

Lappeenranta University of Technology  
Faculty of Industrial Engineering and Management  
Degree Program in Computer Science

Bachelor's Thesis

**Teemu Huovinen**

**SENSITIVITY OF RETINAL IMAGE SEGMENTATION ON  
GROUND TRUTH ACCURACY**

Examiners: Lasse Lensu D.Sc. (Tech.)

Supervisor: Lasse Lensu D.Sc. (Tech.)

# **ABSTRACT**

Lappeenranta University of Technology  
Faculty of Industrial Engineering and Management  
Degree Program in Computer Science

Teemu Huovinen

**Sensitivity of retinal image segmentation on ground truth accuracy**

Bachelor's Thesis

2014

20 pages, 13 figures, 1 table, and 2 appendices.

Examiners: Lasse Lensu D.Sc. (Tech.)

Keywords: eye fundus, image segmentation, sensitivity analysis, ground truth

# **TIIVISTELMÄ**

Lappeenrannan teknillinen yliopisto  
Tuotantotalouden tiedekunta  
Tietotekniikan koulutusohjelma

Teemu Huovinen

**Silmänpohjakuvien segmentoinnin herkkyyssä pohjamerkintöjen tarkkuudelle**

Kandidaatin työ

2014

20 sivua, 13 kuvaaa, 1 taulukko ja 2 liitettä.

Tarkastajat: TkT Lasse Lensu

Hakusanat: silmänpohja, segmentointi, herkkyyssanalyysi, pohjamerkintä

Keywords: eye fundus, image segmentation, sensitivity analysis, ground truth

## PREFACE

I would like to thank the Laboratory of Machine Vision and Pattern Recognition of Lappeenranta University of Technology for giving me this opportunity to carry out my research. I would like to give special thanks to D.Sc. Lasse Lensu for the endless guidance and support during my summers as an intern.

I would like to thank University of Bristol for their retinal image database and its accurate ground truth data, which was extensively used in this thesis.

Lappeenranta, October 19th, 2014

*Teemu Huovinen*

# CONTENTS

<b>1 INTRODUCTION</b>	<b>7</b>
1.1 Background . . . . .	7
1.2 Objectives and Restrictions . . . . .	7
1.3 Structure of the Thesis . . . . .	8
<b>2 RETINAL IMAGES AND THEIR SEGMENTATION</b>	<b>9</b>
2.1 Structure of the eye fundus . . . . .	9
2.2 Optic disc detection . . . . .	10
2.3 Blood vessel detection . . . . .	10
2.4 Color transform . . . . .	12
2.5 Unsupervised methods . . . . .	14
2.6 Supervised methods . . . . .	14
2.6.1 Bayesian classification using Gaussian Mixture Models . . . . .	14
2.6.2 Naive Bayes . . . . .	14
<b>3 SENSITIVITY ANALYSIS OF IMAGE FEATURES</b>	<b>16</b>
3.1 Evaluation methods . . . . .	16
3.2 Effects of inaccurate ground truth . . . . .	17
<b>4 EXPERIMENTS AND RESULTS</b>	<b>18</b>
4.1 The Ground Truth . . . . .	18
<b>REFERENCES</b>	<b>18</b>

## ABBREVIATIONS AND SYMBOLS

<b>RGB</b>	Red Green Blue
<b>TP</b>	True Positive
<b>TN</b>	True Negative
<b>FP</b>	False Positive
<b>FN</b>	False Negative
<b>FPR</b>	False Positive Rate
<b>FNR</b>	False Negative Rate
<b>CLAHE</b>	Contrast Limited Adaptive Histogram Equalization

# 1 INTRODUCTION

## 1.1 Background

The growing amount of diabetes patients and (arguably) more importantly the estimated amount of undiagnosed patients motivate the research for an effective mass screening method for monitoring and early detection of diabetes. The most common complication of diabetes, diabetic retinopathy, causes abnormalities in the eye, and detecting these abnormalities in the eye fundus is a promising mass screening method. [1] When developing a method for detecting these abnormalities, handmade annotations of the objects in the image are used as a ground truth to train classifiers and to evaluate the results. In eye fundus image segmentation research, ground truths are usually done by medical experts of the field (ophthalmologists) marking these abnormalities, such as exudates.

Optimal ground truth would be a pixel-accurate binary representation of the abnormalities, but as ground truths are done by a human hand, such accuracy is not possible. Because the marking of an accurate ground truth takes a good amount of time and patience, it is often necessary to have to settle for rough markings of the present abnormalities. Clusters of exudates are circled, rather than each small finding specified separately.

## 1.2 Objectives and Restrictions

This thesis will address two main questions in its research:

- How will inaccurate ground truth affect features and segmentation methods?
- How will other features than colour perform?

This thesis focuses on exudate detection in terms of segmentation performance, and Bristol database is used as it is readily available and has accurate ground truths of exudates. To enable comparison between features, colour, edge and texture features are used. Blood vessel detection is explored only to create a mask for them. A rough method for optic disk detection is also implemented as a preprocessing step for masking reasons. Structure of the eye fundus is briefly explored for context.

Both supervised and unsupervised segmentation methods are used. In supervised methods, ground truths are used to label observations as either exudate or background. In

unsupervised methods, ground truth is used to evaluate segmentation results. Best parameters for each method are chosen based on their performance.

### 1.3 Structure of the Thesis

Section 2 takes a look at the different features of eye fundus images, and how they are relevant in this thesis. It also explains the theory behind the applied pre-processing and segmentation methods. Section 3 details how sensitivity analysis is done in this thesis, and also explains the used evaluation methods. Section 4 describes the experiments in detail, and presents the results for each experiment.

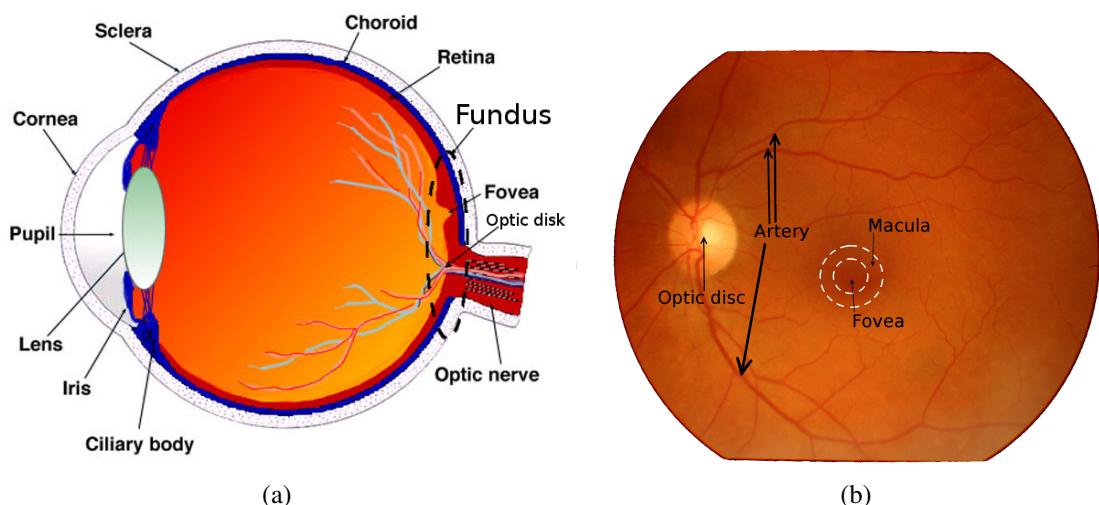
## 2 RETINAL IMAGES AND THEIR SEGMENTATION

### 2.1 Structure of the eye fundus

The function of the human eye on an abstract level can be compared to a camera. Light reflected from an object passes through the cornea, pupil and lens and is focused on the retina (the inner part of the eye). It is then processed by photoreceptors and then transmitted to the brain via the optic nerve. [2] Rather than covering the eye as a whole, this chapter will focus on the structure of the eye fundus.

The eye fundus is the area located in the retina, at the very back of the eye. The location of the eye fundus is illustrated in Figure 1a. Its most noticeable parts are the optic disk, the macula, and the arteries. Macula is a highly light sensitive area in the central region of the retina. Fovea is a round area located in the macula, which is densely populated by cones (cells sensitive to color). [3]

Optic disc is where the optic nerve and main arteries connect with the eye. There are no light sensitive cells inside the optic disc, which creates a blind spot in the retina. Arteries inside the eye together with the choroid (vascular layer surrounding the retina) provide nutritional support to the eye. [1] These parts of the eye fundus are illustrated in Figure 1b.



**Figure 1.** Structure of the human eye (a) Cross section of the eye (modified from [3]) (b) Structure of the fundus

## 2.2 Optic disc detection

Optic disc is very similar to exudates in terms of color and intensity, so detection and masking of the optic disk is an important preprocessing step in exudate detection. There are papers dedicated to the localisation of the optic disk [4], and it is also covered in papers concerning the detection of other parts of the eye fundus, such as exudates [5].

This method is based on the brightness of the optic disk, and the vertical blood vessels inside it. The horizontal image gradient is calculated using Sobel operator, the result is shown in Figure 2a. Image is then divided into slightly overlapping square areas with a side of 140 pixels (size is adjusted when operating close to image borders). The area with the highest sum of gradients is considered as region of interest, i.e. to hold the optic disk. This is because the dark blood vessels inside the bright optic disk result in a strong horizontal gradient. Images with a “camera glare”, i.e. a high intensity strip in the corner of the eye fundus are problematic, as that area also has a high horizontal gradient. Region of interest is shown in Figure 2b.

Inside this area of highest sum of horizontal gradients, the pixel with the highest intensity is considered to be inside the optic disk. This pixel is then used as a center of a circle that will mask out the optic disk. Final masking result is shown in Figure 2c.

## 2.3 Blood vessel detection

The issue of blood vessel detection in fundus images has been a popular topic of research [6] [7] [8]. In the context of this thesis, however, the purpose of blood vessel detection is to create a mask, and to use that mask to remove false positives from exudate segmentation results. For example, edge detection techniques often highlight the borders of vessels as well as exudates. For this purpose, accurate vessel detection itself is not important and including other dark areas of the image is even beneficial.

The mask is formed by first using CLAHE (contrast limited adaptive histogram equalization) [9] to enhance contrast in the green channel of the image, the result for this is shown in Figure 3a. This contrast enhanced image is then thresholded with Otsu’s method [10], which separates the image into foreground and background by minimizing the intra-class variance. This results in all the vessels and other darker areas showing as black (or background), and all brighter areas as white (foreground). This is shown in Figure 3b. To create a binary mask of the darker areas, we use the complement of this thresholded im-



**Figure 2.** Locating and masking the optic disk: (a) Horizontal gradient (b) Region of interest (c) Optic disk masked out

age. Final version of the mask is shown in Figure 3c.

This method is inadequate for blood vessel detection as it also includes other darker areas of the image, such as the fovea. As a mask however, it clearly reduces the amount of false positives in exudate segmentation results. It also doesn't remove true positives, as only the darker areas of the image are included in the mask.



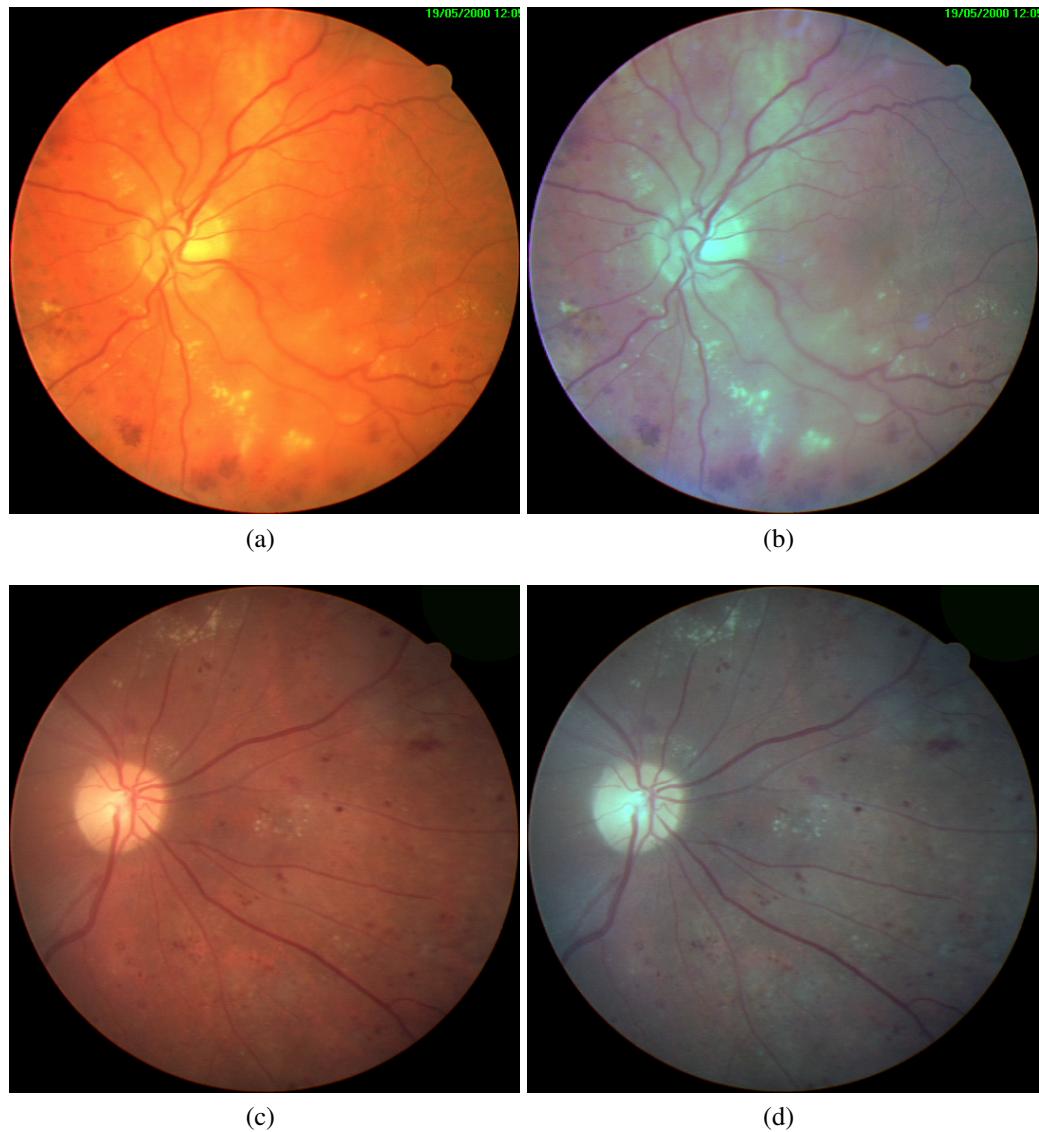
**Figure 3.** Creating the blood vessel mask: (a) CLAHE (b) Thresholding using Otsu's method (c) Final mask, complement of thresholded image

## 2.4 Color transform

The human eye is capable of correcting the effect of varying light sources (illuminants) in perception of color [11]. This is why a red chair will look (more or less) red in the sun and in the moonlight. In contrast, an image of an object taken under different light sources is perceived differently in terms of color by a computer. This is why color features need to be normalized in image sets where there's high variation in color properties.

The illumination and general color properties of the eye fundus images in Bristol database vary quite heavily. To be able to effectively use color features in teaching a classifier,

the variance between image needs to be minimized. This is achieved by estimating the illuminant by applying the gray-world assumption [12]: *the average reflectance in a scene under a neutral light source is achromatic*. Color feature normalization was then done by multiplying each channel with a coefficient defined as  $K_r = I_{avg}/r_{avg}$ , where  $I_{avg}$  is the mean of all RGB-values in the image, and  $r_{avg}$  is the mean of the specific channel. Results are shown in Figure 4.



**Figure 4.** Color space transform: Original images on the left, adjusted images on the right.

## 2.5 Unsupervised methods

An unsupervised method doesn't hold any information of the features it seeks, nor does it have any "expectations" of the image. It performs the same operations on each image with the object of clustering the data into meaningful sets.

Most edge detection methods are unsupervised algorithms. Edge is defined as a sharp change in image and color intensity, and the purpose of an edge detection algorithm is to emphasize and detect those sharp changes in an image. In an eye fundus image, this would include the exudates, but also the borders of the blood vessels and optic disc. As such, edge detection isn't sufficient for exudate detection since these unwanted regions need to be subtracted from the initial result. During this study, this is accomplished with the mask acquired in Section 2.3.

## 2.6 Supervised methods

Before a supervised method (or its classifier) can be used in classification, it needs to be trained. A specific training set of labeled images is used to teach the classifier. The method then extracts features from the training images, and generally forms a model for each used label. For example, to describe a fruit you might use its color and shape. You "teach" the method what each fruit looks like by labeling a set of training images. The classifier learns which features describe each label. Yellow and round equals orange, yellow and long equals banana, green and round equals apple. After training, the classifier can be used to classify new images.

### 2.6.1 Bayesian classification using Gaussian Mixture Models

### 2.6.2 Naive Bayes

Naive Bayes is a probabilistic classifier that gets its name from the assumption that all features are independent from each other. Each feature contributes to the classification regardless of the absence or presence of other features. As the name also implies, the Naive Bayes classifier uses the Bayes' theorem as a basis of classification. [13]

The Naive Bayes classifier calculates the posterior probability for the sample to belong

in each of the possible classes based on its features, and then classifies the sample to the class with the highest posterior probability. In the case of binary pixel classification, classes are foreground and background. In this thesis, exudates represent the foreground and the rest is considered background. The posterior probability is calculated with the Bayes' theorem as follows:

$$p(C_1) = \frac{P(C_1)p(\bar{x}|C_1)}{P(C_1)p(\bar{x}|C_1) + P(C_2)p(\bar{x}|C_2)} \quad (1)$$

where  $p(C_1)$  is the posterior probability and  $P(C_1)$  is the prior probability for the sample to belong in class one, and  $p(\bar{x}|C_1)$  denotes the conditional probability of the feature vector having the values of the sample pixel, if it were of class one.

### 3 SENSITIVITY ANALYSIS OF IMAGE FEATURES

#### 3.1 Evaluation methods

In evaluation, correctly classified samples can be split into true positives and true negatives. Errors are then similarly split into false positives and false negatives. In this thesis, true positives are findings correctly classified as exudates in the segmentation results, and true negatives are consequently findings correctly classified as background. False negatives are classified as background in results, but are actually positive (classified as exudate) in ground truth, and false positives are classified as positive in results, when they are in fact background [14].

There exists a multitude of ways to evaluate segmentation results. In this thesis, it was desired to have a single number describing the "goodness" of segmentation results to enable easier comparison and ranking of results. Different values, such as sensitivity, specificity and precision (defined in [14]) were considered, but these were ignored as they only described one aspect of the result, either the amount of samples correctly positive or negative. The used coefficient would have to describe the "goodness" in a more wholesome way, and for that reason, the Dice coefficient [15], Jaccard index and F-score were considered. These coefficients are defined as follows:

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (2)$$

$$Jaccard = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

$$F - score = \frac{2PR}{P + R} \quad (4)$$

where A is the segmented set, and B is the ground truth, P stands for precision and R stands for recall (sensitivity).

Dice coefficient and F-score values are interchangeable in practice as their values are exactly equal, and Jaccard index behaves similarly to the two. Dice coefficient was chosen as it is very simple to implement and understand. It can have values from the range [0,1], where 0 indicates no similarities, and 1 indicates perfect agreement.

### 3.2 Effects of inaccurate ground truth

When researching feature sensitivity to ground truth accuracy, standard segmentation with unsupervised methods isn't really useful as it doesn't use the ground truth. In this thesis, the ground truth is used in finding the best parameters for unsupervised methods. First, the images used for teaching are segmented with each method and a large set of parameters. The results are then evaluated with the ground truth. Parameters with the highest Dice (see Eq. 2) coefficient are then selected to be used in experiments.

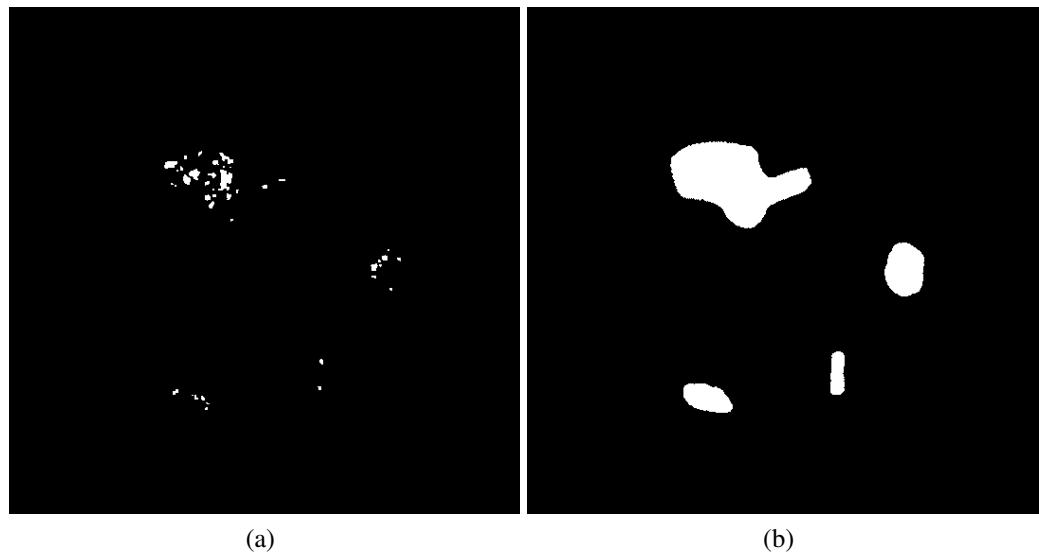
The accuracy of ground truth has a direct impact on training and performance of supervised methods. In general, inaccurate ground truth will mean a significant amount of background samples close to exudates will be categorized as positive. This will result in higher variation of feature distributions and an increase of false positives.

## 4 EXPERIMENTS AND RESULTS

### 4.1 The Ground Truth

The ground truth of exudates in Bristol database is very accurate, and to enable comparison of results and sensitivity analysis, more inaccurate ground truths were made by hand. Instead of the original color images, markings for the inaccurate ground truth were made on the black-and-white images of Bristol ground truth. This was to ensure every exudate present in the Bristol ground truth was also present in the inaccurate ground truth. Markings were done in a way that estimated the way doctors marked their findings when given the freedom to make inaccurate markings. Essentially this means that clusters of exudates are grouped together, and single exudates were more loosely circled. This is illustrated in Figure 5.

To create a basis for more comprehensive testing, different stages of inaccurate ground truths were created by dilating the accurate ground truth. Three different stages were created; one, three and five iterations of dilation by a disk-shaped structuring element, with a radius of 1 pixel. The dilation was restricted to stay inside the inaccurate ground truth.



**Figure 5.** Ground truths: (a) Original accurate ground truth (b) Inaccurate ground truth

## REFERENCES

- [1] Tomi Kauppi et al. *Eye fundus image analysis for automatic detection of diabetic retinopathy*. Lappeenranta University of Technology, 2010.
- [2] Brian A Wandell. *Foundations of vision*. Sinauer Associates, 1995.
- [3] Helga Kolb. *Simple Anatomy of the Retina* url: <http://webvision.med.utah.edu/book/part-i-foundations/simple-anatomy-of-the-retina/>.
- [4] Sribalamurugan Sekhar, Waleed Al-Nuaimy, and Asoke K Nandi. Automated localisation of retinal optic disk using hough transform. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 1577–1580. IEEE, 2008.
- [5] Thomas Walter, J-C Klein, Pascale Massin, and Ali Erginay. A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina. *Medical Imaging, IEEE Transactions on*, 21(10):1236–1243, 2002.
- [6] Frédéric Zana and Jean-Claude Klein. A multimodal registration algorithm of eye fundus images using vessels detection and hough transform. *Medical Imaging, IEEE Transactions on*, 18(5):419–428, 1999.
- [7] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram van Ginneken. Ridge-based vessel segmentation in color images of the retina. *Medical Imaging, IEEE Transactions on*, 23(4):501–509, 2004.
- [8] Meindert Niemeijer, Joes Staal, Bram van Ginneken, Marco Loog, and Michael D Abramoff. Comparative study of retinal vessel segmentation methods on a new publicly available database. In *Medical Imaging 2004*, pages 648–656. International Society for Optics and Photonics, 2004.
- [9] Ali M Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, 38(1):35–44, 2004.
- [10] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.
- [11] David H Foster. Color constancy. *Vision research*, 51(7):674–700, 2011.

- [12] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980.
- [13] Kevin P Murphy. Naive bayes classifiers. *University of British Columbia*, 2006.
- [14] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.
- [15] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.