\mathcal{REF}_I$ we apply the following:*

- *I is true positive (TP_2) if $|\mathcal{RES}_I| > 0$ and $|\mathcal{REF}_I| > 0$.*
- *I is false positive (FP_2) if $|\mathcal{RES}_I| > 0$ and $|\mathcal{REF}_I| = 0$.*
- *I is true negative (TN_2) if $|\mathcal{RES}_I| = 0$ and $|\mathcal{REF}_I| = 0$.*
- *I is false negative (FN_2) if $|\mathcal{RES}_I| = 0$ and $|\mathcal{REF}_I| > 0$.*

Definition 26: *For measuring the performance of the proposed exudate detector within a given dataset we can use the following descriptors:*

- *sensitivity (SEN):*

$$SEN = \frac{\#\ of\ TP_2}{\#\ of\ TP_2 + \#\ of\ FP_2}, \tag{31}$$

- *specificity (SPE):*

$$SPE = \frac{\#\ of\ TN_2}{\#\ of\ FN_2 + \#\ of\ TN_2}, \tag{32}$$

- *false positive rate (FPR):*

$$FPR = 1 - SPE, \tag{33}$$

- *accuracy (ACC):*

$$ACC = \frac{\#\ of\ TP_2 + \#\ of\ TN_2}{\#\ of\ TP_2 + \#\ of\ FP_2 + \#\ of\ TN_2 + \#\ of\ FN_2}, \tag{34}$$

where $\#\ of\ TP_2/FP_2/TN_2/FN_2$ *means the number of* $TP_2/FP_2/TN_2/FN_2$ *in a given test set.*

4.2 Applied algorithms from the relevant literature

In this section, we give a brief overview about the applied existing algorithms as candidate extractor [20], image enhancement methods [11], [25], [56], [76], [78], [79], [80] and the active contour methods [81], [82], [85].

4.2.1 Candidate extractor

Walter *et al.* [20] consider high local contrast and intensity in I_G as the most important properties of exudates. Since there is also a high contrast between the vessels and the background, the method eliminates the vascular system by a simple grayscale morphological closing. In the vessel-free image, the local variation is calculated at each pixel inside a window and the regions with low local variations are excluded. The OD is also eliminated from the image, because it is similar to exudates regarding brightness and contrast. The remaining bright regions are excluded from

the original image and the holes are filled in by morphological reconstruction. The result looks like a healthy image without bright lesions, so when it is subtracted from the original one, the difference image contains the bright exudate candidates only. Finally, thresholding is performed on the remaining candidates to try to eliminate false exudate pixels. This algorithm has three parameters: the size of the window, the contrast threshold and the brightness threshold, which are set as proposed by Walter *et al.* [20]. The boundaries of these extracted regions will be used as initial positions for the active contour segmentation method.

4.2.2 Image enhancement methods

4.2.2.1 Chromaticity Normalization

Chromaticity normalization [76] can normalize the I_G according to the portion of the green among the colors as:

$$I_{CN} = \frac{I_G}{I_R + I_G + I_B}, \tag{35}$$

where I_{CN} is the enhanced intensity channel. This method is usually applied, when the scene is captured by a camcorder and the illumination of the object is not uniform in a video. In our case, it is suited to reduce the bright reflection of retinal images of young patients.

4.2.2.2 Contrast Enhancement

A robust contrast enhancement method has been proposed by Sanchez *et al.* [25]. To follow the instructions of the authors to enhance the image for further analysis, we convert the image from the color space RGB to the

YIQ and modify the channel *Y* by the weighted sum of the channels (*Y, I, Q*) as:

$$Y_{mod} = aY + bI + cQ, \tag{36}$$

where a, b and c are parameters which depend on the characteristics of the image. They are selected $(a = 1.5, b = -1, c = -1)$ to reduce the local luminance variability throughout Y_{mod} and increase the mean contrast levels. After this modification, we convert $Y_{mod}IQ$ back to the color space RGB and it causes the $R_{mod}G_{mod}B_{mod}$. R_{mod} channel of the resulting image is considered as the enhancement version I_{CE} of the input image where the bright regions become brighter, while the dark ones darker.

4.2.2.3 Contrast-Limited Adaptive Histogram Equalization

Contrast-limited adaptive histogram equalization (CLAHE) [56] improves the contrast of the image locally. The sufficiently high contrast of the fundus image is very important, since besides high intensity, the contrast is another useful feature in exudate detection. After applying CLAHE to I_G, the exudates can be better distinguished from the background on the resulting image I_{CL}.

4.2.2.4 Grey-World Normalization

Grey-world normalization [78] divides each color channel by its respective average intensity, so it is suitable to suppress shining along temporal arcades. As I_G contains the largest amount of information about the lesions and anatomical parts, we consider only the gray world normalization of I_G as:

$$I_{GN} = \frac{I_G}{\bar{I}_G}, \tag{37}$$

where I_{GN} is the resulting enhanced channel and $\bar{I}_G$ is the average intensity of the green channel.

4.2.2.5 Illumination Correction

The illumination is usually non-uniform in retinal images due to the variation of the retinal tissues and the spherical shape of the eye. To suppress non-uniform illumination, we apply illumination correction [79]. To perform this image enhancement technique, a spatially large median filter is applied to I_G. To get the corrected image I_{IC}, the blurred image is subtracted from the original one.

4.2.2.6 Illumination Equalization

Besides illumination correction, we also simulate uniform illumination by using illumination equalization [11]. The adjusted intensity values $I_{IE}(p)$ are derived for each pixel as:

$$I_{IE}(p) = I_G(p) + m - \frac{1}{|p_{n\times n}|} \sum_{q \in p_{n\times n}} I_G(p), \tag{38}$$

where m is the desired average intensity and $p_{n\times n}$ is the $n \times n$ pixels local neighborhood of the pixel $p \in I$.

4.2.2.7 White Top-Hat Transformation

White top-hat transformation [80] is a morphological operator designed for extracting bright regions from I_G. Since the opening operator realizes an erosion followed by a dilation, the darker regions will suppress the brighter ones on the opened image. When this relatively dark image is subtracted from the original one, the intensity peaks are enhanced and the exudates can be distinguished better from the background on the resulting image I_{WT}.

4.2.3 Active contour methods

4.2.3.1 Gradient Vector Flow and snake model

A traditional snake [77] is a curve $\mathrm{X}(s) = [x(s), y(s)]$, $s \in [0,1]$, that moves through the spatial domain of an image to minimize the energy function:

$$E_{snake} = \int_0^1 \frac{1}{2}(a|\mathrm{X}'(s)|^2 + b|\mathrm{X}''(s)|^2) + E_{ext}\big(\mathrm{X}(s)\big)ds, \tag{39}$$

where a and b are weighting parameters that control the tension and rigidity of the snake. $\mathrm{X}'(s)$ and $\mathrm{X}''(s)$ denote the first and second derivate of $\mathrm{X}(s)$ with respect to s. The external energy function E_{ext} is derived from the image so that it takes its smaller values at the features of interest, for example boundaries.

A snake that minimizes (39) must satisfy the following Eulerian equation:

$$a\mathrm{X}''(s) - b\mathrm{X}''''(s) - \nabla E_{ext} = 0, \tag{40}$$

where $\mathrm{a}(s) - b\mathrm{X}''''(s)$ can be considered as an internal force E_{int} which discourages stretching and blending while the external potential force $-\nabla E_{ext}$ pulls the snake towards the desired image contour.

To find a solution for (40), the snake is made dynamic by treating X as also a function of time t. Then, the partial derivate of X with respect to t can be set as the left hand side of (40) as follows:

$$\mathrm{X}_t(s,t) = \alpha\mathrm{X}''(s,t) - \beta\mathrm{X}''''(s,t) - \nabla E_{ext}. \tag{41}$$

A solution to (41) can be found by discretizing the equation and solving the discrete system iteratively.

Xu and Prince defined a new non-irrotational external force field, which is called the gradient vector flow (GVF) field in [81]. First, they define a gradient field ∇f based on the edge map of the image having the property that it is larger near the image edges. ∇f consists of vectors pointing toward the edges, but it has a narrow capture range, in general. Furthermore, in homogeneous regions, the field is zero, and therefore no information about nearby or distant edges are available. Finally, GVF is defined to be the vector field $v(x,y) = (u(x,y), v(x,y))$ that minimizes the following energy function:

$$\varepsilon = \int\int \varphi({u_x}^2 + {u_y}^2 + {v_x}^2 + {v_y}^2) + |\nabla f|^2|v - \nabla f|^2 dxdy, \tag{42}$$

where φ is a regularization parameter. This parameter should be set according to the amount of noise on the image. After computing $v(x,y)$, the external potential force $-\nabla E_{ext}$ in the dynamic snake equation (41) can be replaced by $v(x,y)$. Using a force balance condition and the GVF potential force field in (41), a GVF snake can be defined. The GVF field points toward the object boundary in its close surrounding, and varies smoothly over homogeneous image regions.

4.2.3.2 Markovian Segmentation Model

Markov Random Field (MRF) provides a robust tool to find exact boundaries of objects by minimizing a specific energy function. To find the global minimum for the usual energy function is an *NP*-hard problem, however, certain energy functions can be minimized in polynomial time by graph cuts. Lesko *et al*. [82] proposed a segmentation algorithm, which requires user interaction to mark an initial set of foreground/background pixels. Then, segmentation is performed via graph cut in real time.

As a well-known approach, segmentation can be considered as a labeling problem, where labels $\ell_p \in \Lambda = \{0,1\},$ are assigned to the pixels $p \in S = \{p_1, p_2, \dots, p_n\} \subset \mathbb{Z}^2$ based on some of their observed features $H(p)$. Based on the Bayesian theorem, the posterior probability can be factorized as $P(\ell|H) \propto P(H|\ell)P(\ell)$, where the optimal segmentation $\hat{\ell}$ is obtained as the Maximum a Posteriori (MAP) estimate. Based on the Hammersley-Clifford theorem [83], $\hat{\ell}$ can be found with specifying MRF with clique potentials and minimizing Gibbs energy.

The main contribution of [82] is to construct the Gibbs energy function in a way that it can be minimized via standard max-flow/min-cut. Namely, the full gradient information is exploited as magnitude and direction next to the gray-level intensity and only the pairwise interactions (doubleton cliques) are considered. In this way, the constructed Gibbs energy can be represented by a graph and an exact MAP solution can be determined by computing the minimum *s-t-cut* on the graph [84] in polynomial time.

4.2.3.3 Level-Set Method and Chan-Vese energy function

A level-set framework considers the 2D boundary B as an embedded part of a 3D surface and B is represented as the zero-level-set, where the 3D surface is intersected by the image plane. This approach allows the contour to vary iteratively from pixel to pixel by modifying the 3D surface and the image is divided into separate regions so that the defined energy function is minimized.

Whitaker proposed the sparse field method (SFM) [85] which represents the 3D surface by lists of points L_0, L_{-1}, L_{+1}, L_{-2}, L_{+2},... according to the distance of the points from the intersection image plane, where L_0 contains the pixels of the zero-level-set, and L_{-1}/L_{+1} contains the inner/outer adjacent pixels, respectively. The changes of the 3D surface are followed by moving the pixels from/to the appropriate lists. The level-set method can determine the zero-level-set for the next iteration by minimizing the following Chan-Vese energy function:

$$E_{Chan-Vese} = \varepsilon_1(B) + \varepsilon_2(B). \tag{43}$$

The energy function (43) depends only on the difference of the pixel intensities $I(p)$ and the respective average intensities inside $\hat{R}_{in}$ and outside $\hat{R}_{out}$ the contour B as formulated in the following way:

$$\varepsilon_1(B) = \sum_{p \in inside(B)} \left| p - \hat{R}_{in} \right|^2, \tag{44}$$

$$\varepsilon_2(B) = \sum_{p \in outside(B)} \left| p - \hat{R}_{out} \right|^2, \tag{45}$$

The energy function (43) takes its minimum, when both the inside and outside regions are the most homogeneous regarding pixel intensities.

4.3 Methodology of the exudate detection framework

In this section, we introduce an automatic exudate detector framework, which combines the advantage of several image pre-processing methods and applies a novel exudate detection approach using an active contour method. The candidate extractor algorithm [20] is based on grayscale morphology which has high sensitivity, since it basically marks every bright region as an exudate. To preserve up its high sensitivity, next to increasing the specificity, we try to reduce the number of false positive regions. For this purpose, we exclude the non-exudate regions by classification and to enhance the accuracy of this classification step, we apply the most descriptive features of exudates.

The method starts with rescaling the images to normalize the resolution to a common height of 1500 pixels. Next, a rough candidate extractor is applied to retrieve the possible exudate regions. Because of the high similarity in appearance between the exudates and the OD, we exclude OD region from the candidate ones. For the localization of the OD, we apply the ensemble-based method introduced in section 3. After OD removal, we use several proposed image enhancement methods to improve the contrast of the image and to maximize the amount of retrievable information content. Finally, some features are extracted from each candidate, and a properly adjusted Naïve Bayes classifier labels each candidate as an exudate or non-exudate.

4.3.1 Pre-processing step

First, we use a morphological-based method for candidate extraction because of it can extracts candidate regions quite reliably. As we introduced in section 4.2.1, the candidate extractor applies morphological operations such as morphological closing in order to eliminate blood vessels. Standard deviation is calculated, and thresholding is applied to find the candidate regions. Next, morphological reconstruction is applied with a second thresholding to find the contours of the exudates. The partial results of the candidate extraction can be seen in Figure 25.

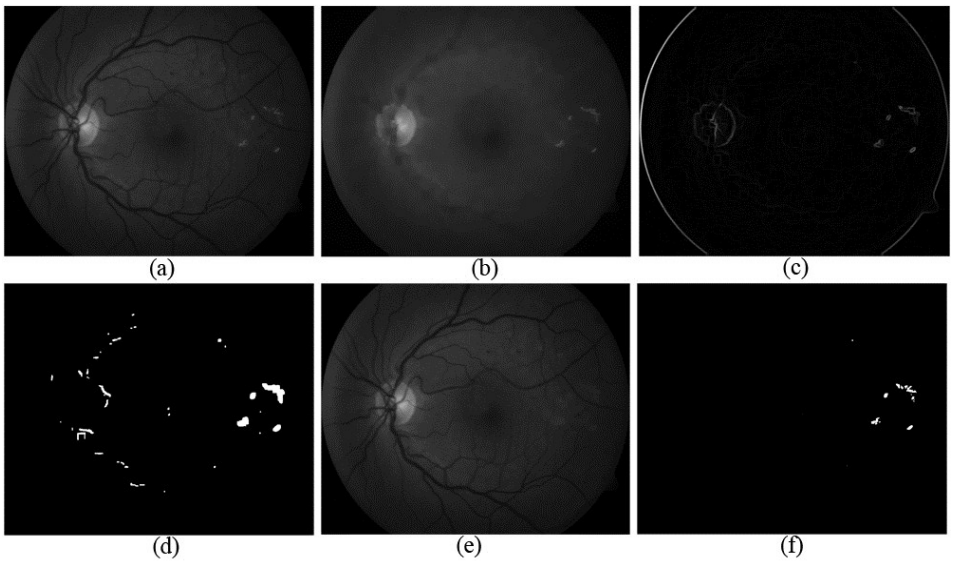

Figure 25: Partial results of the candidate extraction: (a) I_G, (b) vessel free image after morphological closing, (c) the local variation at each pixel, (d) extracting bright regions by adjusted thresholding, (e) result of morphological reconstruction, (f) $\mathcal{C}_I$.

However, this method works improperly on the retinal images of young patients, where shiny regions spread along the temporal arcade

(main vessels). Moreover, the boundaries of the detected exudates are less natural due to the applied structural elements and the method detects several false positives, as well. For these reasons, we use the result binary image $\mathcal{C}_I$ (as given in definition 14) only as an initial mask for a more precise detection step. We also note that, based on the result of our proposed ensemble OD detector, the pixels in OD region are not considered as possible exudate candidates through the detection procedure.

In the corresponding literature [11], the automatic exudate detection algorithms are improved by some image enhancement methods which can make the image more suitable for automatic evaluation. Firstly, we simply extract the I_G from the RGB fundus image and the channel I_I from the color space HSI. Besides using the intensity images I_G and I_I, we apply another seven image enhancement methods (introduced section 4.2.2) which are proposed in the literature of retinal image analyses focusing mainly on exudate detection [20], [24], [26], [27], [30]. The corresponding enhancement algorithms highlight several typical features of exudates. The reason for selecting these specific seven methods [11], [25], [56], [76], [78], [79], [80] also lies in the fact that they provide larger image contrast compared to e.g. [88], [89], [90], [91]. The result of the seven enhancement method with I_G and I_I which compose together $\mathcal{J}$ (as given in definition 20) can be seen in Figure 26.

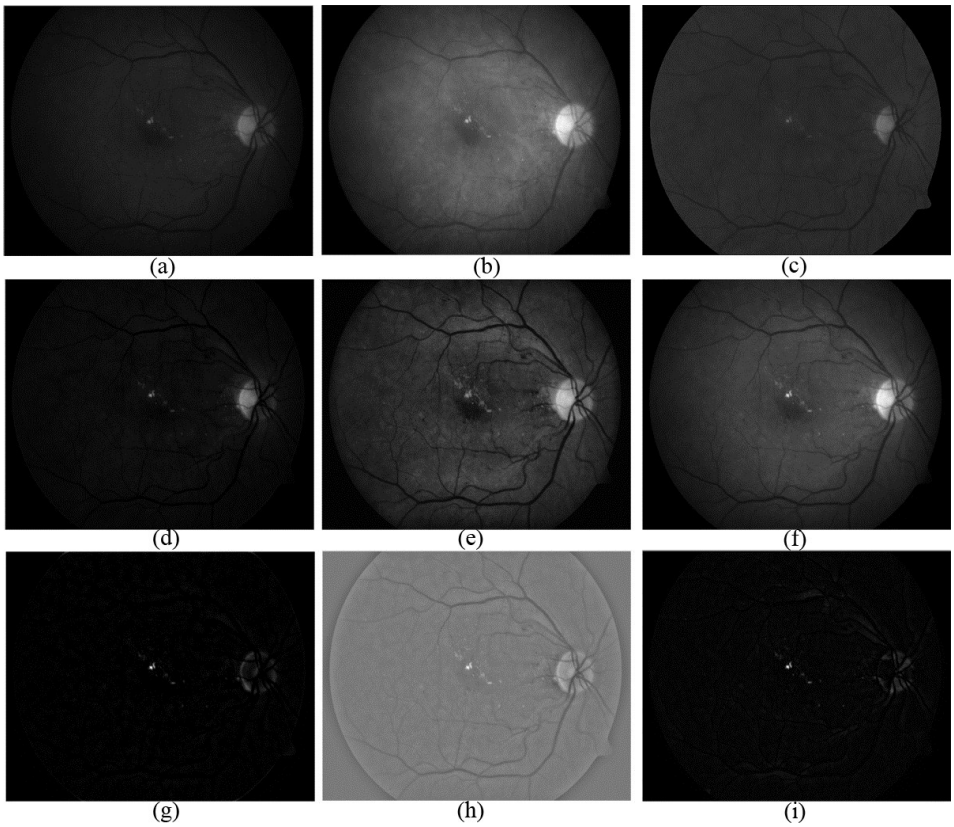

Figure 26: Nine different variants of the input image. (a) I_G, (b) I_H, (c) I_{CN}, (d) I_{CE}, (e) I_{CL}, (f) I_{GN}, (g) I_{IC}, (h) I_{IE}, (i) I_{WT}.

4.3.2 Precise Boundary Detection for each Candidate

Our aim is to determine the most precise boundary for each R_i (i=1,2,…,$|\Re_I|$) to improve the accuracy of the region-wise classification. For this aim, we examine the adaptability of three different ACM [81], [82], [85] from the relevant literature. First, we apply a gradient vector-

flow field based active contour model (known as snake) [81], then we determine the boundary of the exudates by a Markov Random Field and a graph cut algorithm [82]. Finally, the Chan-Vese energy function is minimized by an optimized level-set method [85].

The weaknesses of the GVF snake (introduced in section 4.2.3.1) lie in the hard selection of the parameters and its sensitivity for its initialization. Moreover, it is not an appropriate tool for exudate boundary detection, since the energy function prefers high image gradients and smooth curvature of the contour. However, high contrast may be present close to the vessels in fundus images and the curvature of the contour of the exudates varies irregularly. We applied the GVF snake for this task and its drawbacks can be checked on the result which can be seen in Figure 27.

For the precise boundary detection of the candidates, we also apply the optimized MRF model (introduced in section 4.2.3.2) with a modification to make the method automatic. In order to eliminate user interaction, we define the initial foreground/background pixels as the result of candidate extraction. Based on our experiments, the MRF model initialized automatically provides near precise boundaries of candidates. The result of the MRF model can be seen in Figure 27.

Finally, we apply a level-set method which minimizes the Chan-Vese energy function. This function is suitable for exudate segmentation because it depends only on the difference of the pixel intensities, as it can be seen in Figure 27. However, there are two main drawbacks of the level-set methods introduced in section 4.2.3.3. Namely, handling the 3D surface makes them rather slow, and the definition of the initial contour is also difficult. On the one hand, to reduce the computational time, Whitaker proposed the sparse field method (SFM) [32]. On the other hand, to initialize the zero-level-set, we use the boundary points of the extracted candidates found by the candidate extractor.

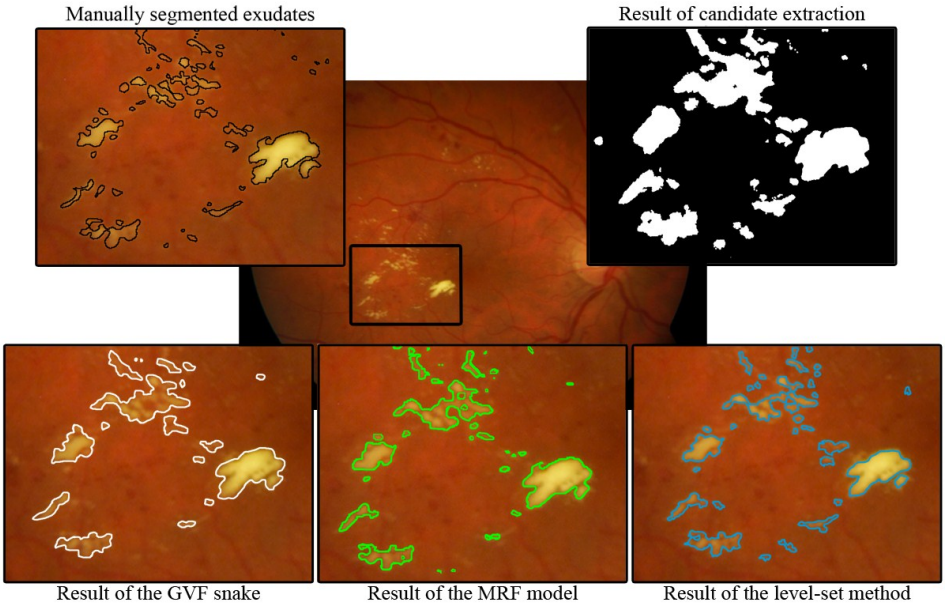

Figure 27: Results of the three different ACMs initialized by $R_i \in \mathfrak{R}_I$ (i=1,2,…,$|\mathfrak{R}_I|$).

To segment the contours of exudates more precisely, we propose the following ensemble of the image enhancement methods. The corresponding enhancement algorithms (introduced in section 4.2.2) highlight several typical features of exudates and to detect the precise exudate boundaries, we propose the application of level-set methods with SFM which minimizes the energy function regarding the each variants of the input image. Finally, we combine the nine extracted contours. The motivation behind this objective is that the features extracted from the precisely segmented regions are more appropriate to differentiate the true exudates from the false ones.

In practice, ACM minimizes their energy function regarding on each $I' \in \mathcal{J}$. ACM is applied separately on the nine disparate enhanced images

to produce nine different contours $B_{i,j}$ for each $R_i \in \mathfrak{R}_I$ (i=1,2,…,$|\mathfrak{R}_I|$) (j=1,2,…,$|\mathcal{J}|$) as shown in Figure 28.

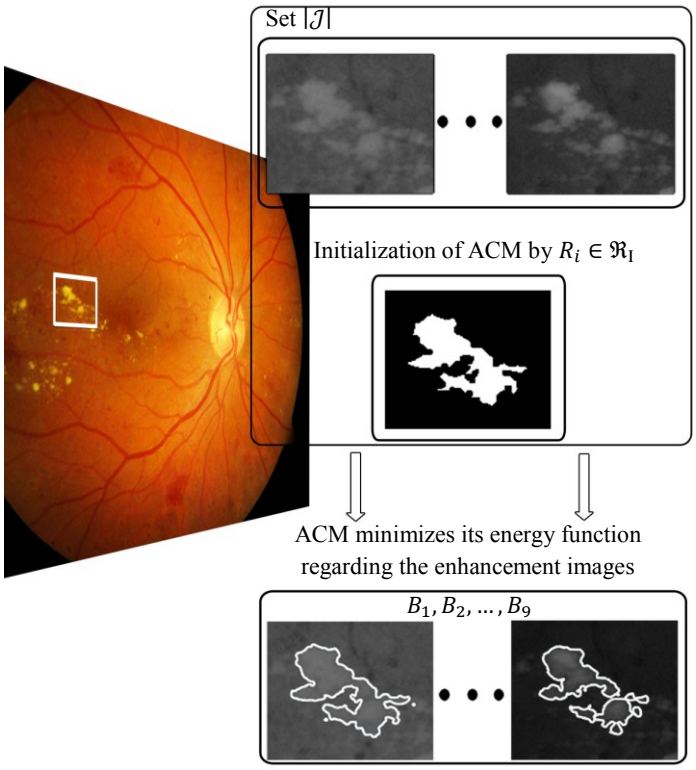

Figure 28: Different contours $(B_1, B_2, \dots, B_9)$ for $R_i \in \mathfrak{R}_I$ (i=1,2,…,$|\mathfrak{R}_I|$).

The next step is to combine these boundaries $B_{i,j}$ (j=1,2,…,$|\mathcal{J}|$) for the determination of the precise boundary (B_i^*) for each $R_i \in \mathfrak{R}_I$ (i=1,2,…,$|\mathfrak{R}_I|$) based on. Precise boundary detection is found to be essential to provide appropriate features for a region-wise classification of true/false candidates. To let the different preprocessors take an effect on the final contour of a specific candidate, we fuse the corresponding extracted information in terms of combining the regions $B_{i,j}$

(i=1,2,…,$|\Re_I|$) (j=1,2,…,$|\mathcal{J}|$) in the following way. First, we create the union of the regions $R_{i,j}$ bounded by $B_{i,j}$ as $\mathcal{R} = \bigcup_{j=1}^{|\mathcal{J}|} R_{i,j}$ and to each pixel $p \epsilon \mathcal{R}$ we assign a score as:

$$Score(p) = \left|\{j : p \epsilon R_{i,j} \ (j = 1,2, \dots, |\mathcal{J}|)\}\right|, \tag{46}$$

where $Score(p)$ is means the number of $R_{i,j}$ (j=1,2,…,$|\mathcal{J}|$) containing p. If p falls out of all the regions $R_{i,j}$ (j=1,2,…,$|\mathcal{J}|$) it gets score 0, whereas it falls inside all of them, it gets score $|\mathcal{J}|$. This assignment leads to nine new regions $R'_{i,j}$ (j=1,2,…,$|\mathcal{J}|$) as:

$$R'_{i,j} = \{p \in \mathcal{R} : Score(p) \geq j\}, (j = 1,2, \dots, |\mathcal{J}|) \tag{47}$$

with $R'_{i,j}$ consisting of pixels having score greater than or equal to j. Note that, we have $R'_{i,1} \supseteq R'_{i,2} \supseteq \dots \supseteq R'_{i,|\mathcal{J}|}$, with $R'_{i,1} = \mathcal{R}$. In this way, we merge the $R_{i,j}$ (j=1,2,…,$|\mathcal{J}|$) and determine new regions $R'_{i,j}$ (j=1,2,…,$|\mathcal{J}|$) with their respective boundaries $B'_{i,j}$ (j=1,2,…,$|\mathcal{J}|$). Besides fusing the regions $R_{i,j}$ (j=1,2,…,$|\mathcal{J}|$), we have found that the final region should represent a stable state as well. That is, we select that $R'_{i,j}$ (j=1,2,…,$|\mathcal{J}|$) as the final combined region (R_i^*), which is the most similar to its neighbors $R'_{i,j-1}$ and $R'_{i,j+1}$ regarding the fusion. As extreme, less meaningful cases, $R'_{i,1} = \bigcup_{j=1}^{|\mathcal{J}|} R'_{i,j}$ and $R'_{i,9} = \bigcap_{J=1}^{|\mathcal{J}|} R'_{i,j}$ are excluded from this analysis. For a precise formulation of this process, we measure the similarities between two adjacent regions by computing their symmetric difference as:

$$d\left(R'_{i,j}, R'_{i,j+1}\right) = \frac{\left|\left(R'_{i,j} \cup R'_{i,j+1}\right) \setminus \left(R'_{i,j} \cap R'_{i,j+1}\right)\right|}{\left|R'_{i,j}\right|} \tag{48}$$

(j=1,2,…,$|\mathcal{J}|$-1), where \ denotes the set difference operator. The denominator $|R'_{i,j}|$ in (48) is applied for scale-invariance. According to our experimental results, that $R'_{i,k}$ with $k \in \{2, \dots, |\mathcal{J}| - 1\}$ should be selected as R_i^* for which:

$$d(R'_{i,k-1}, R'_{i,k}) + d(R'_{i,k}, R'_{i,k+1}) = \min_{j=2,\dots,8} \left(d(R'_{i,j-1}, R'_{i,j}) + d(R'_{i,j}, R'_{i,j+1}) \right) \quad (49)$$

This procedure is performed for each for each $R_i \in \mathfrak{R}_I$ (i=1,2,…,$|\mathfrak{R}_I|$) separately and R_i is replaced with the selected R_i^* to have the set $\mathfrak{R}_I^*$ of the candidates with precisely detected contours. Naturally, some candidates are not true exudates, but we determine the best fitting boundary for each candidate individually to improve the accuracy of the region-wise classification. An example for this procedure is also shown in Figure 29, where the measured similarities are also included.

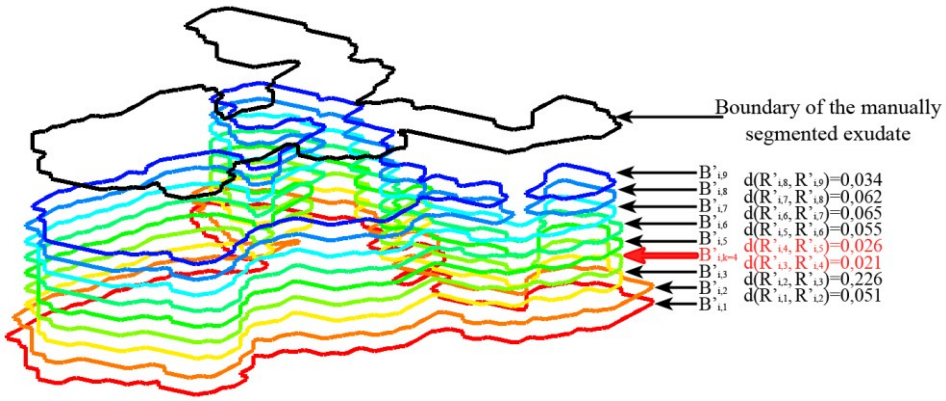

Figure 29: The boundaries $B'_{i,j}$ of the combined regions $R'_{i,j}$ and the boundary of a manually segmented exudate. Red arrow shows the selected region $R_i^* \in \mathfrak{R}_I^*$.

4.3.3 Feature extraction

A candidate extractor has a high sensitivity, since it finds almost all exudates in the input image. However, it marks almost each bright region (mainly the ones which are close to vessels on retinal images of youngsters) as exudates which lead to many false positive hits. If we consider all these candidates as the result of detection of exudates, the specificity of the automatic screening system drops. To exclude the false positive candidates, besides keeping up high sensitivity, we propose a region-wise classification step which labels each candidate region as exudate or non-exudate. This step can be considered as a post-processing step, where each candidate is classified by an optimally adjusted Naïve Bayes classifier based on extracted features.

4.3.3.1 Region-wise descriptors

For classification, we extract descriptors from each $R_i^* \in \Re_I^*$ (i=1,2,…,$|\Re_I^*|$) as it is given in definition 22. $H(R_i^*)$ descriptors are based on the respective intensity values of pixels composing the properly detected candidate and are calculated from the morphological behavior (shape) of the precisely detected region and its boundary. In this way, for candidate classification initially we consider 106 region-wise descriptors listed in Table 5.

	region-wise features	boundary-wise features
*Based on the intensity images I_G, I_I, I_{CN}, I_{GN}, I_{CL}, I_{CE}, I_{IC}, I_{IE}, I_{WT}. (9×10 features)	– mean – standard deviation – minimum value – maximum value – range (difference of minimum and maximum value)	– mean – standard deviation – minimum value – maximum value – range (difference of minimum and maximum value)
Based on the magnitude of gradient image of the green channel. (10 features)	– mean – standard deviation – minimum value – maximum value – range (difference of minimum and maximum value)	– mean – standard deviation – minimum value – maximum value – range (difference of minimum and maximum value)
Morphological (shape) descriptors. (6 features)	– compactness – area – number of holes – elongatedness – eccentricity	– perimeter

*The intensity based descriptors extracted separately for the images I_G, I_I, I_{CN}, I_{CE}, I_{CL}, I_{GN}, I_{IC}, I_{IE}, I_{WT}.

Table 5: The extracted $H(R_i^*)$ descriptors.

In Table 5, we enclosed all the descriptors extracted from the candidate regions of a training dataset TR_2 regardless of their classification performance, so it contains such descriptors that are less efficient for this task. To select the meaningful features for classification, we evaluate their corresponding performance by two-sample *t*-tests.

We note that, we tested some commonly used classifiers (Naïve Bayes, *k*-Nearest Neighbors) with several feature selection methods like PCA, relative entropy, minimum attainable classification error, ROC analysis, Wilcoxon test, etc. based on class separability criteria. We found that the Naïve Bayes classifier and the two-sample *t*-test are the most

efficient in our case. Namely, the Naïve Bayes classifier reached the highest *ACC* value (84.37%) when it used features selected by two-sample *t*-test for labeling. For the sake of completeness, we list the highest *ACC* values of Naïve Bayes, when it used features selected by other methods. The following feature selection methods are included (with highest *ACC* is also indicated): PCA (80.38%), minimum attainable classification error (Bhattacharyya) (80.44%), relative entropy (80.51%), ROC analysis (81.09%) and Wilcoxon test (81.61%). Now we give the proper description of feature selection based on the two-sample *t*-test.

To rank the descriptors based on their performance obtained by the two-sample *t*-test, we used 28 training images from the TR_2. $I \in \mathrm{TR}_2$ were considered as input to the candidate extractor method. Then, we generated $\mathcal{J}$, finally determined the R_i^* by the proposed boundary detection algorithm. The $\sum_{I \in \mathrm{TR}_2} |\Re_I^*| = 1239$ and we labeled manually the 1239 candidates as true exudates (955 true positive) or not (284 false positive) according to the manually segmented binary images described in section 2.1.7. Based on this labeling, the *t*-test can be performed for each descriptor given in Table 5 according to the following formula:

$$dt_j = \frac{(\mu_{j_F} - \mu_{j_T})\sqrt{\frac{nm(n+m-2)}{n+m}}}{\sqrt{(n-1){\sigma_{j_F}}^2 + (m-1){\sigma_{j_T}}^2}}, \tag{50}$$

where t_j denotes the performance of the *j*-th descriptor, n is the number of the true exudate regions (955) and m is the number of the false ones (284) considering the whole set of the candidates for the training set. μ_{j_T} (resp. μ_{j_F}) denotes the mean, while σ_{j_T} (resp. σ_{j_F}) denotes the standard deviation of the *j*-th descriptor of all true (resp. false) exudates respectively.

To find the meaningful descriptors, we ranked them according to their performance values t_j, and tried to find those k number descriptors, which provided highest ACC. For this aim, we divided the set of 1239 manually labeled regions into a training and a test parts. Then, a Naïve Bayes classifier was trained on the first k (k=1,…,106) region-wise features extracted from the training regions. Next, we observed the performance of the classifier on the test regions using these k features. We found that the Naïve Bayes classifier reached the highest ACC value, when the first k=29 descriptors (see them in Table 6) were selected from the ranked list as features (see Figure 30 (a)). For this evaluation, we separated the regions into a training and test set by K-fold cross validation (K=10) and evaluated the performance of the classifier for a given feature set at 10 times. For the sake of completeness, the accuracies of different combination of classifiers (Naïve Bayes, k-Nearest Neighbors) and feature selection methods regarding the numbers of features are enclosed in Fig. 30 (a)(b). These empirical results also serve as further proof for our former claim that the applied Chan-Vese energy function minimized by level-set method is the most appropriate tool for the detection of the irregularly variable contours of the exudates. As it can be observed in Fig. 30 (c)(d)(e)(f), the accuracies of classifiers decrease if the exudate candidates are extracted by the MRF or the GVF snake.

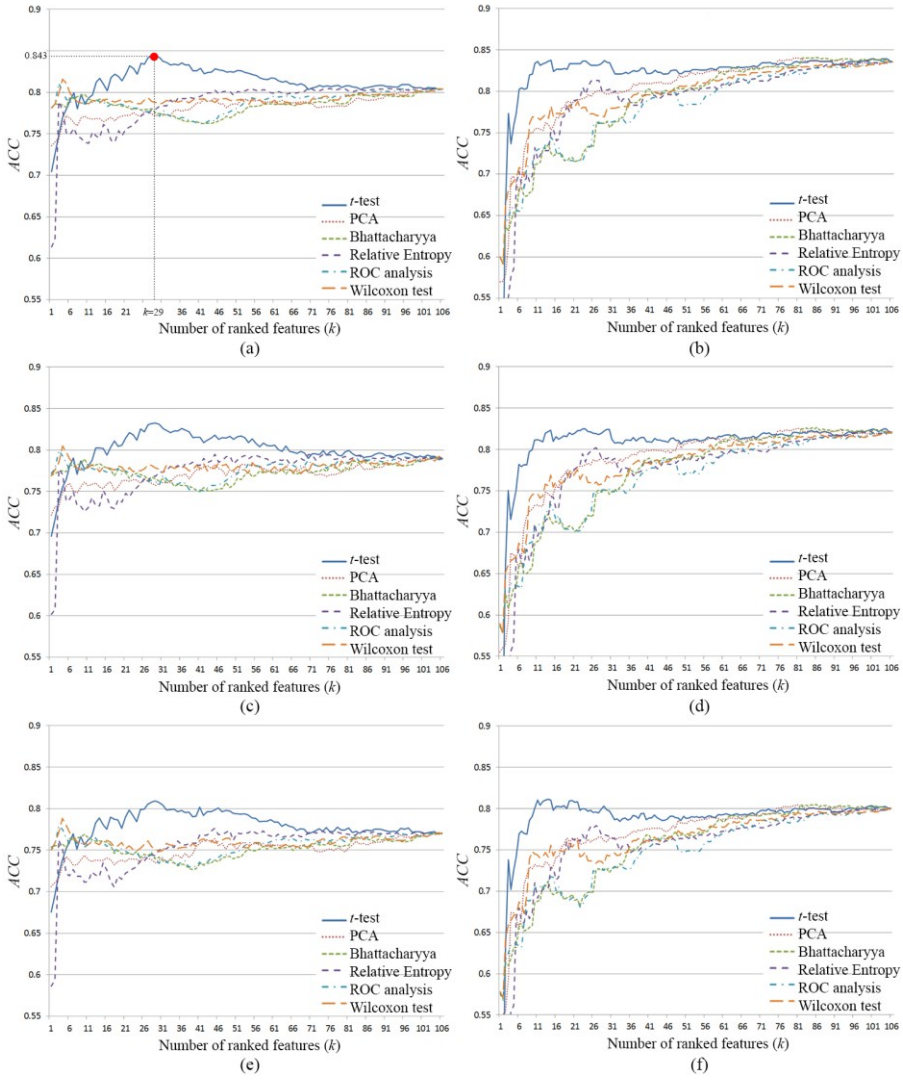

Figure 30: Peformance of the classifiers using different feature selection methods. Naïve Bayes classifier with (a) level-set method, (c) MRF and (e) GVF snake. *k*-Nearest Neighbors classifier with (b) level-set method, (d) MRF and (f) GVF snake.

	region-wise features	boundary-wise features
*Based on the intensity images I_G, I_{CL}, I_{CE}. (3×6 features)	– standard deviation – maximum value – range (difference of minimum and maximum value)	– standard deviation – maximum value – range (difference of minimum and maximum value)
Based on the magnitude of gradient image of the green channel. (5 features)	– mean – maximum value	– mean – minimum value – maximum value
Morphological (shape) descriptors. (6 features)	– compactness – area – number of holes – elongatedness – eccentricity	– perimeter

*The intensity based descriptors extracted separately for the images I_G, I_{CE}, I_{CL}.

Table 6: Selected region-wise descriptors to classify exudate candidates.

4.3.3.2 Pixel-wise descriptors

We investigated also the performance of the pixel-wise features and the pixel-wise classification. As given in definition 22, when H assigns feature vector for each $p \epsilon R_i \in \Re_I$ *(i=1,2,…,*$|\Re_I|$*)* and p, $q \in R_i$ *(i=*1,2,…,$|\Re_I|$*)* then $H(p) \neq H(q)$ is possible. In this way the features are derived from p and its local neighborhood ($p_{n\times n}$).

As we mentioned in section 4, there is a large number of exudate detectors based on pixel-wise classification. The majority of them consider e.g., the pixel intensity, local contrast, average intensity of neighbors of pixels as primary features of exudate pixels for classification. However, the average intensity of the images depends on the ethnicity and the age of the patients. It means that a high intensity value from a dark image is not labeled as an exudate pixel by a classifier

because it is not high enough according to the model which is based on training dataset. To avoid this misclassification, we propose to normalize the intensity values by the average intensity of the OD of the analyzed retina image.

To retrieve the average intensity of the OD, we cluster the pixels under the detected OD region using *Fuzzy C-Means Clustering* algorithm implemented in Matlab [86]. This clustering step is necessary because the detected OD region contains a lot of blood vessel pixels and the *Fuzzy C-Means Clustering* divides the pixels into bright and dark group according to their intensities. We consider the average intensity of the brighter cluster as the average intensity of OD and we normalize individually I_G, I_H, I_I and I_{CL} channels.

After this normalization step, we extract 56 descriptors which may be useful for a pixel-wise classification. For the enumeration of possible descriptors we need to mention the difference of the Gaussian (DoG) filters. Sopharak *et al.* [24] proposed six DoG filters for exudate detection. We apply these filters on I_{CL} and we extract the following descriptors from $DoG_1, \dots, DoG_6$ too. We extract the mean, standard deviation, maximum value, range (difference of maximum and minimum) of the intensities in $p_{n\times n}$ regarding I_G, I_H, I_I, I_{CL} and $DoG_1, \dots, DoG_6$ images to each candidate pixel. Moreover, we extract some descriptors that are based on the strength of the edge in $p_{n\times n}$. For this, we apply the Frei-Chen edge detector [87] on I_{CL} and we retrieve the highest gradient value, average and standard deviation of the strength of the edge pixels. Finally, we take into account the number of pixels of the candidate region which contains the pixel and we normalize this pixel number by pixel number of the detected OD.

These descriptors are retrieved for each $p \epsilon R_i$ *(i=1,2,...,*$|\Re_I|$*)*. Most of the descriptors are appropriate to be used for distinguishing between exudate and non-exudate pixels. However, there are some irrelevant

descriptors and using them can decrease the generalization performance of the classifier. To select the most significant descriptors we also evaluated the pixel-wise features by two-sample *t*-test.

4.3.4 Boosted Naïve Bayes classifier

The simple Naïve Bayes classifier reaches relatively high accuracy regarding the correct labeling of the exudate candidates. As for implementation, the built-in Matlab R2010b class of Naïve Bayes [92] is used. To increase further the performance of the classifier, we apply the adaptive boosting (AdaBoost) technique [93]. To realize the idea that the performance of ensemble learning is usually better than single learning, to set up an ensemble of classifiers, the set TR_2 is separated into two disjoint sets T_1 and T_2. The first classifier learns on T_1 and classifies the elements of T_2. In the next turn, the new classifier is trained on mainly the previously misclassified elements to teach it for the instances that are hard to classify, and so on. Finally, these classifiers make a decision about a label of a new instance by weighted majority voting, where the weights come from the individual accuracies of the classifier. In this way, an ensemble of several Naïve Bayes classifiers can achieve 10% higher accuracy in labeling the candidates as exudates or non-exudates in our framework.

4.4 Experimental results

The test part of DIARETDB1 includes 29 (from 61) and that of HEI-MED 54 (from 169) images which contain exudates according to the publicly available annotation of clinical experts. Based on this knowledge, we measured the performance of the proposed method.

For the sake of completeness, we evaluate the performance of the proposed framework when it used pixel-wise classification based on pixel-wise features (introduced in section 4.3.3.2) and region-wise classification based on region-wise features (introduced in section 4.3.3.1). In the case of region-wise descriptors, we considered following cases; when the boundary of the candidates are determined by gradient vector flow field and snake model (introduced in section 4.2.3.1), Markovian segmentation model (introduced in section 4.2.3.2) or level-set method (introduced in section 4.2.3.3). Finally, we used the proposed fusion of image enhancement method with the most precise ACM (level-set method and Chan-Vese function) for the precise boundary detection and evaluate the classification of candidates based on features set extracted from candidates with precise segmented contours.

Based on the measures (given in definition 26), we can compare the proposed exudate detector framework with some other state-of-the-art exudate detector algorithms quantitatively. As we can see in Table 7 and 8, in these tests the proposed algorithms can outperform the algorithms [20], [21], [22], [23], [24], [25], [28] involved in our comparative study with respect to the *ACC* value. Note that, *ACC* can also be derived from *SEN* and *SPE* (as given in definition 26), so *ACC* is high if and only if both *SEN* and *SPE* are high. Moreover, it can be observed that the application of all the nine pre-processing methods simultaneously and the detection of the precise boundary of the candidates lead to a meaningful improvement in region-wise classification, as well.

	SEN	*SPE*	*ACC*
Region-wise classification after Precise Boundary Detection	0.92	0.68	**0.82**
Sopharak *et al.* [21]	0.73	0.87	**0.79**
Sánchez *et al.* [25]	0.66	0.92	**0.77**
Region-wise classification after level-set method used	0.90	0.54	**0.75**
Region-wise classification after MRF used	0.90	0.54	**0.75**
Walter *et al.* [20]	1.00	0.32	**0.72**
Region-wise classification after GVF snake used	0.81	0.54	**0.70**
Exudate detection by pixel-wise classification	0.69	0.68	**0.69**
Welfer *et al.* [22]	0.79	0.55	**0.69**
Sopharak *et al.* [23]	1.00	0.14	**0.64**
Jaafar *et al.* [28]	1.00	0.06	**0.61**
Sopharak *et al.* [24]	1.00	0.02	**0.59**

Table 7: Comparative results for image-level classification for the proposed method on the dataset DIARETDB1.

	SEN	*SPE*	*ACC*
Region-wise classification after Precise Boundary Detection	0.87	0.86	**0.86**
Sopharak *et al.* [21]	0.65	0.90	**0.82**
Sánchez *et al.* [25]	0.62	0.90	**0.81**
Region-wise classification after level-set method used	0.83	0.79	**0.80**
Region-wise classification after MRF used	0.81	0.78	**0.79**
Welfer *et al.* [22]	0.70	0.84	**0.79**
Region-wise classification after GVF snake used	0.74	0.77	**0.76**
Exudate detection by pixel-wise classification	0.61	0.82	**0.75**
Sopharak *et al.* [23]	0.91	0.68	**0.75**
Walter *et al.* [20]	0.93	0.65	**0.74**
Jaafar *et al.* [28]	0.88	0.65	**0.72**
Sopharak *et al.* [24]	0.93	0.60	**0.70**

Table 8: Comparative results for image-level classification for the proposed method on the dataset HEI-MED.

For the sake of completeness, we also enclose the receiver operating characteristics (ROC) results of the proposed method in Figure 31 to demonstrate its robustness at image level on the datasets DIARETDB1 and HEI-MED. To create the ROC curve, we applied different threshold levels for the weighted majority voting result of the boosted Naïve Bayes classifiers to accept exudate candidates as true ones. As in [21], [22], [23], [24] and [28] the authors did not define an adjustable parameter, a complete ROC analysis cannot performed regarding them. Instead, we indicate only their single available (*SEN*, 1-*SPE*) figures in Figure 31. However, a full comparative ROC analysis with [20], [25], pixel-wise classification and the proposed method with/without precise boundary

detection is included in the figure. The competitiveness of the proposed method can be observed regarding ROC analysis, as well.

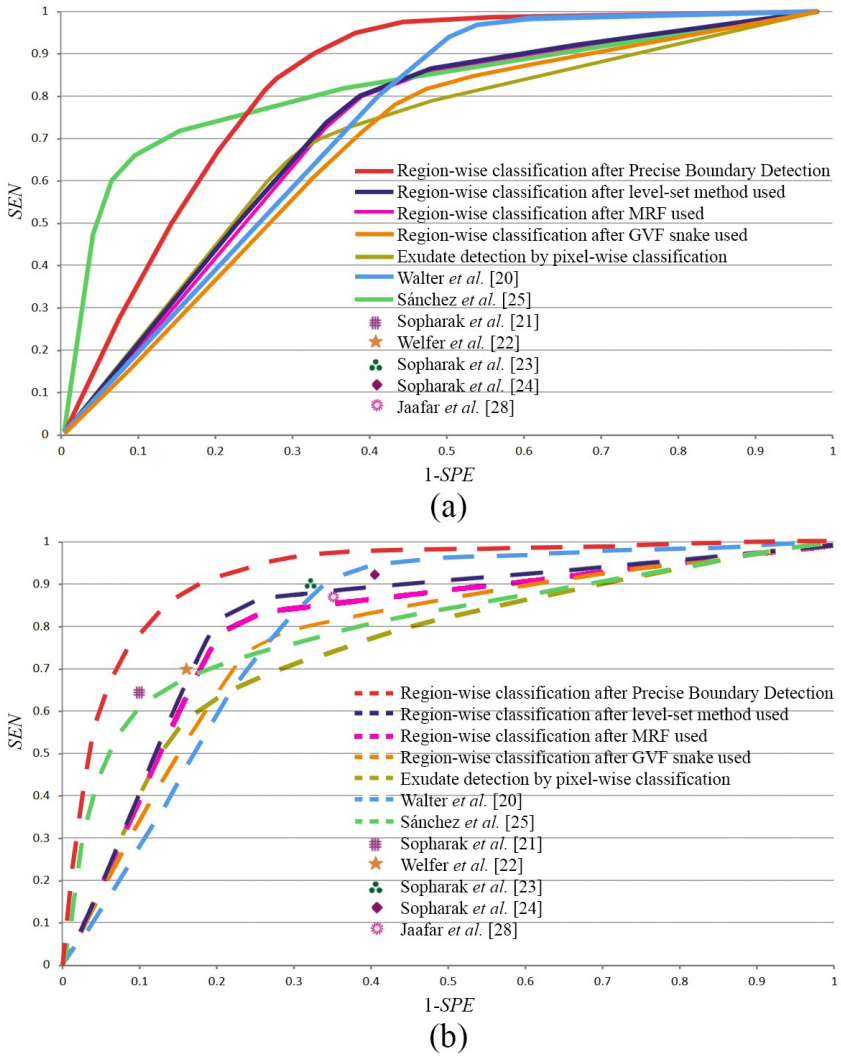

Figure 31: Comparative analysis for receiver operating characteristics (ROC) on the datasets (a) DIARETDB1, (b) HEI-MED.

For the sake of completeness, Figure 32 demonstrates the differences between the results of the proposed algorithm and the state-of-the-art methods involved in our comparative study. Figure 32 includes the final outputs of [20], [21], [22], [23], [24], [25], [28] and the proposed method for the same sample image.

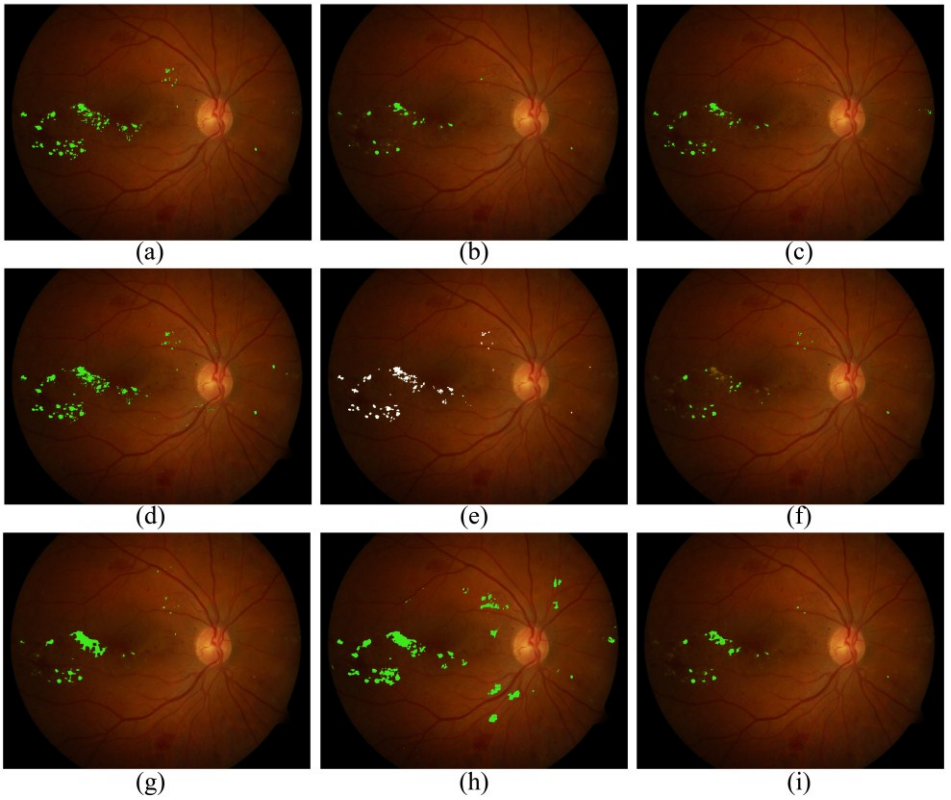

Figure 32: Visual comparison of state-of-the-art methods with the proposed one. Result of (a)Precise Boundary Detection, (b) [21], (c) [25], (d) [20], (e) manual segmentation, (f) [22], (g) [23], (h) [24], (i) [28].

4.5 Discussion

The proposed combined framework is dedicated to the detection of bright lesions caused by diabetic retinopathy, especially for exudates. That is, the selected pre-processing methods enhance the contrast between the bright lesions and their dark background. The candidate extractor method finds the regions which might contain these bright lesions, while the consequent region-wise classifier is trained to select the bright patches with irregular contours as exudates. Naturally, if we change the components of the approach appropriately, it could be applied also for the segmentation of expanded dark lesions like hemorrhages.

According to Abramoff *et al.* [94], such screening algorithms cannot be recommended for clinical practice. However, our proposed methodology with its high *ACC* at image-level can be a solid component of a complex system to make the decision about the further clinical investigation of the patient. Precise boundary detection has an implicitly important influence on clinical practice to increase image-level accuracy.

The presented method was implemented in Matlab, and we used a single core 2.4GHz CPU with 2GB memory for testing. The applied candidate extractor algorithm provides a large number of candidates and this is the reasons for the currently relatively high computational time. Using the proposed level-set methods takes 31 seconds per image (17 seconds per image by using the MRF and 51 seconds per image by using the GVF snake), which could be drastically decrease e.g., by parallel implementation. For the sake of completeness, we enclose the computational times of the re-implemented algorithms, which are involved in our comparative studies: Sopharak *et al.* [21] - 6.63 s, Sánchez *et al.* [25] - 27 s, Welfer *et al*. [22] - 9.75 s, Sopharak *et al*. [23] - 86 s, Walter *et al*. [20] - 6.42 s, Sopharak et al. [24] – 17 s, Jaafar et al. [28] – 12 s per image. These times do not include the OD detection step.

5 Conclusion

In this PhD thesis, we have proposed some approaches to improve the accuracy of single object detection by using an ensemble of individual algorithms. The first approach is based on simple majority voting, where each member algorithm has a single candidate. Even if some of the algorithms cannot detect the exact location of the object, a majority voting-based combination will be capable of detecting the center point of the object in most cases appropriately. In the second approach, we have modified the algorithms so that they produce more than one candidate. These candidates are ranked according to confidence values assigned by the algorithms. Then, a weighted graph is defined from these candidates by taking the geometric properties of the target object also into consideration. The location of the center of the object can be detected by finding the maximal weighted clique. Finally, the combination of the probability maps by the Bayesian models makes also possible the approximate detection of the single object region.

In the second part of the thesis, we have presented an exudate segmentation approach which is based on the combination of grayscale morphology, active contour method and region-wise classification. The result of a grayscale morphology-based exudate detection method is considered as the initial mask for ACM. We investigated the applicability of three different ACMs and different image enhancement results are involved for minimizing the energy functions. The proposed exudate segmentation method finds the contours more precisely and reduces the number of false positive pixels and improves the reliability of the region-wise features. The candidates are labeled as exudate or non-exudate through a region-wise classification step. For this task, we extract carefully selected descriptors for each candidate. For feature selection, we

used a two-sample *t*-test, while for classification the Naïve Bayes classifier is optimized by AdaBoost.

Acknowledgement

Some elements of this thesis was supported by the Janos Bolyai Grant of the Hungarian Academy of Sciences, and by the TECH08-2 project DRSCREEN- Developing a computer-based image processing system for diabetic retinopathy screening of the National Office for Research and Technology of Hungary (contract no.: OM-00194/2008, OM-00195/2008, OM-00196/2008). This work was also supported by the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 project and the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP-4.2.4.A/ 2-11/1-2012-0001 'National Excellence Program'. Moreover, the MESSIDOR dataset is kindly provided by the Messidor program partners (see http://messidor.crihan.fr).

I am grateful to my thesis advisor András Hajdu for his encouragement and guidance. I am also thankful to the Doctoral School of Informatics of the University of Debrecen for having provided the opportunity to conduct researches and to write my thesis.

I am honored to have worked with the members of the DRSCREEN project, especially the co-authors of my papers. I am grateful to my colleague Attila Szalkay for proofreading this thesis.

Finally, I would like to thank my family and especially my wife Mariann for their support and patience throughout the years.

References

[1] International Diabetes Federation, “IDF Diabetes Atlas, 6th edn.” *International Diabetes Federation,* Brussel, Belgium, 2013.

[2] J. E. Shaw, R. A. Sicree, and P. Z. Zimmet, “Global estimates of the prevalence of diabetes for 2010 and 2030”, *Diabetes Research and Clinical Practice*, vol. 87, no. 1, pp. 4-14, 2010.

[3] R. J. Qureshi, L. Kovacs, B. Harangi, B. Nagy, T. Peto, A. Hajdu, “Combining algorithms for automatic detection of optic disc and macula in fundus images,” *Journal of Computer Vision and Image Understanding*, vol. 116, no. 1, pp. 138-145, 2012.

[4] F. Mendels, C. Heneghan, and J. P. Thiran, “Identification of the Optic Disk Boundary in Retinal Images using Active Contours”, *Proceedings of the Irish Machine Vision Image Processing Conference*, pp. 103–115, Sept. 1999.

[5] S. Sekhar, W. Al-Nuaimy, and A. Nandi, “Automated Localisation of Retinal Optic Disk using Hough Transform”, *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1577–1580, 2008.

[6] M. D. Abramoff, M. Niemeijer, “The Automatic Detection of the Optic Disc Location in Retinal Images using Optic Disc Location Regression”, *Proceedings of the IEEE Engineering in Medicine and Biology Society*, pp. 4432–4435, 2006.

[7] L. Shijian. "Accurate and efficient optic disc detection and segmentation by a circular transformation." *IEEE Transactions on Medical Imaging*, vol. 30, no. 12, pp. 2126-2133, 2011.

[8] M. Lalonde, M. Beaulieu, L. Gagnon, “Fast and Robust Optic Disk Detection using Pyramidal Decomposition and Hausdorff-Based

Template Matching", *IEEE Transaction on Medical Imaging*, vol. 20, pp. 1193-1200, 2001.

[9] A. Sopharak, K.T. New, Y.A. Moe, M.N. Dailey, B. Uyyanonvara, "Automatic Exudate Detection with a Naive Bayes Classifier", *Proceedings of the Conference on Embedded Systems and Intelligent Technology*, pp. 139–142, 2008.

[10] M. Niemeijer, M. D. Abràmoff , B. van Ginneken, "Fast detection of the optic disc and fovea in color fundus photographs", *Medical Image Analysis*, vol. 13, pp. 859–870, 2009.

[11] A. Hoover and M. Goldbaum, "Locating the Optic Nerve in a Retinal Image using the Fuzzy Convergence of the Blood Vessels", *IEEE Transaction on Medical Imaging*, vol. 22, pp 951-958, 2003.

[12] S. Ravishankar, A. Jain, A. Mittal, "Automated Feature Extraction for Early Detection of Diabetic Retinopathy in Fundus Images", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 210-217, 2009.

[13] X. Zhu, R. M. Rangayyan, "Detection of the optic disc in images of the retina using the Hough transform". *Proceedings of the IEEE Engineering in Medicine and Biology Society*, pp. 3546-3549, 2008.

[14] K. A. Vermeer, F. M. Vos, H. G. Lemij, A. M. Vossepoel, "A model based method for retinal blood vessel detection". *Computers in Biology and Medicine*, vol. 34, no. 3, pp. 209-219. 2004.

[15] P. Bankhead, C. N. Scholfield, J. G. McGeown, T. M. Curtis, "Fast retinal vessel detection and measurement using wavelets and edge location refinement". *PloS ONE*, vol. 7, no. 3, art. no. e32435, 2012.

[16] Gy. Kovács, A. Hajdu, “Template Matching Based Segmentation of Retinal Vessels Using Generalized Gabor Functions.” Under review by *Medical Image Analysis*; submitted in 2013.

[17] M. Niemeijer, B. van Ginneken, J.J. Staal, M.S.A. Suttorp-Schulten, M.D. Abramoff, “Automatic Detection of Red Lesions in Digital Color Fundus Photographs”, *IEEE Transactions on Medical Imaging*, vol. 24, pp. 584-592, 2005.

[18] B. Antal, A. Hajdu, “Improving microaneurysm detection in color fundus images by using context-aware approaches”, *Computerized Medical Imaging and Graphics*, vol. 37, no. 5, pp. 403-408, 2013.

[19] I. Lázár and A. Hajdu, “Retinal microaneurysm detection through local rotating cross-section profile analysis”, *IEEE Transaction on Medical Imaging*, vol. 32, no. 2, pp. 400-407, 2013.

[20] T. Walter, J. C. Klein, P. Massin and A. Erginay, “A contribution of image processing to the diagnosis of diabetic retinopathy – detection of exudates in color fundus images of the human retina”, *IEEE Transactions on Medical Imaging,* vol. 21, no. 10, pp. 1236–43, 2002.

[21] A. Sopharak, B. Uyyanonvara S. Barman and T. H. Williamson, “Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods”, *Computerized Medical Imaging and Graphics*, vol. 32, no. 8, pp. 720–727, 2008.

[22] D. Welfer, J. Scharcanski and D. R. Marinho, “A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images”, *Computerized Medical Imaging and Graphics*, vol. 34, no. 3, pp. 228–235, 2010.

[23] A. Sopharak, B. Uyyanonvara and S. Barman, "Automatic exudate detection from nondilated diabetic retinopathy retinal images using fuzzy c-means clustering", *Sensors*, vol. 9, no. 3, pp. 2148–2161, 2009.

[24] A. Sopharak, M. N. Dailey, B. Uyyanonvara, S. Barman, T. Williamson, K. T. Nwe and Y. A. Moe, "Machine learning approach to automatic exudate detection in retinal images from diabetic patients", *Journal of Modern Optics*, vol. 57, no. 2, pp. 124-135, 2010.

[25] I. Sánchez, R. Hornero, M. I. López, M. Aboy, J. Poza and D. Abásolo, "A novel automatic image processing algorithm for detection of hard exudates based on retinal image analysis", *Medical Engineering & Physics*, vol. 30, no. 3, pp. 350-357, 2008.

[26] M. Niemeijer, B. van Ginneken, S. R. Russell, M. S. Suttorp-Schulten, and M. D. Abramoff. "Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis" *Investigative ophthalmology & visual science*, vol. 48, no. 5, pp. 2260-2267, 2007.

[27] M. García, C. I. Sánchez, M. I. López, D. Abásolo, R. Hornero, "Neural network based detection of hard exudates in retinal images", *Computer Methods and Programs in Biomedicine*, vol. 93, no. 1, pp. 9-19, 2009.

[28] H. F. Jaafar, A. K. Nandi and W. Al-Nuaimy, "Detection of exudates in retinal images using a pure splitting technique*", Proceedings of the IEEE Engineering in Medicine and Biology Society*, pp. 6745-6748, 2010.

[29] S. Ali, D. Sidibé, K. M. Adal, L. Giancardo, E. Chaum, T. P. Karnowski, and F. Mériaudeau, "Statistical atlas based exudate

segmentation", *Computerized Medical Imaging and Graphics*, vol. 37, no. 5, pp. 358-368, 2013.

[30] B. Harangi, R. J. Qureshi, A. Csutak, T. Peto, A. Hajdu, "Automatic Detection Of The Optic Disc Using Majority Voting In A Collection Of Optic Disc Detectors," *Proceedings of the IEEE International Symposium on Biomedical Imaging,* pp. 1329-1332, 2010.

[31] R. J. Qureshi, L. Kovacs, B. Nagy, B. Harangi, A. Hajdu, "Automatic detection of the fovea and optic disk in digital retinal images by combining algorithms," *Proceedings of the International Conference on Applied Informatics*, pp. 175-184, 2010.

[32] L. Kovacs, B. Harangi, B. Nagy, A. Hajdu, R. J. Qureshi, "Gráf alapú vakfolt és sárgafolt detektálás retina felvételeken," *Proceedings of the Képfeldolgozók és Alakfelismerők Országos Konferenciája*, pp. 329-341, 2011.

[33] B. Harangi, A. Hajdu, "Improving the accuracy of optic disc detection by finding maximal weighted clique of multiple candidates of individual detectors," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 602-605, 2012.

[34] B. Harangi, A. Hajdu, "Detection of the optic disc in fundus images by combining probability models", *Computers in Biology and Medicine*, accepted.

[35] B. Harangi, B. Antal, A. Hajdu, "Automatic Exudate Detection with Improved Naïve-Bayes Classifier," *Proceedings of the IEEE International Symposium on Computer-Based Medical System*, pp. 1-4, 2012.

[36] B. Harangi, I. Lazar, A. Hajdu, "Automatic Exudate Detection Using Active Contour Model and Regionwise Classification," *Proceedings*

of the IEEE Annual International Conference of the Engineering in Medicine and Biology Society, pp. 5951-5954, 2012.

[37] B. Harangi, A. Hajdu, "Aktív kontúr használatával és régió alapú osztályozással pontosított exudátum detektáló algoritmus," *Proceedings of the Képfeldolgozók és Alakfelismerők Országos Konferenciája*, pp. 379-392, 2013.

[38] B. Harangi, A. Hajdu, "Detection of Exudates in Fundus Images Using a Markovian Segmentation Model," *Proceedings of the IEE Annual International Conference of the Engineering in Medicine and Biology Society*, pp. 130-133, 2014.

[39] B. Harangi, A. Hajdu, "Improving automatic exudate detection based on the fusion of the results of multiple active contours," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 45-48, 2013.

[40] B. Harangi, A. Hajdu, "Automatic exudate detection by fusing multiple active contours and regionwise classification," *Computers in Biology and Medicine*, vol. 54, pp. 156-171, 2014.

[41] P. J. Saine, "Focusing the fundus camera: A clinical approach", *Journal of Opthalmic Photography*, vol. 14, pp. 7-24, 1992.

[42] T. Louwies, L. I. Panis, M. Kicinski, P. De Boever, T. S. Nawrot, "Retinal microvascular responses to short-term changes in particulate air pollution in healthy adults", *Environ Health Perspect*, vol. 121, no. 9, pp. 1011-1016, 2013.

[43] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, I. J. Constable, "Retinal image analysis: concepts, applications and potential", *Progress in retinal and eye research*, vol. 25, no. 1, pp. 99-127, 2006.

[44] M. D. Saleh, C. Eswaran, "An automated decision-support system for non-proliferative diabetic retinopathy disease based on MAs and HAs detection", *Computer Methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 186-196, 2012.

[45] P. J. Saine, M. E. Tyler, "Ophthalmic photography: retinal photography, angiography, and electronic imaging", *Butterworth-Heinemann*, 2002.

[46] L. Gagnon, M. Lalonde, M. Beaulieu, M. C. Boucher, "Procedure to detect anatomical structures in optical fundus images", *Proceedings of the SPIE Medical Imaging*, vol. 4322, pp. 1218-1225, 2001.

[47] T. Kauppi, V. Kalesnykiene, J. K. Kamarainen, L. Lensu, I. Sorri, H. Uusitalo, H. Kälviäinen, and J. Pietilä, "DIARETDB0: Evaluation Database and Methodology for Diabetic Retinopathy Algorithms", *Technical report*

[48] T. Kauppi, V. Kalesnykiene, J. K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, and J. Pietilä, "Diaretdb1 diabetic retinopathy database and evaluation protocol", *Proceedings of the Conference on Medical Image Understanding and Analysis*, pp. 61-65, 2007.

[49] J. 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*, vol. 23, pp. 501–509, 2004.

[50] http://messidor.crihan.fr

[51] L. Giancardo, F. Meriaudeau, T.P. Karnowski, Y. Li, S. Garg, Jr K. W. Tobin and E. Chaum, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets", *Medical Image Analysis*, vol. 16, no. 1, pp. 216-226, 2012.

[52] R. J. Trudeau, “Introduction to Graph Theory”, *Dover Books on Mathematics Series*, 1993.

[53] S. G. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation”, *IEEE Transaction on Pattern Analysis Machine Intelligence*, vol. 11, pp. 674–693, 1989.

[54] J. Canny, “A computational approach to edge detection”, *IEEE Transaction on Pattern Analysis Machine Intelligence*, vol. 6, pp. 679-698, 1986.

[55] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, “Comparing images using the Hausdorff distance”, *IEEE Transaction on Pattern Analysis Machine Intelligence*, vol. 15, pp. 850–863, 1993.

[56] R. C. Gonzales and R.E. Woods, “Digital Image Processing”, *Addison Wesley Publishing Corporation*, Massachusetts, 1993.

[57] N. Otsu, “A Threshold Selection Method from Gray-Level Histograms”, IEEE Transaction on System Man and Cybernetics, vol. 9, no. 1, pp. 62-66, 1979.

[58] Hough, P.V.C., “Method and Means for Recognizing Complex Patterns”, *U.S. Patent No. 3069654*, 1962.

[59] M. Niemeijer, J. J. Staal, B. van Ginneken, M. Loog, M. D. Abramoff, “Comparative study of retinal vessel segmentation methods on a new publicly available database”, *Journal of Medical Imaging*. vol. 5370, pp. 648-656, 2004.

[60] R. O. Duda and P. E. Hart, “Use of the Hough transformation to detect lines and curves in pictures”, *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[61] D. Kumlander, “A new exact algorithm for the maximum-weight clique problem based on a heuristic vertex-coloring and a backtrack search”, *Proceedings of the 5th International Conference on Modelling, Computation and Optimization in Information Systems and Management Sciences*, pp. 202-208, 2004.

[62] R. T. Clemen and R.L. Winkler, “Aggregating probability distributions.” *Advances in Decision Analysis*, pp. 154-176, 2007.

[63] R. A. Jacobs, “Methods for combining experts' probability assessments.” *Neural computation*, vol. 7, no. 5, pp. 867-888, 1995

[64] J. Kittler, M. Hatef, R.P.W. Duin and J. Matas. "On combining classifiers." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226-239, 1998.

[65] M. Stone, “The opinion pool”, *The Annals of Mathematical Statistics*, vol. 32, no. 4, pp. 1339-1342, 1961.

[66] P.A. Morris, “Decision analysis expert use” *Management Science*, vol. 20, no. 9, pp. 1233-1241, 1974.

[67] P.A. Morris, “Combining expert judgments: A Bayesian approach” *Management Science*, vol. 23, no. 7, pp. 679-693, 1977.

[68] K. Pearson, “Note on regression and inheritance in the case of two parents.” *Proceedings of the Royal Society of London*, vol. 85, pp. 240-242, 1895.

[69] D. Howard, P. L. Lam, “Managerial economics: an analysis of business issues”, *Pearson Education*, 2001.

[70] N. Friedman, D. Geiger, M. Goldszmidt, “Bayesian network classifiers.” *Machine learning*, vol. 29, no. 2, pp. 131-163, 1997.

[71] D. M. Chickering, "Learning Bayesian networks is NP-complete." *Springer - Learning from data: Artificial Intelligence and Statistics V*, pp. 121-130, 1996.

[72] H. Zhang, L. Jiang, J. Su, "Hidden Naive Bayes." *Proceedings of the National Conference on Artificial Intelligence*, vol. 20, no. 2, pp. 919-924, 2005.

[73] J. Xu, O. Chutatape, P. Chew, "Automated optic disk boundary detection by modified active contour model", *IEEE Transactions on Biomedical Engineering*, vol. 5,4 no. 3, pp. 473-482, 2007.

[74] H. Li, O. Chutatape, "Boundary detection of optic disk by a modified ASM method", *Journal of the Pattern Recognition*, vol. 36, no. 9, pp. 2093-2104, 2003.

[75] M. Sonka, V. Hlavac and R. Boyle, "Image Processing, Analysis, and Machine Vision - 2nd ed.", *Thomson-Engineering*, 1998.

[76] J. B. Martinkauppa and M. Piettikäinen, "Facial Skin Color Modeling", *Handbook of face recognition*, Springer, pp. 113-135, 2005.

[77] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Internation Journal of Computer Vision*, vol. 1, pp. 321–331, 1987.

[78] K. A. Goatman, A. D. Whitwam, A. Manivannan, J. A. Olson and P. F. Sharp, "Colour normalisation of retinal images", *Proceedings of the Conference on Medical Image Understanding and Analysis*, pp. 49-52, 2003.

[79] A. Frame, P. Undrill, M. Cree, J. Olson, K. McHardy, P. Sharp and J. Forrester, "A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein

angiograms", *Computers in Biology and Medicine*, vol. 28, pp. 225–238, 1998.

[80] F. Meyer, "Contrast feature extraction", *Proceedings of the Symposium on Quantitative Analysis of Microstructures in Material Science, Biology and Medicine*, pp. 374-380, 1977.

[81] C. Xu and J.L. Prince, "Gradient Vector Flow: A New External Force for Snakes", *Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 66-71, 1997.

[82] M. Lesko, Z. Kato, A. Nagy, I. Gombos, Zs. Török, L. Vígh Jr. and L. Vígh, "Live Cell Segmentation in Fluorescence Microscopy via Graph Cut", *Proceedings of International Conference on Pattern Recognition*, pp. 1485-1488, 2010.

[83] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, No. 6, pp. 721–741, 1984.

[84] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, No. 2, pp. 147–159, 2004.

[85] R. T. Whitaker, "A Level-Set approach to 3D reconstruction from range data", *International Journal of Computer Vision*, vol. 29, no. 3, pp. 203-231, 1998.

[86] S. L. Chiu, "Fuzzy model identification based on cluster estimation", *Journal of intelligent and Fuzzy systems*, vol. 2, no. 3, pp. 267-278, 1994.

[87] W. Frei and C. Chen, “Fast Boundary Detection: A Generalization and New Algorithm”, *IEEE Transaction Computers*, vol. 26, no. 10, pp. 988-998, 1977.

[88] A. A. A. Youssif, A. Z. Ghalwash and A. S. Ghoneim, “Comparative Study of Contrast Enhancement and Illumination Equalization Methods for Retinal Vasculature Segmentation”, *Proceedings of International Biomedical Engineering Conference*, pp. 21-24, 2006.

[89] A. A. A. Youssif, A. Z. Ghalwash and A. S. Ghoneim, “A Comparative Evaluation of Preprocessing Methods for Automatic Detection of Retinal Anatomy”, *Proceedings of International Conference on Informatics and Systems*, pp. 24-26, 2007.

[90] M. Foracchia, E. Grisan and A. Ruggeri, “Luminosity and contrast normalization in retinal images”, *Journal of Medical Image Analysis*, vol. 9, no. 3, pp. 179-90, 2005.

[91] H. Peregrina-Barreto, A. M. Herrera-Navarro, L. A. Morales-Hernández and I. R. Terol-Villalobos. “Morphological rational operator for contrast enhancement”, *Journal of the Optical Society of America A*, vol. 28, no. 3, pp. 455-464, 2011.

[92] G. H. John and P. Langley. "Estimating continuous distributions in Bayesian classifiers", Proceedings of *Uncertainty in Artificial Intelligence*, pp. 338-345, 1995.

[93] R. Polikar, “Ensemble based system in decision making”, *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21-45, 2006.

[94] M. D. Abramoff, M. Niemeijer, M. S. Suttorp-Schulten, M. A. Viergever, S. R. Russell and B. van Ginneken, “Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes”, *Journal of Diabetes Care*, vol. 31, no. 2, pp. 193-198, 2008.

Appendix

A Summary

In this PhD dissertation two methods, which were elaborated to meet the demands arising from practical application, have been introduced in connection with the automatic detection of one or several objects in digital images. An automated diabetic retinopathy (DR) screening system has to be capable of detecting automatically the abnormalities caused by DR and to localize the anatomical parts of the retina in fundus image. Thus, the localization of the optic disc (OD) and detecting the signs of DR in the images taken of the fundus can be made feasible by integrating the elaborated methods into a DR screening system.

In this thesis first, a complex method elaborated for locating a single object which appears in a digital image has been introduced. As an individual detector may be wrong when it localizes the object, we studied and adapted some of the state-of-the-art OD detectors and finally organized them into an ensemble framework in order to combine their strengths and maximize the accuracy of the localization of the OD.

The elaborated method is the result of multi-stage development process, during which two further approaches have been developed based on different fusion techniques. These two approaches have been also introduced in this dissertation as well. The applicability of the fusion method depends on the number of possible locations given by the individual algorithms. Thus, we distinguished the following cases: one algorithm puts up one candidate, one algorithm puts up several candidates and one algorithm assigns a confidence value to each pixel of the image. In the last case, the probability value assigned to the pixel indicates the certainty that the given pixel can be regarded as the center point of an object.

In the three cases mentioned above, the following approaches have been elaborated; majority voting-based complex system, a method based on search of maximum weighted clique in weighted graph and a method based on probability maps fused by the Bayesian model, respectively. In this dissertation, following the detailed introduction of the elaborated approaches, the connection between the quantity of the information deriving from the individual algorithms and the accuracy of the localization of the object has been demonstrated. Regarding the practical application of the last method, the accuracy of the localization of the optic disc (OD) is close to the 100 percent accuracy.

The method for the detection of objects appearing in image multiple times, which combines the mainstream approaches within a single framework, has been proposed as the second main focus of this thesis. The elaboration of the algorithm is justified by a practical problem which is the detection of a certain type of signs called exudates, which come into being at the early stage of diabetic retinopathy and which are visible in the image taken of the retina.

B Összefoglaló

Ezen PhD disszertációban két módszert mutattunk be digitális képen történő objektum(ok) automatikus detektálásához kapcsolódóan. Ezen módszerek kidolgozását gyakorlati alkalmazás fejlesztése során felmerülő igények kapcsán dolgoztuk ki. Egy a cukorbetegség szemszövődményeit automatikusan szűrő rendszernek a betegség okozta szemszövődmények detektálásán kívül egyéb fontos funkcionalitásai is vannak, mint például a retina anatómiai képleteinek lokalizálása. Ezért a dolgozatban bemutatásra került módszerek egy ilyen szűrő rendszer felállításához, mint szükséges és lehetséges alkotóelemek lettek kidolgozva. A módszerek egyrészt alkalmasak a szemfenék felvételeken a vakfolt helyének meghatározása, másrészt a kapcsolódó elváltozások jelenlétének detektálására.

A dolgozatban elsőként, általános terminológiát használva, egy egyszeres előfordulási valószínűséggel megjelenő, egyetlen objektum detektálására kidolgozott összetett módszer került bemutatásra. Mivel az egyszerű detektáló-algoritmusok hibázhatnak az objektum helyének meghatározása során, ezért megvizsgáltuk ezen algoritmusok adaptálhatóságát, és hogy milyen módszerek mentén szervezhetnénk őket össze egy összetett rendszerbe abból a célból, hogy kombináljuk azok erősségeit és maximáljuk a lokalizálás pontosságát.

A kidolgozott módszer egy több lépcsős fejlesztés eredménye, mely fejlesztés során további két, különböző fúziós alapokon nyugvó megközelítés is született, amelyek szintén bemutatásra kerültek A kidolgozott fúziós módszerek alkalmazhatósága függ attól, hogy az egyedi algoritmusok hány lehetséges középpontot jelölnek meg a működésük végeztével. Így tehát a következő esetek különböztettük meg: 1 algoritmus – 1 jelöltet, 1 algoritmus – több jelölt, 1 algoritmus – valószínűségi mező. Az utóbbi esetben a kép minden egyes pixeléhez

rendelt valószínűségi érték azt fejezi ki, hogy az adott pixel mennyire biztosan tekinthető az objektum középpontjának.

A fenti esetekhez a felsorolásnak megfelelően a következő megközelítések kerültek bemutatásra: egyszerű többségi szavazáson alapuló összetett módszer, súlyozott gráfban történő maximális súlyú részgráf keresésén alapuló módszer, illetve valószínűségi mezők Bayesian modell használatával történő fuzionálásán alapuló módszer. A dolgozatban a kidolgozott megközelítések részletes bemutatását követően láthatóvá vált, hogy milyen kapcsolat van az egyedi algoritmusokból kinyert információmennyiség és az objektum detektálásának a pontossága között. A módszer gyakorlati alkalmazását tekintve, a retina képen történő vakfolt helyének meghatározása megközelítette a 100%-os pontosságot.

A dolgozat másik fontos eredménye egy olyan összetett algoritmus megalkotása volt, amely magas pontossággal képes eldönteni egy adott bemeneti kép esetén, hogy olyan objektumok, amelynek előfordulása egy képen belül többszörös is lehet, szerepel-e a képen vagy sem. Amennyiben az objektumok valóban megtalálhatók a felvételen, akkor az algoritmus kimenete a szegmentált objektumokat a lehető legprecízebb kontúrral szolgáltatja. Az algoritmus kidolgozását gyakorlati probléma indokolta, mégpedig a diabéteszes retinopátia korai stádiumában a retina felvételen megjelenő, úgynevezett exudátum típusú változások jelenlétének detektálása.

List of publications

Journal papers

B. Harangi, A. Hajdu, “Detection of the Optic Disc in Fundus Images by Combining Probability Models,” *Computers in Biology and Medicine*, accepted.

B. Harangi, A. Hajdu, “Automatic exudate detection by fusing multiple active contours and regionwise classification,” *Computers in Biology and Medicine*, vol. 54, pp. 156-171, 2014, IF. 1.475.

R. J. Qureshi, L. Kovacs, B. Harangi, B. Nagy, T. Peto, A. Hajdu, “Combining algorithms for automatic detection of optic disc and macula in fundus images,” *Journal of Computer Vision and Image Understanding*, vol. 116, no. 1, pp. 138-145, 2012, IF. 1.232.

Conference proceedings

B. Harangi, A. Hajdu, “Detection of Exudates in Fundus Images Using a Markovian Segmentation Model,” *Proceedings of the IEE Annual International Conference of the Engineering in Medicine and Biology Society*, pp. 130-133, 2014.

B. Harangi, A. Hajdu, “Improving automatic exudate detection based on the fusion of the results of multiple active contours,” *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 45-48, 2013.

B. Harangi, A. Hajdu, “Aktív kontúr használatával és régió alapú osztályozással pontosított exudátum detektáló algoritmus,” *Proceedings of the Képfeldolgozók és Alakfelismerők Országos Konferenciája*, pp. 379-392, 2013.

B. Harangi, I. Lazar, A. Hajdu, “Automatic Exudate Detection Using Active Contour Model and Regionwise Classification,” *Proceedings of the IEEE Annual International Conference of the Engineering in Medicine and Biology Society*, pp. 5951-5954, 2012.

B. Harangi, B. Antal, A. Hajdu, “Automatic Exudate Detection with Improved Naïve-Bayes Classifier,” *Proceedings of the IEEE International Symposium on Computer-Based Medical System*, pp. 1-4, 2012.

B. Harangi, A. Hajdu, “Improving the accuracy of optic disc detection by finding maximal weighted clique of multiple candidates of individual detectors,” *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 602-605, 2012.

L. Kovacs, B. Harangi, B. Nagy, A. Hajdu, R. J. Qureshi, “Gráf alapú vakfolt és sárgafolt detektálás retina felvételeken,” *Proceedings of the Képfeldolgozók és Alakfelismerők Országos Konferenciája*, pp. 329-341, 2011.

R. J. Qureshi, L. Kovacs, B. Harangi, B. Nagy, A. Hajdu, “Graph Based Detection of optic disc and fovea in retinal images”, *Proceedings of International Workshop on Soft Computing Applications*, pp. 143-148, 2010.

R. J. Qureshi, L. Kovacs, B. Nagy, B. Harangi, A. Hajdu, “Automatic detection of the fovea and optic disk in digital retinal images by combining algorithms,” *Proceedings of the International Conference on Applied Informatics*, pp. 175-184, 2010.

B. Harangi, R. J. Qureshi, A. Csutak, T. Peto, A. Hajdu, “Automatic Detection Of The Optic Disc Using Majority Voting In A Collection Of Optic Disc Detectors,” *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 1329-1332, 2010.

Other conference proceedings (not part of thesis)

J. Toth, L. Kovacs, B. Harangi, Cs. Kiss, A. Mohacsi, Z. Orosz, A. Hajdu, "An Online Benchmark System for Image Processing Algorithms", *Proceedings of the IEEE International Conference on Cognitive Infocommunications*, pp. 377-382, 2014.

B. Harangi, B. Nagy, A. Hajdu, 'Improving the detection of excessive activation of ciliaris muscle by clustering thermal images", *Proceedings of the Conference* of *Quantitative InfraRed Thermography*, pp. 1-6, 2012.

B. Nagy, B. Harangi, B. Antal, A. Hajdu, "Ensemble-based exudate detection in color fundus images", *Proceedings of the International Symposium on Image and Signal Processing and Analysis,* pp. 700-703, 2011.

B. Harangi, T. Csordás, A. Hajdu, "Detecting the excessive activation of the ciliaris muscle on thermal images", *Proceedings of the International Symposium on Applied Machine Intelligence and Informatics*, pp. 329-331, 2011.

B. Harangi, T. Csordás, A. Hajdu, "Detecting the excessive activation of the ciliaris muscle on thermal images", *Proceedings of the International Conference on Applied Informatics*, pp. 449-450, 2010.

Harangi B., Csordás T., Hajdu A., "A ciliaris izom túlzott működésének vizsgálata szomatoinfrával készített képeken", *Proceedings of the Képfeldolgozók és Alakfelismerők Társaságának Országos Konferenciája*, pp. 1-6, 2009.