FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Digital Image Colorimetry for Determination of Sulfonamides in Water

Paulo Jorge Teixeira Silva

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Hélder Filipe Pinto de Oliveira

Co-Supervisor: Marcela Alves Segundo

Co-Supervisor: Sílvia Neto Bessa

Co-Supervisor: Patricia S. Peixoto

June 25, 2017

Resumo

Agentes antimicrobianos são considerados poluentes emergentes na água, devido ao seu potencial para acelerar a propagação de genes de resistência bacteriana e ao seu efeito nocivo para o ecossistema através da morte ou inibição da microbiota natural. As sulfonamidas (SAs) são um importante grupo antimicrobiano amplamente utilizado em medicina humana e veterinária. Estudos demonstraram que as SAs estão altamente presentes no solo sem bioacumulação. Além disso, estes compostos parecem ser bastante resistentes à biodegradação em águas superficiais, o que pode beneficiar a contaminação do meio aquático envolvente. Dito isto, a monitorização dos níveis de concentração de SAs na água é muito importante para determinar o risco de contaminação do meio aquático.

Vários métodos para a determinação de SAs na água já foram desenvolvidos. A maioria deles é baseada no acoplamento de cromatografia líquida de alta performance (LC) e espectrometria de massa (MS). LC-MS é frequentemente utilizado devido à sua alta sensibilidade e especificidade. No entanto, esta abordagem é muito dispendiosa e não permite análises in situ. Assim, é necessário o desenvolvimento de métodos capazes de realizar triagem no campo.

Métodos baseados em colorimetria de imagens digitais já são usados em testes de campo, análises forenses e monitorização ambiental. Estes métodos são alternativas promissoras para técnicas de triagem de campo porque são métodos rápidos, de baixo custo, portáteis e de fácil manuseamento.

Nesta dissertação, propõe-se uma técnica de colorimetria de imagens digitais que consiste em analisar a resposta de cor de imagens digitais com um algoritmo de processamento de imagem automático para estabelecer uma relação entre as características de cor de um analito e sua concentração. Garantido este relacionamento, um protótipo para uma aplicação Android é fornecido. A aplicação Android é uma alternativa rápida, fácil e eficiente para a detecção de SAs na água.

Para desenvolver a referida técnica, foram estudados e comparados vários métodos de segmentação de imagem e métodos de correção de cor, para determinar a melhor solução a implementar no produto final. O objetivo final desta dissertação foi criar um algoritmo independente do dispositivo utilizado capaz de determinar a concentração de SA numa determinada amostra de água. Este objetivo não foi totalmente alcançado porque, devido ao baixo desempenho dos métodos de correção de cor usados, a constância de cores entre os vários dispositivos utlizados não foi assegurada. No entanto, uma alternativa dependente do dispositivo utilizado foi criada. Esta alternativa pode ser usada para determinar a concentração de SA de uma determinada amostra com um erro médio inferior a 10% que é aceitável para este tipo de aplicações.

No entanto, como esta foi, segundo o meu melhor conhecimento, uma abordagem inovadora para resolver o problema em questão, estes primeiros resultados sugerem que, no futuro, é possível atingir o objetivo final, melhorando o desempenho dos métodos de correção de cor implementados.

Palavras-Chaves: Química Analítica baseada em Visão Computacional, Processamento de Imagens, Segmentação de Imagens, Correção de cor.

Abstract

Antimicrobial agents are considered emerging pollutants in water, because of their potential to accelerate spread of bacterial resistance genes, and due to their harmful effect to ecosystem through death or inhibition of natural microbiota. Sulfonamides (SAs) are an important antimicrobial group that is widely used in both human and veterinary medicine. Studies have demonstrated that SAs are highly available in the soil with no bioaccumulation. Furthermore, these compounds seem to be quite resistant to biodegradation in surface water which can promote contamination of aquatic environment. Thus, monitoring of SA levels in water is very important to determine their aquatic risk assessment.

Several methods for determination of SAs in water have been developed. Most of them are based on the coupling of high-performance liquid chromatography (LC) and mass spectrometry (MS). LC-MS is widely used due to their high sensitivity and specificity, however, this approach is very expensive and does not allow in situ analysis. Hence, development of field deployable screening methods is required.

Methods based on digital image colorimetry have been broadly applied for point-of-care tests, forensic analysis and environmental monitoring. These digital image based methods are very promising as field screening techniques because they are fast, low cost, portable and easy handling methodologies.

In this dissertation, a digital image colorimetric technique is proposed. This technique consists in analyzing the color response of digital images with an automatic image processing algorithm to establish a relationship between the color characteristics of an analyte and its concentration. Once this relationship has been established, a prototype for an Android application is provided. The Android application is a quick, easy and efficient alternative for screening SAs in water.

To develop such technique, various image segmentation methods and color correction methods were studied and compared, in order to determine the best solution to implement in the final product. One of the goals of this dissertation was to create a device-independent algorithm capable of determining the SA concentration in a given sample of water. Although a satisfactory device-independent algorithm was not achieved, due to the under-performance of the color correction methods used, color constancy throughout the multiple devices could not be guaranteed. However, a device-dependent alternative was created. This alternative can be used to determine the SA concentration of a given sample with an average error lower than 10%, which is satisfactory for such an application.

Since this was, to the best of my knowledge, an innovative approach to solve the problem at hand, these first results suggest that, in the future, it is possible to achieve the ultimate objective by improving the performance of the color correction methods implemented.

Keywords: Computer Vision Analytical Chemistry, Image Processing, Image Segmentation, Color Correction.

Agradecimentos

Depois de todos estes anos de universidade, reservo esta página da minha dissertação para agradecer às pessoas que me apoiaram e que tornaram esta fase da minha da minha vida em algo inesquecivél. Aproveito ainda para agradecer às pessoas que contribuiram para a realização deste trabalho.

Primeiramente, gostaria de agradecer ao meu orientador, o Professor Hélder Oliveira, não só pela motivação e orientação oferecida ao longo deste trabalho prático, mas também pela paciência que demonstrou ter comigo, quando nem sempre sou a pessoa mais fácil de orientar.

Gostaria de agradecer ainda à minha co-orientadora, Sílvia Bessa, pelas suas sugestões de abordagem aos problemas que me iam surgindo e também pela enorme disponibilidade que demonstrou para me ajudar.

Quero agradecer também à minha outra co-orientadora, Patricia Peixoto, pela sua boa disposição, pelos bons momentos passados durante as experiências laboratoriais e pela ajuda prestada durante o processo de escrita desta dissertação.

Deixo também uma palavra de agradecimento à Professora Marcela Segundo, pelos inúmeros conhecimentos que me transmitiu durante a duração deste trabalho.

Desejo ainda agradecer ao INESC TEC e à Faculdade de Farmácia da Universidade do Porto pela disponibilização de excelentes condições de trabalho.

Sem assinalar ninguém em particular, gostaria de agradecer aos amigos e colegas com quem partilhei estes últimos anos da minha vida e que, de uma forma ou de outra, contribuiram para criar memórias que jamais irei esquecer.

Por último, mas não menos importante, gostaria de agradecer aos meus pais, não só pelo suporte financeiro ao longo do meu percurso académico, mas acima de tudo pelo afeto e carinho que sempre demonstraram ter por mim. Espero um dia conseguir retribuir um pouco daquilo que vocês sempre me deram. Quero também deixar uma palavra especial de apreço aos meus irmãos, pela sua companhia e prontidão em ajudar-me sempre que necessário. Considero-me um pessoa extremamente feliz por fazer parte desta família. É com grande orgulho que vos dedico a conclusão da minha jornada académica.

A todos, o meu Muito Obrigado!

Paulo Silva

"Intellectual growth should commence at birth and cease only at death"

Albert Einstein

Contents

List of Figures

List of Tables

Abbreviations and Symbols

CE	Capillary Electrophoresis
CC	Color Corretion
CVAC	Computer Vision Analytical Chemistry
FAQ	Frequently Asked Questions
HPLC	High Performance Liquid Chromatography
HTTP	Hypertext Transfer Protocol
LC	Liquid Chromatography
LOD	Limit of Detection
LOQ	Limit of Quantification
MS	Mass Spectometry
MSE	Mean Squared Error
OS	Operating System
p-DMACA	p-Dimethylaminocinnamaldehyde
PHP	PHP: Hypertext Preprocessor
RGB	Red-Green-Blue
ROI	Region of Interest
RMSE	Root Mean Squared Error
StD	Standard Deviation
SA	Sulfonamide
SMZ	Sulfmethoxazole
SQL	Structured Query Language
SSE	Sum of Squares Due to Error
SSR	Sum of Squares of the Regression
SST	Total Sum of Squares
UHPLC	Ultra High Performance Liquid Chromatography
WB	White Balance
XML	Extensible Markup Language

Chapter 1

Introduction

1.1 Context

Antimicrobial resistance is one of the most serious health threats. Infections from resistant bacteria are now too common, and some pathogens have even become resistant to multiple types or classes of antibiotics (antimicrobials used to treat bacterial infections). The loss of effective antibiotics will undermine our ability to fight infectious diseases and manage the infectious complications common in vulnerable patients undergoing chemotherapy for cancer, dialysis for renal failure, and surgery, especially organ transplantation, for which the ability to treat secondary infections is crucial.

The misuse of antibiotics is the single most important factor leading to antibiotic resistance around the world. Antibiotics are among the most commonly prescribed drugs used in human medicine. However, up to 50% of all the antibiotics prescribed for people are not needed or are not optimally effective as prescribed. Antibiotics are also commonly used in food animals to prevent, control, and treat disease, and to promote the growth of food-producing animals.

The other major factor in the growth of antibiotic resistance is spread of the resistant strains of bacteria from person to person, or from the non-human sources in the environment, including food. There are four core actions that will help fight these deadly infections: preventing infections and preventing the spread of resistance; tracking resistant bacteria; improving the use of today's antibiotics and promoting the development of new antibiotics and developing new diagnostic tests for resistant bacteria.

1.2 Motivation

Antibiotic resistance is a worldwide problem. New forms of antibiotic resistance can cross international boundaries and spread between continents with ease. Many forms of resistance spread with remarkable speed. World health leaders have described antibiotic resistant microorganisms as "nightmare bacteria" that "pose a catastrophic threat" to people in every country in the world. Each year in the United States, at least 2 million people acquire serious infections with bacteria

that are resistant to one or more of the antibiotics designed to treat those infections. At least 23,000 people die each year as a direct result of these antibiotic-resistant infections and many more die from other conditions that were complicated by an antibiotic resistant infection [9].

When first-line and then second-line antibiotic treatment options are limited by resistance or are unavailable, healthcare providers are forced to use antibiotics that may be more toxic to the patient and frequently more expensive and less effective.

Because of the increasing seriousness of this threat, creating monitoring programs for determination of antibiotic levels in the environment is important. The commonly used methods to monitor the antibiotic levels in the environment are expensive and difficult to use on a large scale. This discourages the authorities from deploying these monitoring programs. Therefore, there is a need for cheaper and easier to use solutions.

Thanks to the advancements made in computer vision-based analytical chemistry, more specifically in the field of digital image colorimetry, and the advent of the smarthphone, it is now possible to create affordable and easy to use solutions for large scale environmental monitoring. A innovative approach to determine SAs concentration, based on digital image colorimetry for hand-held devices is proposed in this thesis.

1.3 Objectives

The main goal of this thesis is to create a digital image colorimetry method to determine the SAs concentration in a sample of water. To do this, we need to establish a relationship between the color characteristics in a digital image of an analyte and its concentration. After that relationship has been established, a quick, easy and efficient way to detect the presence of SAs in water will be provided. Once detected, the polluted areas may be decontaminated therefore preventing the spread of bacterial resistant genes.

1.4 Contributions

The proposed solution consists in developing a digital image-based colorimetry Android application prototype for screening of SAs in water by analyzing the color response with an automatic image processing algorithm.

1.5 Document Structure

This dissertation is divided into 6 chapters. Following this first introductory chapter, comes a chapter that describes the state-of-the-art and contains a bibliographic review of the literature upon this work was based.

In chapter 3, a description of the materials, methods and procedures used in this work to achieve the proposed objectives is given.

In chapter 4, the results of the methods and procedures described in the previous chapter are presented and discussed.

Chapter 5 offers information regarding the creation of the prototype for the Android application.

Finally, in chapter 6, the final conclusions of this thesis are divulged. A critical evaluation of the work is done and suggestions for future works are given.

Chapter 2

State of the Art

2.1 Sulfonamide Detection

Several analytical methods for determination of sulfonamides (SA) residues have been developed. These methods can be classified into three groups - quantitative, screening and confirmatory [1]:

- Quantitative methods are mostly based on chromatography and different electrokinetic separation methods such as capillary electrophoresis, that allow for the quantification of SAs. However, these methods usually require complex procedures, advanced laboratory equipment and trained operators, therefore these methods are time-consuming and expensive.
- Screening methods can detect substances or chemical constituents at the level of interest and usually provide semi-quantitative results. These methods must allow for the reliable checking of samples, and only those samples indicating the presence of the analyte should be selected for a thorough analysis. The advantages of screening methods are a low percentage of false compliant samples, short analysis time and high throughput, ease of use, acceptable selectivity and low cost.
- The main feature of confirmatory methods is their ability to reliably identify a compound. Since these methods are as expensive as they are complex, they are advised after a thorough screening process or when extremely precise measurements are required. An example of such method is mass spectrometry (MS) detection, which is used to identify and quantify a substance and can also be used to confirm a compound's molecular structure. MS allows for the confirmation of the composition of the compound and provides its structural information.

High-performance liquid chromatography (HPLC) has been widely used for the quantification of SAs due to its high sensitivity and broad linear range. Currently, the most widely used analytical methods for SAs are based on a reversed-phase HPLC (RP-HPLC) separation. Among these methods, HPLC-MS/MS, which can also be considered to be a confirmatory method, has become the main analytical technique for the identification of SAs due to its higher selectivity

and sensitivity than other instrumental methods. One of the advantages of MS/MS is the fact that complete HPLC separation of the target analytes is not necessary for selective detection. However, it is always advisable to have good chromatographic separation in order to reduce matrix effects that typically result in the suppression or, less frequently, in the enhancement of analyte signals. Therefore, short HPLC columns are generally used, considerably speeding up the analysis. Because MS detection is incompatible with most mobile phases, volatile organic modifiers should be used when HPLC is coupled to MS. In addition to HPLC, ultra-high-performance liquid chromatography (UHPLC) has increasingly been used for the rapid separation of SAs. Rapid multi-residue screening of veterinary antibiotics including SAs in food and environmental samples is one of the most promising applications of UHPLC technology. A major drawback of a multi-residue method using HPLC-MS/ MS is the excessive cost of the organic solvents and the HPLC-MS/MS equipment.

A highly selective and sensitive method was developed for the simultaneous detection of twelve SAs in beef and milk by immunoaffinity chromatography purification coupled to UHPLC-MS/MS [10]. The limit of detection (LOD) for the studied SAs ranged from 0.4 to 2.0 mg/L. Shi et al. developed a novel method for the rapid separation and determination of twenty-four SAs in bovine milk by UHPLC-MS/ MS [11]. The samples were treated with acetonitrile followed by homogenization, ultrasound treatment and centrifugation. The analytes were detected by multiple reaction monitoring (MRM) in the positive ion scan mode. Liu et al. have presented a fast and sensitive UHPLC-MS/MS method for the simultaneous quantitative determination of 16 SAs in animal feeds [12].

Capillary electrophoresis (CE) is another good quantitative analytical approach that is mainly used when only small amounts of a sample are available [1]. Some advantages of CE are its high separation efficiency, ability to analyze several samples simultaneously in multicapillary systems, and low consumption of reagents and accessories. Farooq et al. [13] used CE for the separation and detection of sulfonamides such as Sulfamethoxazole (SMZ), sulfamerazine (SMR) and sulfadimethoxine (SDM) in chicken and beef tissue samples.

Another method used for SAs determination are microbiological assays. The microbiological inhibition assays are based on the use of SA-sensitive bacteria as indicators. These assays assess the ability of microbes to reproduce in milk. When growth of these bacteria is suppressed, which is determined directly or indirectly by the metabolic activity of the bacteria, a conclusion is made about the presence of drugs. The presence of the residual substances is determined from inhibition plots of the bacterial growth obtained from an agar-diffusion method that uses cylinders, wells on the agar surface or discs of filter paper. These types of techniques usually meet the requirements of a screening approach. The microbiological method was used for screening samples of milk for their SA and antibiotic contents before these compounds could be detected by enzyme immunoassay [14] or HPLC [15].

Immunoassays are characterized by high specificity, high sensitivity, simplicity and cost effectiveness, which make them particularly useful for routine uses. These assays are based on a specific reaction between an antibody and an antigen, and they are capable of detecting the low

concentration of residues in short amount of time and often do not require difficult extraction or clean-up steps. This is a very promising technique, and large number of papers have been devoted to the detection and identification of SAs using different immuno-methods. Enzyme-linked immunosorbent assays (ELISA) are the most widely used immunoassays due to their high sample throughput. These methods can drastically reduce the number of analyses required to detect SA contamination in food samples [16].

G.G. Mohamed et al. developed a simple, sensitive and accurate spectrophotometric method for the detection of SMX in different pharmaceutical preparations [17]. The charge-transfer reactions between SAs as p-electron donors and 7,7,8,8-tetra cyanoquinodimethane, 2,3-dichloro-5,6-dicyano-1,4-benzoquinone and 2,5-dichloro-3,6-dihydroxy-1,4- benzoquinone as p-acceptors result in highly colored complexes. p-Dimethylaminocinnamaldehyde (p-DMACA) has been suggested as a spectrophotometric reagent for the detection of SAs [18] and it has been shown that in acetonitrile medium, this reagent participated in a condensation reaction with SMX and SMZ, forming colored products that can be measured spectrophotometrically. This technique can be used as a screening method since it produces a small number of false non-compliant samples, as proven by a confirmatory HPLC method [19].

In general, the reported methods of SA determination can be grouped depending on the type of analytical technique applied. The corresponding percentages are depicted in Figure 2.1. As it can be seen, HPLC-MS(/MS) is the most commonly used analytical method (38%), followed by HPLC with other detectors (22%) and electrophoresis (15%).

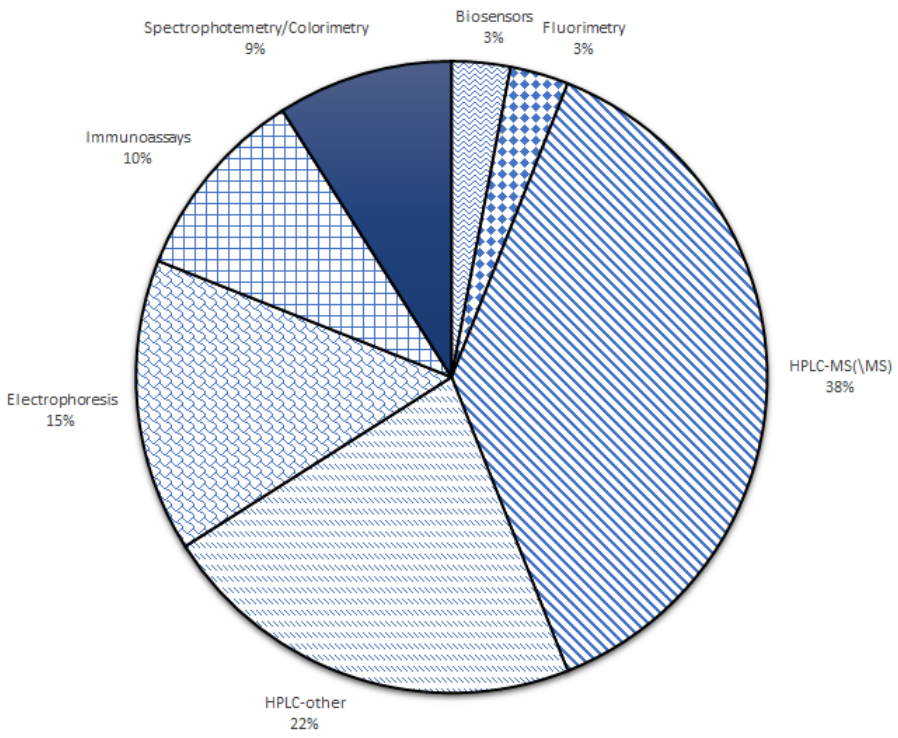


Figure 2.1: Analytical methods that have usually been applied to the detection of sulfonamides over the past years [1].

The currently proposed analytical approaches for the detection of SAs are mainly based on HPLC–MS or HPLC-MS/MS. Great advances in HPLC-MS/MS have made it a key technique for the determination of not only SAs but also other antibiotic residues [1]. These hybrid approaches have made a great contribution to the analysis of trace organic contaminants, including SAs, and have contributed to the development of multi-analyte techniques for the detection of a wide range of substances in a single analytical run [13]. These methods seem poised to be the most frequently used techniques for the purposes of analysis in the future. The main disadvantages of these methods are their complex equipment and excessive costs. This fact currently stimulates a great interest in the development of screening methods based on microbiological, immunoassays and biosensors, which have the main advantages of low cost, short analysis times and the possibility of in situ use [16].

The clear trend in this field is the miniaturization of the screening systems (chips, microarrays, microtiter plates) as well as their automation. These features will maintain the sustainable progress of these methods in the near future.

2.2 Chromogenic reaction between SAs and p-DMACA

Condensation of primary aromatic amines with aldehydes is widely used in organic synthesis (preparation of secondary amines and heterocyclic compounds, protection of aldehyde group, etc) [20] and analytical chemistry (formation of colored Schiff bases) [21]. It is known that micellar media formed by surface-active substances are capable of essentially changing the state of reactants in solution, the rate and mechanism of reactions, and their direction [22]. The catalytic effect of surfactant micelles in the condensation of aldehydes with aliphatic and aromatic amines was the subject of a number of publications. For example, Martinek et al. [23] and Krivova et al. [24] reported on the effect of sodium dodecyl sulfate micelles on the condensation of aromatic amines (aniline, benzidine, p-anisidine) with p-dimethylaminobenzaldehyde. The authors showed that the reaction rate increases due to increase in the reactant concentrations in the micellar pseudophase. The influence of salt on the mechanism and equilibrium in the condensation of hydrazine and phenylhydrazine with vanillin [25] and p-dimethylaminobenzaldehyde was also studied [26]. It was found that the apparent rate constants increased by more than 2 orders of magnitude for the reaction with hydrazine and by a factor of 20-30 for phenylhydrazine, which was attributed to increase of the reactant concentrations in the surfactant micelles.

R. K. Chernova et al. examined the effect of surfactant micelles on the condensation of p-dimethylamino cinnamaldehyde with aniline, nitroanilines, aminophenols, toluidines, and amino benzoic acids. The mechanism of condensation of primary aromatic amines with aldehydes was studied in sufficient detail [2]. The condensation includes two steps: the first step is nucleophilic addition of amine at the carbonyl carbon atom to give intermediate -amino alcohol, and the second step is elimination of water with formation of C=N bond (see Figure 2.2). The final products are the corresponding colored Schiff bases that vary according to the intensity of the reaction.

N. N. Gusakova et al. [3] studied the chromogenic reaction between SAs and p-DMACA based on condensation of these compounds in acidic medium, which results in colored Schiff bases (Figure 2.3) and concluded that the analytical effects caused by the anionic surfactants in the p-dimethylamino cinnamic aldehyde–primary aromatic amine systems form a basis for the procedures of the photometric determination of aniline and its derivatives in environmental samples, pharmaceutical formulations and biological fluids. The procedures exhibit low detection limits (fractions of the maximum permissible concentrations) and they are simple and precise.

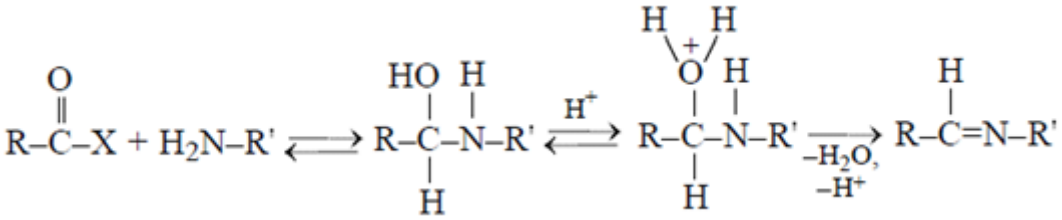


Figure 2.2: Condensation of p-dimethylaminocinnamaldehyde with aniline, nitroanilines, aminophenols, toluidines, and aminobenzoic acids [2].

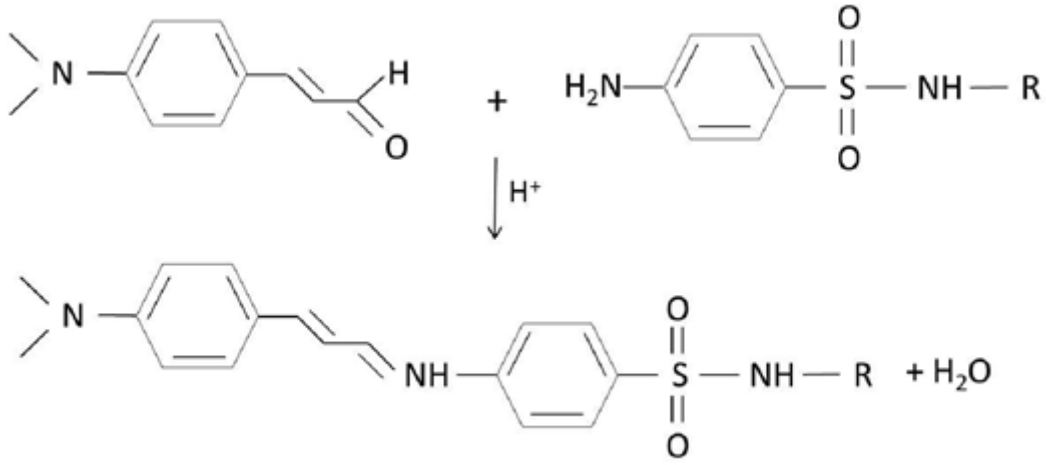


Figure 2.3: Schematic representation of DMACA and SAs reaction in acidic medium [3].

2.3 Computer vision-based analytical procedures

Chemical analysis based on color changes recorded with imaging devices is gaining increasing interest due to its several significant advantages, such as simplicity of use, and the fact that it is easily combinable with portable and widely distributed imaging devices, resulting in friendly analytical procedures in many areas that demand out-of-lab applications for in situ and real-time monitoring. These characteristics make them very promising as field screening techniques for point-of-care test, forensic analysis and environmental monitoring [4].

Among the analytical properties, color, as perceived by the observer, has been used in chemical analysis as one of the main organoleptic properties that encodes qualitative and semi-quantitative information for separation and identification of reactions or to indicate the end point in titrations. A large number of tests using both reagents in solution and reagent-impregnated papers or cloths in which a color appears or changes in the presence of an analyte have been proposed since the 18th century and up to the present time (the naked eye test).

These procedures are mainly qualitative, many of them are not very sensitive, influenced by the background lighting and subjective in any case. Moreover, the quantitative or semi-quantitative approaches are usually comparison based, using naked-eye vision or color charts. After the fast development of instruments and materials in the second half of the 20th century, these chemical tests became limited to a few, although important, specific areas such as urinalysis and immunochromatographic tests. However, the advent of color detectors and image sensors that duplicate and enhance human vision by electronically perceiving and understanding an image and the coupling of color reactions and processes with image sensors and software, along with the development of new selective chemistries, has opened the door to new strategies and opportunities in analytical chemistry. In the current field of chemical analysis, the use of color as an optical property of matter that contains chemical information and can be objectively measured is mainly connected with two domains: the use of imaging devices as analytical instruments and the development of fast analytical systems to be used in out-of-the-lab applications.

Thanks to the wide distribution and low price of these devices, they are being used as analytical tools instead of spectrophotometers, flame photometers or other equipment, and in diverse areas, including lab teaching and routine analysis. These devices are also used along with sensor systems involving the appearance or disappearance of color or any property that can be measured through a color change, i.e. luminescence, developing complete analytical systems that include chemical recognition and the transduction of the signal and especially consider the processing capacity of some current devices. Likewise, color has been used in a set of sometimes divergent and often overlapping strategies that may include: point-of-care testing [27], lab-on-a-chip [28], paper based devices [29], test express methods [30], or one-shot sensors [5]. Represent in Figure 2.4 is a typical computer vision-based analytical approach.

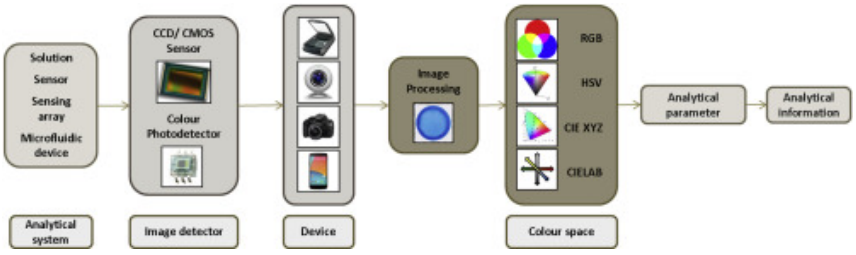


Figure 2.4: The computer vision-base analytical chemistry approach [4].

The advent of image sensors and the later explosive development of consumer electronics at the end of the 20th and beginning of the 21st century placed numerous devices with optical detection capabilities in the hands of end users. Their characteristics of sensitivity, ubiquity, afford-

ability and portability make them attractive as potential analytical tools, combining recognition of the event and signal processing in a single device. Mobile phones and tablets are particularly interesting because of the combination of their connectivity and their ability to take pictures with computational power. This made it possible to develop new concepts that can be included in health information technology and telemedicine and mobile health services through on-site processing and remote processing via data transfer to a centralized facility for archiving and analysis, e.g. rapid-diagnostic-test (RDT) for the quantification of immunoassay test strips [31]. Thanks to this rapid development of the technology available, the interest in CVAC has been exponentially growing.

The accumulated index of publications (Figure 2.5) shows that the number of papers published to date has steadily grown. Therefore, the appearance of imaging devices, combined with the analysis of color properties, has made it possible to objectify color for chemical information purposes.

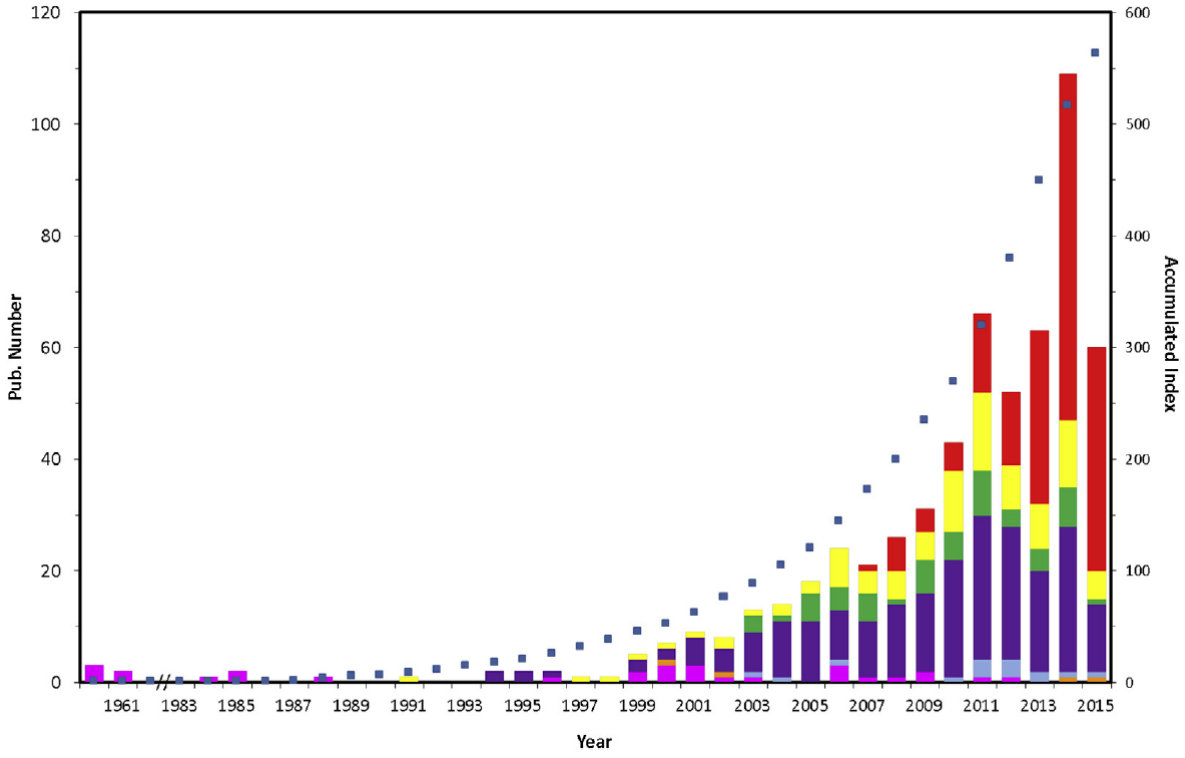


Figure 2.5: Number of CVAC paper published between 1960 and 2015 [5]

Computer vision refers to the use of devices for optical, non-contact sensing to automatically receive and interpret the image of an object in order to obtain information, control machines and/or process images [32]. It has been widely used to examine, monitor and control a very broad range of applications from scientific, commercial, industrial, and military to biomedical applications. Typically, the images are acquired with a physical image sensor, and computing hardware and software are used to manage and analyze the images with the objective of performing a prede-

fined visual task [33]. The computer vision systems needed for chemical analysis purposes, also known as computer vision-based analytical chemistry (CVAC), can vary but generally consist of a few different basic components: an analytical device that modifies its color characteristics in contact with a sample, an imaging device (digital camera, scanner, webcam, smartphone or other) connected to a computer or microcontroller, an illumination system, and control and processing software. Additionally, the use of portable and widely distributed devices such as smartphones and tablets and their corresponding application software is an attractive factor that may contribute to the spread of the systems.

Digital image processing uses a series of mathematical operations, such as noise reduction, grey level correction and blurring correction, to enhance the quality of the image and remove defects. Then an image analysis algorithm selects the regions of interest (ROI), i.e. a membrane that contains the immobilized reaction product or a colored solution after reaction, from the background and extracts the color information from them. The image segmentation process that makes it possible to obtain the ROI containing the chemical information can be done by thresholding, edge-based segmentation or region-based segmentation. From the color information, the analytical parameter is built, and qualitative or quantitative analytical information is extracted by applying single calibration, pattern recognition and/or multivariate analysis.

With these kinds of colorimetric procedures, understood here in a broad sense as those based on color measurement and able to perform chemical identification, classification and/or quantitative multi-analyte analysis, the change of color is evaluated as it relates to the magnitude of interest. However, their efficiency primarily depends on the proper choice of recognition chemistry and the sensing platforms, with the needed analytical elements: membranes, supports, cases, microfluidic devices, and others, which are used to determine the optical properties of the analyte.

2.4 Colorimetry and Color Spaces

The quantification of colors is performed by colorimetry, the science that studies the processes related to color by defining methods that allow the decomposition, analysis, and description of visible light to quantify color information for processing and analysis with several purposes in many scientific and technological fields [34].

Colors can be quantified using a number of different methods or color spaces that specify color from the three primary colors, similar to human color perception. As Helmholtz has stated about the trichromatic theory of color [34], any color can be formed by combining three properly chosen primary colors. Therefore, most of the color spaces use three component representation in order to accurately describe the human vision system.

RGB Color Space

RGB stands for red, green, and blue, which are the three color components used to completely define a color in this additive color space. The RGB color space is the most common color space

in capture devices since the human visual system works in a similar way, and, for this reason it is used as the initial color space from which the coordinates of some other spaces are obtained. Most color and image detectors have been conceived and designed to provide information in RGB coordinates. RGB is a device-dependent not globally standardized color space since the response of the device to each red, green, and blue wavelength depends on the manufacturer. For this reason, an RGB tuple can visually represent two different colors in two different devices. Valuable attempts at standardization have been made, such as sRGB and Adobe RGB [35]. However, regardless of their simplicity and direct measurement using electronic image detectors, their high illuminance dependence impedes their use in some high-resolution applications. The RGB color space has been widely used in the quantification of solutions and immobilized substances using digital processing of images and specific software that makes it possible to extract the colorimetric information by selecting the areas of interests in a image [36].

Instead of using the three coordinates separately, it is common in the literature to use a combination of the RGB coordinates or define a magnitude called intensity to define the colorimetric contribution of the incident light over the sensing elements of the detector. For example, Park et al. use the total color intensity of a pixel as a variable for oxygen determination. The total color intensity of a pixel was defined as the square root of $\mathrm{R}^2 + G^2 + B^2$.1001[37].

In a work by Meier et al. [38] the simultaneous analysis of pH and pO2 is carried out by capturing one single image and using the R/B ratio as the parameter for oxygen and G/B for pH, using a membrane with three fluorescent dyes with emission peaks matching the three channels of the detector, two for oxygen (R) and pH (G) and a third (B) as a reference.

Martinez et al. [39] also used the R/B ratio to quantify gaseous oxygen using an inkjet-printed RFID tag for monitoring.

XYZ Color Space

In 1931, the International Commission on Illumination (CIE) established the CIE XYZ or CIE 1931 color space, which is defined by two coordinates that define chromaticity, X and Z, and a third one, Y, that defines luminance, also known as tristimulus values. Hence, for a given value of luminance, Y, X and Z describe all the possible chromaticities of the space. XYZ space was designed to comprise all the colors that human visual system is able to see. In the field of colorimetry, it is usually used as an intermediate space to allow transformations between RGB and other color spaces.

The XYZ color space is mainly used in spectrophotometers and digital color analyzers that directly provide the tristimulus values of the light transmitted or reflected by the analyzed samples. Although widely used in chemical analysis, it is not perceptually uniform, which makes it more difficult to compare two colors. There are several works in the literature that use this space as a starting point in analytical systems to obtain other color coordinates by applying linear transformations. For example, Suzuki et al. [40] used this color space to analyze the results obtained by a colorimetric sensor for ions Li(I), NH4(I) and protein determination.

HSV Color Space

The HSV color space is a hue-oriented color space that completely describes a color by means of three different components: hue (H), saturation (S) and value (V). Hue is commonly defined as the dominant wavelength of the spectral radiance of a color. Saturation defines the purity of the color, and, finally, value, which is usually also referred to as brightness, represents the intensity of light present in the described color [41]. HSV values can be easily obtained from the RGB color space transforming this color space into a perceptually uniform space and thus classifying colors purely regarding their relation to how the human eyes perceives them.

This color space has been applied in the last decade in different studies of colorimetric chemical analysis. Just as with the CIE XYZ color space, the HSV is not uniform and it is not possible to establish differences between colors in order to compare them. However, the main feature and interest of this color space is the possibility of representing the hue in a single parameter, avoiding the use of redundant information when the color coordinates are used with bitonal or polytonal optical chemical sensors [42]. For users without experience, hue-based spaces provide ease of use against other color spaces. Furthermore, it has been proved that the hue component is more robust to light variations in the illumination, which could affect the measurements [43]. In this respect, Paciornik et al. [44] proposed a study of several color coordinates to model Hg(II) response using spot tests, showing that the hue coordinate has a better linear response to changes in concentration and a larger slope.

CIELAB Color Space

Another commonly used colorspace is CIELAB, defined in 1942 by Richard S. Hunter. Unlike the RGB, CIELab color is designed to approximate human vision. It aspires to perceptual uniformity, and its L component closely matches human perception of lightness. Thus, it can be used to make accurate color balance corrections by modifying output curves in the a and b components, or to adjust the lightness contrast using the L component. In the CIELAB color space, the color is described by the coordinates L*, a* and b*. The L* coordinate represents the luminance, while the a* and b* coordinates represent the change of color from red to green and yellow to blue, respectively. By using this space, or a similar space called CIELUV, a uniform perception can be obtained, meaning that a minimum change in the chromaticity coordinates produces an equal change in the visual perception of the color, not like in the color spaces mentioned prior to this, which is extremely valuable since the chromaticity is an objective specification of the quality of a color regardless of its luminance. The main advantage of the uniformity of this space is the possibility of extracting differences between two colors using L*a*b* coordinates by applying the Euclidean distance in order to determine how different they are with respect to each other. Therefore, it is especially useful for differential measurements of colors. However, in order to obtain the L*a*b* coordinates from the RGB color space, an intermediate step using the XYZ space is necessary, making it harder to process this information which requires more consumption of resources, i.e. processing power, since it's not as immediate as other spaces. In analytical chemistry, other color spaces have been used in combination with colorimetric analysis to provide information related to the analyte. As explained above, most of them are derived from the RGB or tristimulus color space (CIE XYZ), which are the most used spaces in image detectors and optical sensors.

In summary, the selection of the working color space in each case depends on multiple factors such as the instrument used, illumination or the color change developed by the optical device, which makes it necessary to study all of them to define which one best fits the model. In general, when the color change occurs in a single color shift (i.e. from clear to a determined final color), the values used are commonly RGB intensities, making it possible to apply histogram and difference map methods to analyze the images. However, if the change occurs between two different colors, as happens with bitonal optical chemical sensors (from red to yellow, purple to blue, etc.), it is common to use other color spaces such as HSV.

2.5 Image Acquisition, processing and analysis

To acquire the picture, several factors have to be taken into account since different approaches can be proposed to capture the image in optical sensor analysis, i.e. the disposition of the system, the light source, sensing element and image device, if the technique is based either on absorption, transmission or reflection. The disposition of the elements and the parameters of the imaging device used have to be considered during the acquisition procedure. Usually, in digital cameras or

smartphones with built-in cameras, a wide range of parameters can be set such as white balance, exposure time, aperture, ISO sensitivity, among others. To avoid using the devices' automatic settings, it is necessary to establish the optimal working point of the parameters.

One essential part of a computer vision system is the illumination used during the acquisition procedure considering the color dependency on light. Most of the equations in the literature related to color spaces are determined by the use of a specific illuminant as the light source for the color measurements. In some works, a specific illuminant that emits a determined wavelength is required to excite the sensing membrane luminescence or to study the behavior of the samples in controlled light conditions to simplify the equipment processing. Most of the systems developed for colorimetric determination usually reproduce an environment with controlled illumination conditions to carry out the acquisition of the measurements to obtain the color information and relate it to the analyte, avoiding ambient light interference. For instance, if commercial digital cameras or phone cameras are used, they are typically held in a fixed position with some types of clamps in an environment of controlled lighting.

2.6 Digital Image Processing

The processing and analysis of the color characteristics enclosed in digital images make it possible to obtain its analytical parameters.

Image processing involves a series of operations to enhance and extract valuable information from the acquired image. When colorimetric analysis is combined with image detectors and included as part of systems for chemical analysis, it is necessary to process the color components in an external computer, either with public domain image processing programs such as ImageJ, commercial graphics editors such as Adobe Photoshop, dedicated software, or customized processing programs (using technical computing platforms such as Matlab, Java, or LabVIEW, for example).

Among the different image processing techniques used in computer vision, the most common in CVAC procedures is image segmentation, which aims to extract the ROI containing the analytical information from the background. There are different segmentation techniques, such as pixel-based, edge-based and region-based segmentation, the most common is pixel-based segmentation and among these the thresholding technique which retains pixels from a certain part of the image with gray-scale level or color intensity, typically in separate and normalized RGB channels, higher than a certain value defined by a threshold fraction. One of the methods used for thresholding is Otsu's method, which assumes that the image contains two classes of pixels following a bi-modal histogram: ROI and background pixels. Usually edge-based segmentation techniques rely on the use of masks, which make it possible to search for sudden changes in a certain image parameter, i.e., a border or edge, through the definition of gradients over this parameter. Thus, different masks, combined with the mathematical operations associated with the mask application, define the different types of edges to be found. Different border detection methods have been

used in CVAC such as differential detection, Prewitt operators, and Sobel operators. Typically, edge detection is carried out over a grayscale image, by transformation from RGB to gray-scale, although Garcia et al. [45] used the V coordinate of the HSV space, which corresponds to image brightness, in order to reduce the computational needs.

The most common segmentation technique in CVAC is manual selection of the ROI, which is generally circular or square, in such a way that the edge of the area, where coffee-ring effects, reagent or product concentration, or leaching can occur, is not included. To this end, masks are used in the processing of the image. They are usually obtained from the subtraction of the background of the sensor, sensor array or solution, that is, a negative or binary image of the sensor. The same mask can be used for all the images of the same series, thus saving processing time.

Image analysis uses digital data obtained from ROI to extract chemical information: qualitative and/or quantitative mono or multi-analyte information or classification as well as to highlight the chemical features of images. Once the ROI is obtained, the analysis of the color components of the ROI has to be processed. The processing consists of extracting colorimetric information to be related to the analyte. Several analytical colorimetric parameters can be used for this purpose in order to establish a relationship between the analyte and color. Multiple examples for each color space were mentioned according to the chromatic coordinates employed for the quantification of the color information.

One of the most important aspects when comparing the analytical information between samples from the same series is to ensure that the lighting conditions stay constant, otherwise, for the same sample, two different values are obtained, since color is illuminant-dependent. This requirement is extremely difficult to satisfy, especially in a situation where we do not have control over it, i.e. natural lighting. The most common solution to this problem, is to implement color correction methods such as white-balance or individual color channel correction using a Macbeth Color Checker (see Figure 2.6).

Figure 2.6: Macbeth Color Checker

Between these two methods the most successful is arguably, the individual color channel correction. This method places a color checker in the same image as the analyte and uses the 24 color patches to correct the color channels, estimating the correction factor with a least-square regression algorithm using the obtained values and the known and expected values of the color

checker. Although not as effective as the previous method, white-balancing is still a relevant way to do color correction, since it requires much less processing power and is therefore much faster.

The captured RGB image can sometimes be processed by first converting the information to grayscale. As demonstrated by Murdock et al. [46] when using the same image for both analyses (RGB or grayscale), there is less sensitivity and resolution of the system with the latter. In some cases, in colorimetric analysis, there is no simple relationship between color and analytical information, as in the case of classification or identification using colorimetric arrays. In such a case, two main methods are used for the classification and the determination of analytes once the average of the colorimetric information for each spot or sensing element is obtained: principal component analysis (PCA, converting a set of related data into a set of linear uncorrelated variables) and hierarchical cluster analysis (HCA, finding a hierarchy between different clusters obtained from the data) [47].

An artificial neural network [48] and expert system [49] can also be used to establish a relationship between the sensing array and the analyte. Sometimes such a complex analysis of the information is not necessary and a relation between the used color component and the concentration of the analyte is directly settled. Normally, the average value for each color channel is used as the representative value of the specific channel. However, if any heterogeneity exists in the surface to process, the mean value can be altered by unwanted regions of the picture where the chemical membrane is not homogeneous. For this reason, some works have included the statistical mode as analytical parameter [50].

2.7 Summary

The increasing number of consumer electronic devices available on the market has lead to faster development of analytical methods based on imaging devices. The first applications to do chemical analysis appeared in the early 21st century and started with the use of scanners and, to a lesser extent, the use of CCD cameras and webcams. However, from the scanned image it is necessary to obtain color information from ROI and perform calculations to extract the analytical data. Hence, the emergence of cell phones and, later, smartphones and their processing capability, along with their ubiquity and affordable price due to mass production, has enabled an exponential development of CVAC procedures.

Chapter 3

Digital Image Colorimetry for SA Determination

In this chapter, the materials, methods and procedures used to create a digital image colorimetry system for determining the SA concentration in a given sample are presented. It starts by describing the chemical reaction used to create the colored product that the system is going to analyze. The sample extraction procedure and the image database acquisition protocol are also explained here. Following that, the algorithms used to process the images and the metrics used to evaluate them are described in detail. Finally, this chapter ends with a brief summary.

3.1 Chemical Reaction

The chemical reaction between sulfonamides and p-DMACA in a micellar medium is an important, if not the most important step of all those about to be described in this chapter, since it is the starting point for the work presented in this thesis. DMACA, an aromatic aldehyde, has been applied as a reagent for spectrophotometric determination of primary aromatic amines, including SAs [51–54]. This reaction is based on the condensation of DMACA and SAs resulting on the formation of a colored product, which is a Schiff base, in acidic medium (see Figure 3.1). The intensity of this colorimetric reaction is proportional to the SAs concentration.

Using colorimetric analysis, this colored product can be analyzed and a relationship between the intensity of the reaction and the intensity of the color can be established and studied.

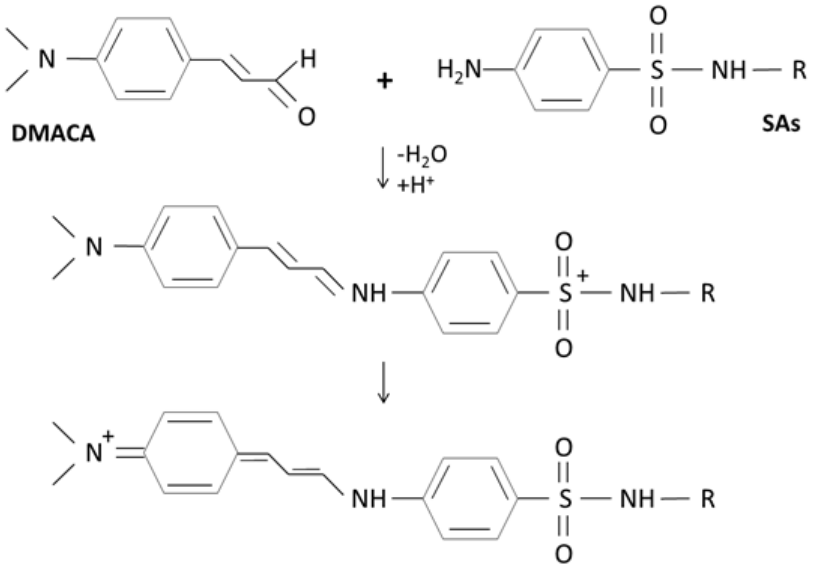


Figure 3.1: Schematic of colorimetric reaction between p-dimethylaminocinnamaldehyde (DMACA) and sulfonamides (SAs).

3.2 Materials and Methods

Analytical grade sulfamethoxazole (SMZ) was purchased from Sigma Aldrich (St Louis, MO, USA). Methanol, chloroform and water from arium water purification systems (resistivity > 18 M cm, Sartorius, Goettingen, Germany) was used for the preparation of solutions. Stock solution of SMZ 250 mg L^{-1} was prepared by dissolution of accurately weighed mass in methanol. Intermediate solution of SMZ 1.00 mg L^{-1} was prepared by dilution of stock solution in sulfuric acid 0.005 M. Stock solution of p-dimethylaminocinnamaldehyde (DMACA) 0.44 g L^{-1} was prepared by dissolving 11 mg in 3.5 mL of 0.6 M HCl solution and by completing the volume up to 25 mL with methanol. DMACA working solution (0.22 g L^{-1}) was prepared by dilution of DMACA stock solution in chloroform (1:1, v/v). SMZ working solutions in the concentration range of 5-150 μg L^{-1} were prepared in quadruplicate by dilution of appropriate volumes of intermediate solution of SMZ 1 mg L^{-1} sulfuric acid 0.005 M.

3.3 Extraction Procedure

For the extraction procedure of SMZ, the mixed-mode ion exchange polystyrene divinylbenzene sulfonated (SDB-RPS) disk from EmporeTM (Bellefonte, PA, USA) was cut (13 mm diameter) and placed inside a polypropylene disk holder, (Swinnex® filter holder, SX0001300, EMD Millipore Corporation, Billerica, MA). Four units were attached in parallel to propulsion tubes (Tygon®, 1.02 mm i.d.) fitted in a peristaltic pump (Gilson Minipuls 2, Villiers-le-Bel, France). Standards and samples (10 mL) were loaded at 0.8 mL min^{-1}. Then, disks were dried by propelling air through them at 2.0 mL min^{-1} during 6 min. After each set of four parallel extractions,

the pump tubes were washed with c.a. 5 mL of ultrapure water. Please see Figure 3.2 for a schematic representation of extraction and color evaluation procedure.

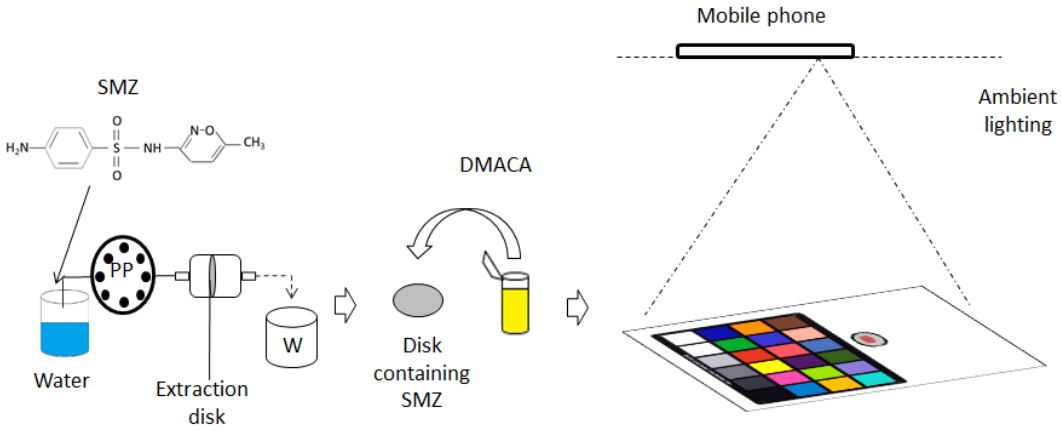


Figure 3.2: Schematic view of extraction procedure and colorimetric analysis of sulfamethoxazole.

3.4 Database Acquisition Protocol

For the acquisition of the image database, three different smartphones were used: Huawei P9 Lite, Vodafone Smart Ultra 6 and Vodafone Smart Mini (Figure 3.3). The specifications for each of these smartphones are described in Table 3.1.

(a) Huawei P9 Lite [55] (b) Vodafone Smart Ultra 6 [56] (c) Vodafone Smart Mini [57]

Figure 3.3: Smartphones used for the acquisition of the database.

Table 3.1: Smartphone specifications.

Characteristic	**Huawei P9 Lite (Camera 1)**	**Vodafone Smart Mini (Camera 2)**	**Vodafone Smart Ultra 6 (Camera 3)**
Operative System	Android OS, v6.0 (Marshmallow)	Android OS, v4.1 (Jelly Bean)	Android OS, v5.1.1 (Lollipop)
CPU	Octa-core (4x2.0 GHz Cortex-A53 + 4x1.7 GHz Cortex-A53)	1.0 GHz Cortex-A9	Octa-core (4x1.5 GHz Cortex-A53 + 4x1.0 GHz Cortex-A53)
GPU	Mali-T830MP2	PowerVR SGX531	Adreno 405
Primary Camera (back)	13 megapixels, f/2.0, autofocus, LED flash	2 megapixels, NO flash	13 megapixels, autofocus, LED flash
Secondary Camera (front)	8 megapixels, f/2.0	None	5 megapixels, NO flash

As seen in Table 3.1, the camera resolution is not the same for all the tested smartphones. This was done on purpose, because using three different image resolutions in the database acquisition process could help in the creation a more robust image processing algorithm.

The smartphones were placed on top of a homemade support (Figure 3.4) that guaranteed that the images were captured from a constant distance (21 cm) from the working bench. For this step, the acquisitions were performed under common uncontrolled artificial lighting. This was decided, because the proposed final product is an Android application capable of fast in-situ analysis where the scene illumination cannot always be controlled.

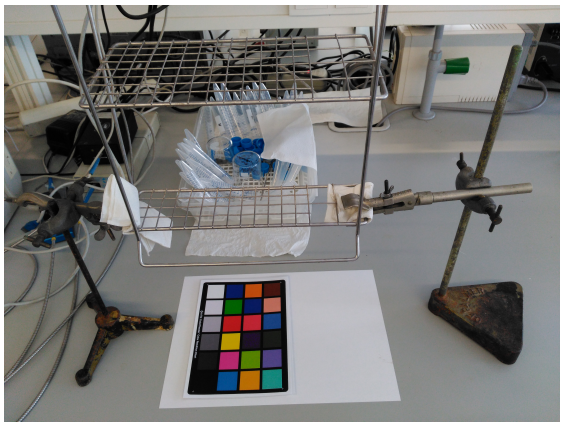


Figure 3.4: Support used for the databse acquisition

In order to eliminate illumination-based variance in the acquisitions, an unofficial high quality print of the 24-patch color rendition chart (x-Rite color rendition Chart®) was used (Figure 3.5). This color rendition chart is the industry standard for color correction in digital images and camera sensor calibration.

Figure 3.5: X-Rite color rendition Chart®

Each of the color patches in this color rendition chart has a unique color value. By using optimization algorithms and ground truth values for each color patch, it is possible to normalize the acquisition images in order to obtain color constancy throughout. This color rendition chart was placed in the working bench next to extracted disks over a white A4 sheet of paper.

The image acquisition was initiated 90s after the addition of a 10 μL of DMACA 0.22g L^{-1} solution to the disk surface. The waiting time will be later validated by a color stability test. The sequence of the image acquisition was kept constant: camera 1 (Huawei P9 Lite), camera 2 (Vodafone Smart Mini) and camera 3 (Vodafone Smart Ultra 6). For each of these cameras the images were acquired in duplicates. An example of an image from the database is shown in Figure 3.6.

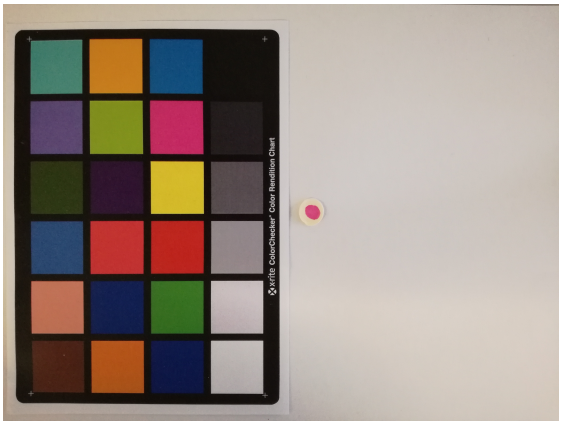


Figure 3.6: Example of an image from the acquisitions database

In this work, two different image databases were acquired. The first set of images was acquired in December 6, 2016 and was intended to be used in the development of the image processing algorithms and to establish a relation between color and concentration of SMZ. The second set of images was acquired in February 21, 2017, and was intended to validate the previous results.

The correspondence between each individual image in the database and the SA concentrations was made through the use of timestamps.

The SMZ concentration values used for both of these databases acquisitions are shown in Table 3.2.

Table 3.2: SMZ concentrations used in the database acquisition

SMZ Standard Concentration [μgL^{-1}]	**SMZ standard volume [mL]**	**SMZ concentration [μgL^{-1}]**	**Sample volume [mL]**
Blank	0	0.0	10
1000	0.025	2.5	10
1000	0.050	5.0	10
1000	0.075	7.5	10
1000	0.100	10.0	10
1000	0.125	12.5	10
1000	0.150	15.0	10
1000	0.175	17.5	10
1000	0.200	20.0	10
1000	0.225	22.5	10
1000	0.250	25.0	10
1000	0.300	30.0	10
1000	0.400	40.0	10
1000	0.500	50.0	10
1000	0.800	80.0	10
1000	1.000	100.0	10
1000	1.500	150.0	10

3.5 Image Processing and Analysis

Image processing consists in the processing of images using mathematical operations by using any form of signal processing for which the input is an image, a series of images, or a video, such as a photograph or video frame and the output of these operations may be either an image or a set of characteristics or parameters related to the image [58]. In this work, this framework is divided into two big sections: image segmentation and color correction. In computer vision, image segmentation is the process of partitioning a digital image into multiple segments with the goal of simplifying and/or changing the representation of an image into something more meaningful and easier to analyze [59, 60]. This is accomplished by assigning a label to every pixel in an image so that pixels with the same label share certain characteristics. For the colorimetric determination proposed, there are 2 key objects that should be isolated from the original image: the colored product of the chemical reaction in the polypropylene disk and the color rendition chart, including each of its 24 color patches. Segmenting the colored product in the disk will facilitate the extraction of the colorimetric data that will later be analyzed. Isolating the color patches in the color rendition chart is crucial since this will make it possible to extract the information needed to

color correct the images and guarantee color constancy throughout them. All the work pertaining to image processing was done on the MATLAB R2017a software.

3.5.1 Sample Segmentation

After carefully analyzing the images from the database, a simple method for the segmentation of the sample was developed. As shown in the example-image (Figure 3.6) the part of the image we want to isolate is a circular spot, that is darker than its surrounding pixels. With that being said, an image processing technique known as image thresholding was used. Image thresholding is one of the simplest method used in image segmentation. It converts a grayscale image into a binary image. This is accomplished by replacing each pixel in an image with a white pixel if the image intensity is greater than a fixed constant, known as the threshold, or a black pixel is the intensity is lower [59]. There are a variety of methods that can be used to define the value of this threshold. Among them, the most commonly used one is the Otsu's Method, named after Nobuyuki Otsu. This algorithm assumes that an image contains two classes of pixels (foreground and background pixels), it then calculates the optimum threshold separating the two classes [61].

Going back to Figure 3.6, it can be safely assumed that these two classes are easily identifiable and therefore easy to separate. It was established that the white sheet of paper is the background, while all the remaining objects in the image make up the foreground. For this reason, the Otsu's Method was the approach chosen to calculate the threshold value. After converting the original image into a grayscale image, this newly calculated threshold can be used to convert the image into a binary image (black and white). From here we need to apply some image processing morphological operations such as image opening and closing, image dilation and erosion in order to remove some leftover undesirable objects. The last step in this segmentation algorithm is correctly identifying the foreground objects. If everything goes according to plan there should only be two foreground objects: the color rendition chart and the sample. Since the color chart is a large rectangular object and the sample is smaller and circular, labelling these objects should be easy. Once the desired target has been found, a mask is created to crop out our region-of-interest (ROI) from the original RGB image and extract its color information. The pipeline for this algorithm is shown in Figure 3.7.

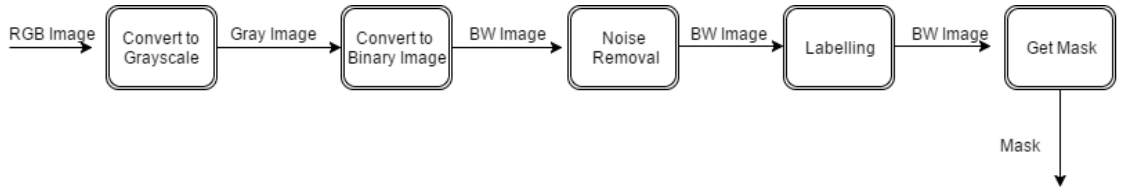


Figure 3.7: Sample segmentation algorithm pipeline

3.5.2 Color Checker Segmentation

This algorithm is very similar to the one described above in section 3.5.1, with the only difference appearing in the final step. Instead of creating a mask from the small circular object, the

mask is created using the large rectangular object and, just as before, the mask is used to crop out this new ROI from the original RGB image.

To extract any meaningful information from the color checker it is necessary to isolate all the 24 color patches, but before this process can begin, it is necessary to ensure that the all the color checker images are horizontally oriented and the white patch appears in the bottom-left corner. This is done so that the color patches can be correctly identified based on their relative position to each other. To do so, an image rotation algorithm based on the Hough Transform was proposed. The Hough transform is a feature extraction technique used in image analysis, computer vision and digital image processing [62]. The linear Hough transform algorithm uses a two-dimensional array, called an accumulator, to detect the existence of lines and is described by the following equation:

$$r = x \cdot cos(\theta) + y \cdot sin(\theta) \tag{3.1}$$

The dimension of the accumulator is equal to the number of unknown parameters, i.e., two, considering quantized values of r and θ in the pair **(r,θ)**. For each pixel at **(x,y)** and its neighborhood, the Hough transform algorithm determines if there is enough evidence of a straight line at that pixel. If so, it will calculate the parameters **(r,θ)** of that line, and then look for the accumulator's bin that the parameters fall into, and increment the value of that bin. By finding the bins with the highest values, typically by looking for local maxima in the accumulator space, the most likely lines can be extracted [59]. The result of the linear Hough transform is a matrix similar to the accumulator, in which one dimension of this matrix is the quantized angle θ and the other dimension is the quantized distance r. Each element of the matrix has a value equal to the sum of the points or pixels that are positioned on the line represented by the quantized parameters **(r, θ)** so that the elements with the highest values indicate the straightest lines represented in the input image [63]. By using the Hough Transform on a binary image of the color rendition chart, it is possible to find the angle, θ, that produces the straightest lines, and by rotating the image by θ, it is assured that the images will be horizontally orientated. This happens because the horizontal lines will be longer than the vertical ones thus producing a greater value for the quantized parameter r. Once the horizontal orientation of the image if guaranteed, it is necessary to determine the relative position of the white color patch. By using a very high threshold value in the binarization process of the image, it is possible to isolate the white patch and determine if it is necessary to horizontally flip the image to make sure that the white patch appears on the bottom left corner.

Isolating all the color patches is no easy task. Unlike before there is no clear distinction between the background and foreground, since the difference in intensity between some of the darker patches and the black background of the chart is so small and these patches can easily be misclassified as background. Three different alternative approaches are proposed: edge detection-based segmentation (Canny method), the use of an adaptive threshold (instead of a global one), and segmentation based on the intensity profile of the color checker image. After each of this methods, some noise removal operations were performed, such as: image opening/closing and image ero-

sion/dilation.

Edge-based Segmentation (Canny Method)

The Canny edge detector, was developed by John F. Canny in 1986, is an edge detection operator that uses a multi-stage algorithm to detect object edges in digital images [64]. The Canny edge detection algorithm can be broken down to 5 different steps [65]:

1. Apply Gaussian filter to smooth the image to remove the noise;
2. Find the intensity gradients of the image;
3. Apply non-maximum suppression to get rid of spurious response to edge detection;
4. Apply double threshold to determine potential edges;
5. Finalize the detection of edges by suppressing all the other edges that are weak and not connected to strong edges.

Applying this algorithm to a grayscale image of the color checker should result in a binary image in which the white pixels correspond to the edges of the patches. Using a morphological operation known as image filling, it is possible to replace the black pixels inside the edges detected by the Canny operator with white ones resulting in a binary image that correctly displays the foreground objects. At this point, a mask for each of these objects is extracted and used to isolated the individual color patches in the original RGB image. The pipeline for this algorithm is show in Figure 3.8.

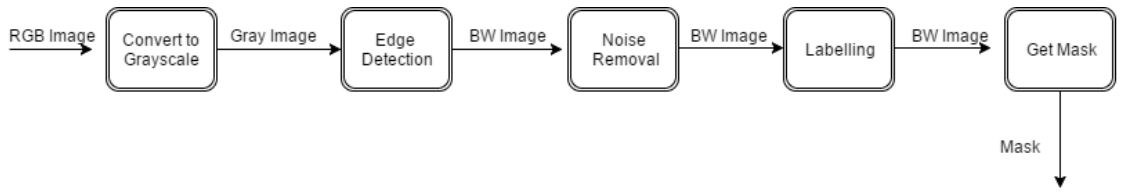


Figure 3.8: Edge Based (Canny) Segmentation Pipeline

Adaptive Thresholding

While using a global thresholding method is unadvised because the color patches present a wide range of intensities that makes it difficult to find an optimal threshold value, there is an alternative that can be explored – adaptive thresholding. Whereas the conventional thresholding operator uses a global threshold for all pixels, adaptive thresholding changes the threshold dynamically over the image. Local adaptive thresholding, selects an individual threshold value for each pixel based on the range of intensity values of its local neighborhood. This allows for thresholding of an image producing a binary image that correctly separates background and foreground objects.The pipeline for this algorithm is show in Figure 3.9.

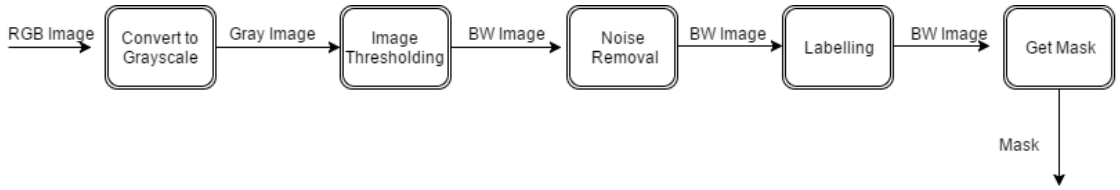


Figure 3.9: Adaptive Threshold 1 Pipeline

Segmentation based on the intensity profile

This last approach uses prior knowledge about the objects that need to be segmented. Since the total number of color patches and their relative position to each other in the color chart is known, the algorithm described below was proposed:

1. Apply Gaussian filter to smooth the image to remove the noise;
2. Draw 4 horizontal lines and 6 verticals lines over all the color patches;
3. Get the intensity profile of the image across the previously drawn lines;
4. Find the local minima and maxima in those intensity profiles.

The location of the minima and maxima of the intensity profiles correspond to the locations where there is a sudden change in the image intensity and therefore are classified as the edges of the color patches. Now that the position of the edges is identified, a mask to isolate each individual patch can be extracted.

This algorithm is a lot similar to the Canny method, except it has the upside of being optimized for the segmentation problem in question. The pipeline for this algorithm is show in Figure 3.10

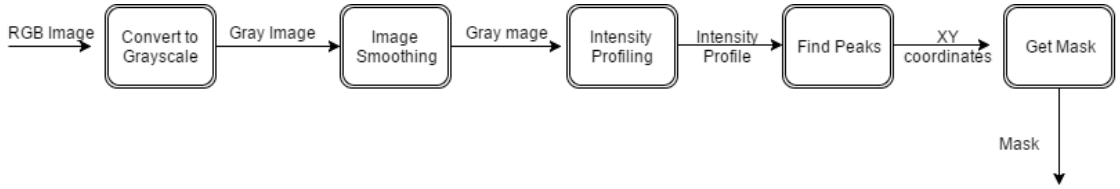


Figure 3.10: Intensity Profile based Segmentation Pipeline

3.6 Color Constancy

Color constancy is a property of the vision system allows humans to perceive colors of objects as being essentially independent of the light which illuminates them [66]. Similarly, computer vision color constancy aims to estimate actual color in an acquired scene image disregarding the illuminant [67].

Illumination-based variance in the colorimetric analysis of the images from the database is inevitable. This variance can be very problematic when trying to model certain behaviors, such as

the color intensity response of the chemical reaction studied in this paper. Not only can the ambient lighting change in an instant, the way these images are acquired can be camera specific. This means that each of the cameras used can produce its own interpretation of the scene being captured. These problems were expected, and so, to mitigate their overall influence on the results a color rendition chart was placed on the working bench during the database acquisition process. With the help of this chart, the images in the database can be color corrected by comparing the color information extracted from the color patches in each of the images with their reference values. This will make it so that the same object present in different images captured by different cameras will produce similar colorimetric readings, hopefully guaranteeing color constancy through the images of the database.

3.6.1 White Balance

The first approach used in this work is probably the simplest. This approach attempts to secure color constancy through a technique known as white balance (WB). A WB algorithm aims to render white objects as white under any light source [68]. The WB algorithm here used was proposed by Garud et al. [69] and uses a combination of a static method of illuminant estimation called White Patch Retinex (WPR) method [70] [71] and linear transformation model of WB correction of the images [72]. The main advantage of this method is that it does not require training data, is simple to implement and tune and has a very low computational complexity and Garud et al. [69] concluded that this algorithm was a promising primary method for WB correction. Its main disadvantage is that is focus primarily on correction in the white color spectrum and is not as effective for correcting other colors.

3.6.2 Color Correction

Unlike the method mentioned above, the next methods for color correction studied in this thesis require the use of the color rendition chart and are based in a process known as RGB Matching. RGB Matching changes the color intensities reported by the smartphone cameras so that they match those snapped at some initial time-point (same system), or to another system (the “gold standard” system).

This approach here presented uses the RGB values extracted from all the 24 color patches of the color rendition chart and compares them to the reference values provided by the chart’s manufacturer that were precisely measured with a spectrophotometer. By performing a least-square regression between these two datasets we can estimate the parameters that will more likely approximate the measured RGB values with the reference values.

Another approach considered in this work is also based in color matching, but unlike the RGB-to-RGB transform described before, this method uses a RGB-to-XYZ transformation. Much like before, the XYZ tri-stimulus ground truth values used are specific to that color chart used and published by its manufacturer. This transformation is then applied to a white-balanced RGB image to obtain the color-corrected XYZ image, which is the linear, device-independent version of the

raw camera output [73]. It then estimates the 3 3 transformation matrix T by computing it as a least-square regression problem [73] (3.2):

$$T = V^{XYZ}_{GroundTruth}[V^{RGB}_{Linear}]^{+} \tag{3.2}$$

Where the superscript + denotes the Moore–Penrose pseudoinverse of the matrix. Here $V^{XYZ}_{GroundTruth}$ and V^{RGB}_{Linear} are 3 N matrices where N is the number of patches in the calibration target (24). This transformation T is then applied to a white-balanced RGB image (see 3.3) to obtain the color-corrected XYZ image, which is the linear, device-independent version of the raw camera output [73].

$$I^{XYZ}_{Corrected} = T \cdot I^{RGB}_{Linear} \tag{3.3}$$

3.7 Color/Concentration Model

3.7.1 Color Space Comparison

Establishing a relationship between the intensity of the colored product formed by the chemical reaction described in section 3.1 and the SA concentration used in that reaction is one of the goals of this dissertation. This can be done by correlating the color values extracted from the sample versus the SA concentration for each of the parameters that describe the following color spaces: RGB, HSV, XYZ, CIELAB. This is done to determine the best individual parameters to describe the changes in the SA concentration. The ideal variable to map this should produce a curve as close to linear as possible and, most importantly, it should be monotonic, i.e. it should be always decreasing or increasing.

Once the parameters that best fit the criteria mentioned above, a mathematical function that best describes the relationship between color and concentration is found via linear regression.

3.7.2 Color Stability

Once the function that best describes the relationship between color and concentration has been found it is important to know when it can be used correctly. To understand this, a simple experiment was proposed that consisted in studying the evolution of the colorimetric readings over time for the same sample. Starting from the moment the reagent was applied to the disk, an image was captured every minute for 20 minutes. This process was repeated multiple times for different concentration values (see Table 3.3). With this experiment, we can check how long it takes for the colored product of the reaction to stabilize and if there is any degradation of color over time. By doing so we can determine the timeframe for which the developed model is valid.

Table 3.3: SMZ concentrations used in color stability test

SMZ Standard Concentration [μg L^{-1}]	**SMZ standard volume [mL]**	**SMZ concentration [μg L^{-1}]**	**Sample volume [mL]**
1000	0.050	5.0	10
1000	0.100	10.0	10
1000	0.250	25.0	10
1000	0.500	50.0	10

3.8 Evaluation Metrics

3.8.1 Image Segmentation

The quality of the segmentation algorithms used in this work will be evaluated by comparing the results of these algorithms with ground truth annotations. The ground truth used for this comparison is the result of manually segmenting the ROI from the original images. By comparing these two results we can calculate the precision, recall and f-score values of the segmentation methods. Precision is the probability that a (randomly selected) element is relevant, while recall is the probability that a (randomly selected) relevant element is retrieved. Precision and recall are defined by Equations (3.4) and (3.5) respectively.

$$Precision = \frac{Truepositives}{TruePositives + FalsePositives} \tag{3.4}$$

$$Recall = \frac{Truepositives}{TruePositives + FalseNegatives} \tag{3.5}$$

Both precision and recall are based on an understanding and measure of relevance [74]. In statistical analysis of binary classification, the F-score is a measure of a test's accuracy [74]. It considers both the precision and the recall of the test to compute the score. This score is the weighted average (Eqution 3.6) of the precision and recall, that has its best value at 1 and worst at 0. The traditional F-score is the harmonic mean of precision and recall — multiplying the constant of 2 scales the score to 1 when both recall and precision are 1 [74]:

$$F - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{3.6}$$

3.8.2 Color Constancy

The quality of the methods used to achieve color constancy can be measured by comparing the color measurements from the color patches of the color rendition chart in the color corrected images and the published references values from each patch. The best method should be the one that presents the smaller the difference between these two numbers.

3.8.3 Color/Concentration Model

For the evaluation of the quality of the curve fitting, several metrics were used:

1. **The sum of squares due to error (SSE):** This statistic measures how successful the fit is in explaining the variation of the data. R-square is defined as the ratio of the sum of squares of the regression (SSR) and the total sum of squares (SST). R-square can take on any value between 0 and 1, with a value closer to 1 indicating that a greater proportion of variance is accounted for by the model [75].

2. **R-square:** This statistic measures how successful the fit is in explaining the variation of the data. R-square is defined as the ratio of the sum of squares of the regression (SSR) and the total sum of squares (SST). R-square can take on any value between 0 and 1, with a value closer to 1 indicating that a greater proportion of variance is accounted for by the model [75].

3. **Root mean squared error (RMSE):** This statistic is also known as the fit standard error and the standard error of the regression. It is an estimate of the standard deviation of the random component in the data [75]. Just as with SSE, an MSE value closer to 0 indicates a fit that is more useful for prediction.

3.9 Summary

In this chapter, the steps necessary for the colorimetric analysis of digital images for determination of SAs were presented. In the first section, the chemical reaction between p-DMACA and SAs was described. Following that, the extraction procedure for obtaining the sample was explained and the necessary materials were listed. Additionally, the image acquisition protocol used to create the image database was described.

Also in this chapter, a comprehensive description of all the stages of the image processing portion of the work was given. It starts off by listing all the elements that needed to be segmented: the sample and the color checker (along with each individual color patch). One method for segmenting the sample and three different methods for segmenting the color checker were proposed. This chapter also provides information on the importance of color constancy throughout the images and presents three alternatives for image correction.

The next section of this chapter gives information on how the function that relates color and concentration is to be created and proposes an experiment that analysis the behavior of the color of the sample over time in order to validate its limitations. Finally, in the last section of this chapter, this work's evaluation metrics were enumerated. These metrics will later be used to judge and compare the different methods used for processing the images.

Chapter 4

Results

In this chapter, all the results obtained throughout the duration of this work are presented. It starts by describing the contents of the image database, and follows that by comparing the different methods proposed for image segmentation and color correction. Following that, a color space comparison is given. This comparison is used to determine which space is best suited to describe the relationship between SA concentration and color measurements. Once that color space is determined, the curve fitting process to model this relationship is explained. Using the results of that curve fitting process, the validation of this model is presented. Finally, to close out this chapter, the main results and conclusions from each of these sections are highlighted.

4.1 Database Acquisition

Acquiring all the images to create an image database was a long process that involved a lot of repeated motions. However, by the end of this process, an extensive image database had been compiled. This database consisted in a total of 408 images captured with 3 different smartphone cameras and occupies 4.5 GB of memory storage. In these images, 17 different SMZ concentrations are represented (see Table 3.2). There are 24 images for each SMZ concentration (3 for each camera). Please see Figure 4.1 for a visual representation of the color variation between different concentrations.

Figure 4.1: Color variation between different concentrations

4.2 Image Processing and Analysis

In this section, the results from segmenting the various elements from the original images are presented. It starts by displaying the results of the sample segmentation algorithm and follows that by comparing the results of the different methods for segmentation the color checker.

4.2.1 Sample Segmentation

The goal of this step in the image segmentation process is to isolate the sample in each of the images that make up the image database. A successful segmentation should result in a RGB image that only displays the colored product created by the chemical reaction.

Displayed in Figure 4.2 are the results for several intermediate stages in the sample segmentation algorithm.

(a) Binary Image (b) Sample Cropping (c) Define Contour

Figure 4.2: Sample Segmentation Algorithm

This algorithm was tested in 30 randomly selected images from the database, including 10 images per smartphone camera. In order to evaluate the quality of this procedure, the results of

the automatic segmentation were compared to the results of the manual segmentation of these images (ground truth). The average precision score, recall score and f-score for the 30 images are displayed in Table 4.1.

Table 4.1: Results obtained from the evaluation of algorithm used for the sample segmentation

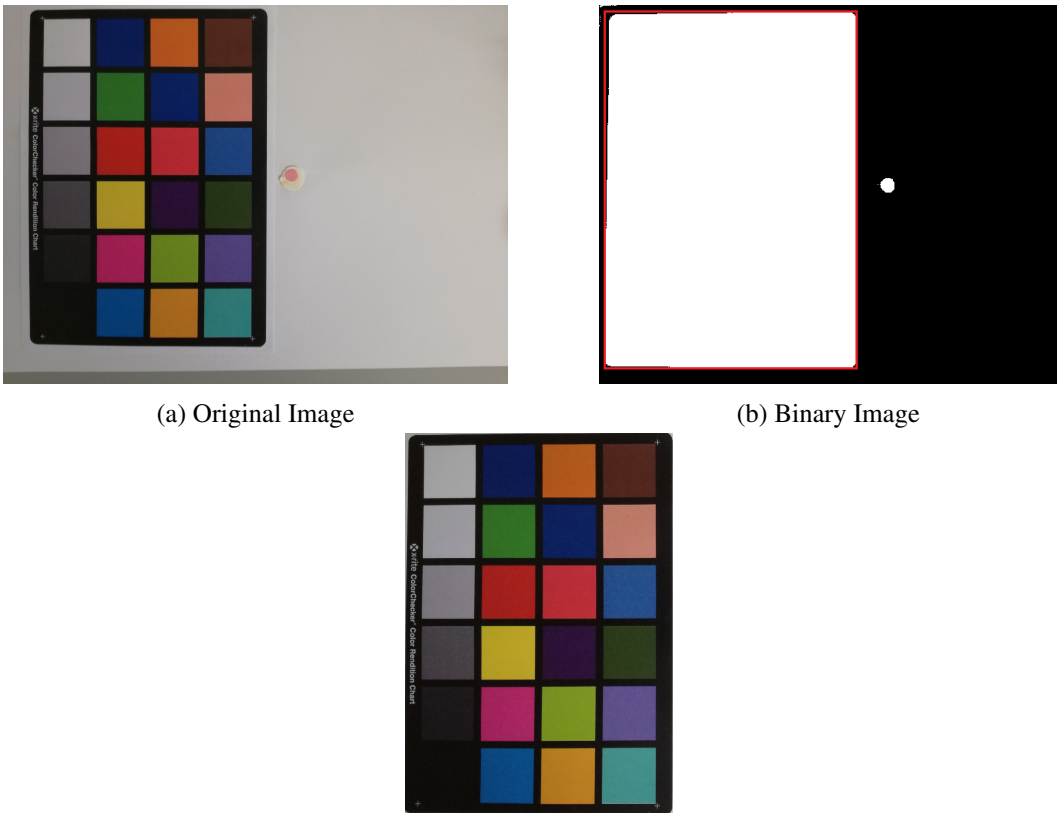
Precision			**Recall**			**F-score**
Mean(%)	**Std**	**Median(%)**	**Mean(%)**	**Std**	**Median(%)**	
93.05	1.21	92.94	90.65	0.92	90.64	0.93

Overall, this algorithm presented highly satisfactory results and these scores mean that the algorithm can correctly isolate the sample from the original image. With an average precision score of 93.05%, it is safe to assume that this segmentation algorithm does a good job in selecting the important information from the images. An average recall score of 90.65, while satisfactory, means that there is some slight under-segmentation of the sample, but not enough for it to be problematic.

4.2.2 Color Checker Segmentation

The color checker segmentation algorithm can be divided into three major stages: isolating the color checker, rotating the color checker and isolating the 24 color patches.

Isolating the color checker from the original image proved to be quite easy thanks to the sharp contrast between the black background of the color checker and the white of the blank sheet of paper. This contrast between the two different areas made it possible to use a high threshold value for the binarization operation, that resulted in a very clean binary image with very little or close to no noise that correctly isolated the color checker (Figure 4.3).



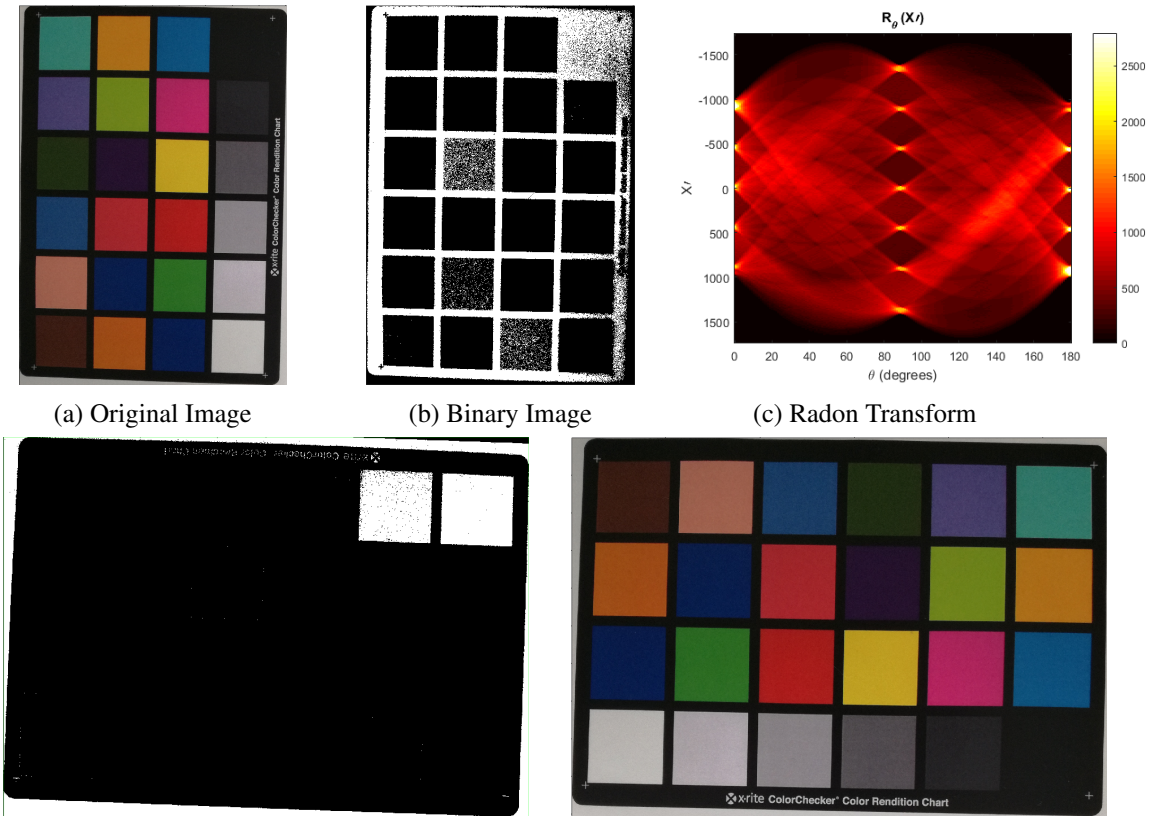
(a) Original Image

(b) Binary Image

(c) Color Checker Cropping

Figure 4.3: Color Checker Segmentation Algorithm

Once the color checker has been isolated, this new image has to be rotated so that it has a horizontal orientation and the white color patch appears in the bottom-left corner. Guaranteeing that all the color checker images are correctly positioned is important, since this makes identifying the individual color patches, that will later be isolated that much easier, because the same color patch will have the same relative position throughout all the images. The results for the intermediate stages of this rotation algorithm are displayed in Figure 4.4.



(a) Original Image (b) Binary Image (c) Radon Transform

(d) Find brightest patches (e) Final Image

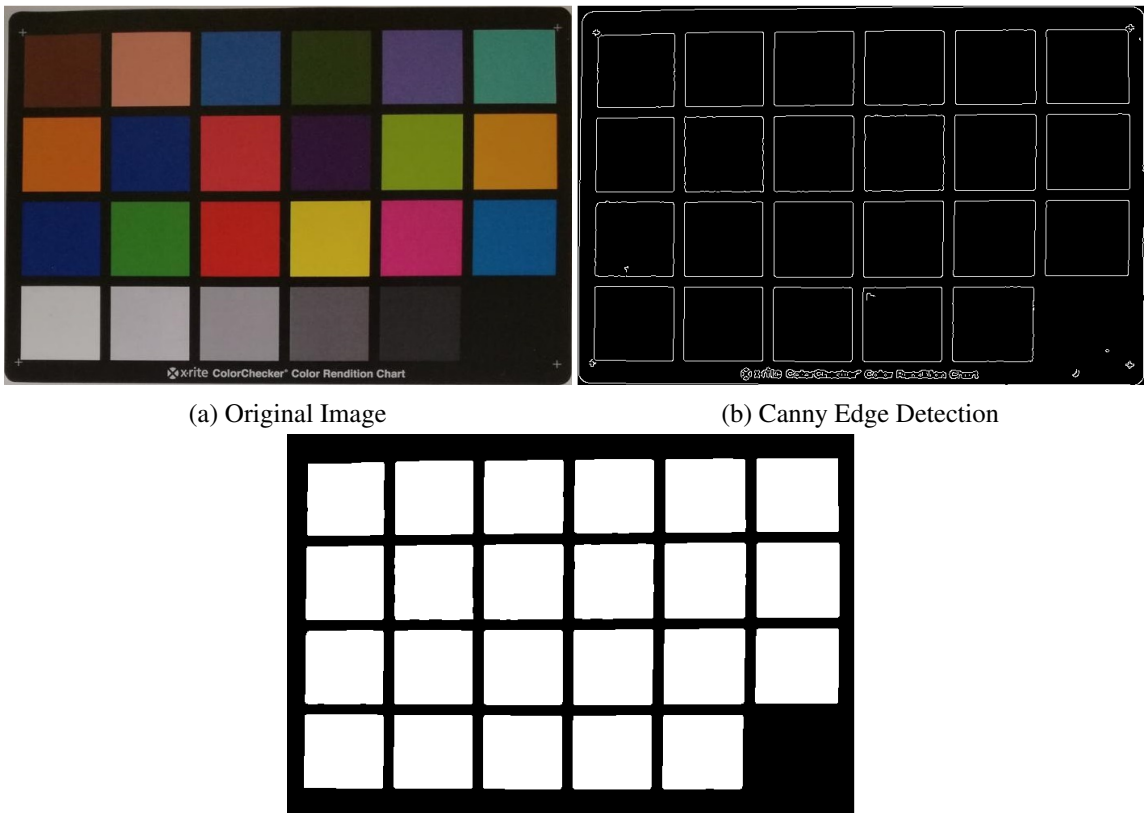
Figure 4.4: Image Rotation Algorithm

As seen on Figure 4.4 the rotation algorithm starts by converting the original RGB image of the color checker into a binary image and then performing a Radon Transform in order to determine the angle, θ, that is going to be used to rotate the RGB image. After rotating the RGB image by theta, the algorithm isolated the brightest patch of the color checker. It does so that converting the image into a binary using an high threshold value. If the brightest patches are on the top-right quadrant, the original RGB image is rotated by an additional 180^{o}.

Once this process is complete, all that is left to do is to segment each individual color patch. As stated in chapter 3, three different methods to achieve this were proposed. Now, the results for each of these methods will be presented.

Method 1: Edge-based Segmentation (Canny Method)

This first method uses a canny edge detection algorithm to find the edges of the 24 color patches present in the color checker. It does so by finding the image coordinates where there are sudden differences in intensity. The results for the intermediate stages of this segmentation algorithm are displayed in Figure 4.5.



(a) Original Image (b) Canny Edge Detection

(c) Color Patch Detection

Figure 4.5: Edge-based Segmentation (Canny Method)

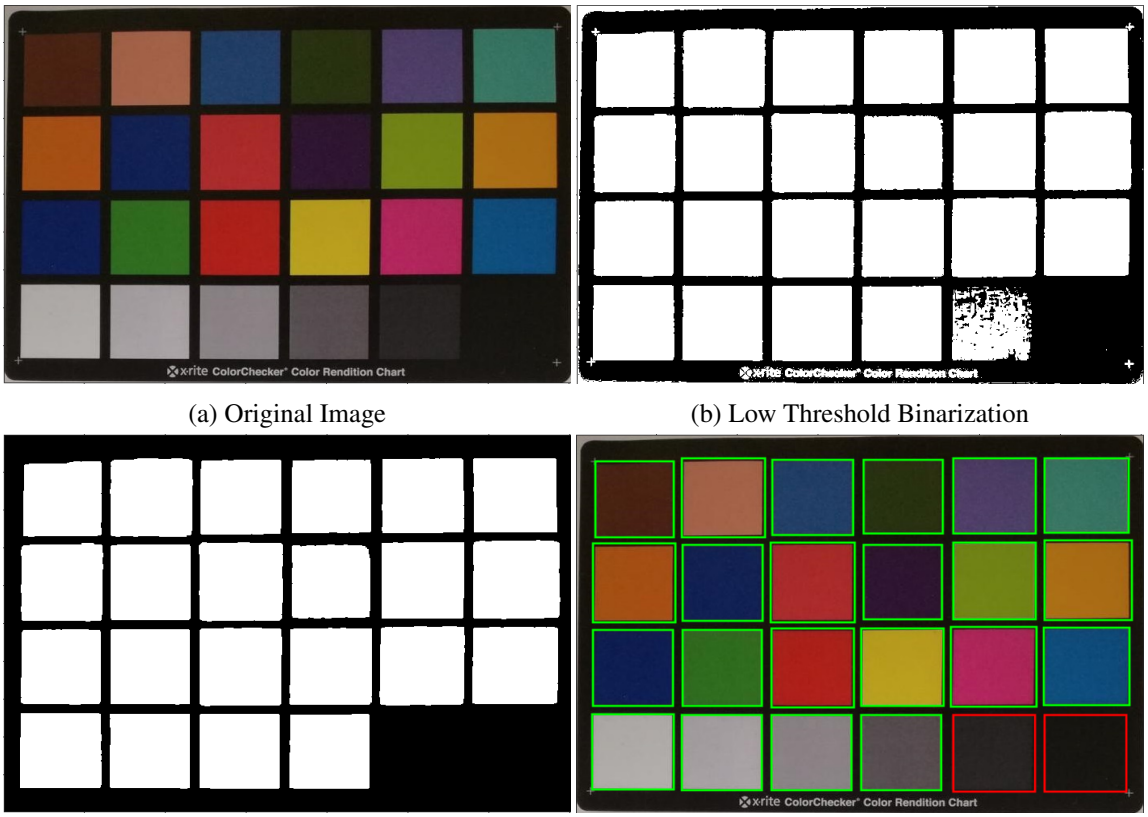
As seen on Figure 4.5, this algorithm can only isolate 23 out of the 24 color patches. This was an expected outcome, since the difference in intensities between the color checker background and the darkest patch is almost non-existent. Therefore, the position of the last patch was estimated by combining prior knowledge of its relative position in the chart and information gathered by the correctly isolated patches. With this information it is possible to predict where the patch should be.

Method 2: Adaptive Thresholding

This second method uses a technique known as local adaptive thresholding. The individual threshold value for each pixel is calculated based on the range of intensity values of its local neighborhood. Using this technique, the original RGB image of the color checker is convert into a binary image.

After analyzing some preliminary results, it was determined that the sensitivity of this local operator required a lot of fine tuning. Using a higher sensitivity meant that this method could isolate a considerable number of patches, but did so by generating a lot of noise in the binary image. While using a lower sensitivity generated very little, or close to, no noise, but failed to detect a lot of the color patch. Therefore, it was decided to divide this method into two separate steps. One would use a highly sensitive local operator to calculate the individual threshold values

for each pixel, which meant that after the binarization of the image it needed a lot of noise removal (Figure 4.6). The other one would use a lower sensitivity operator meaning that there would be a need to estimate the position of the undetected patches, and to do so there had to be at least one detected patch for each row and column of the color checker chart (Figure 4.7).



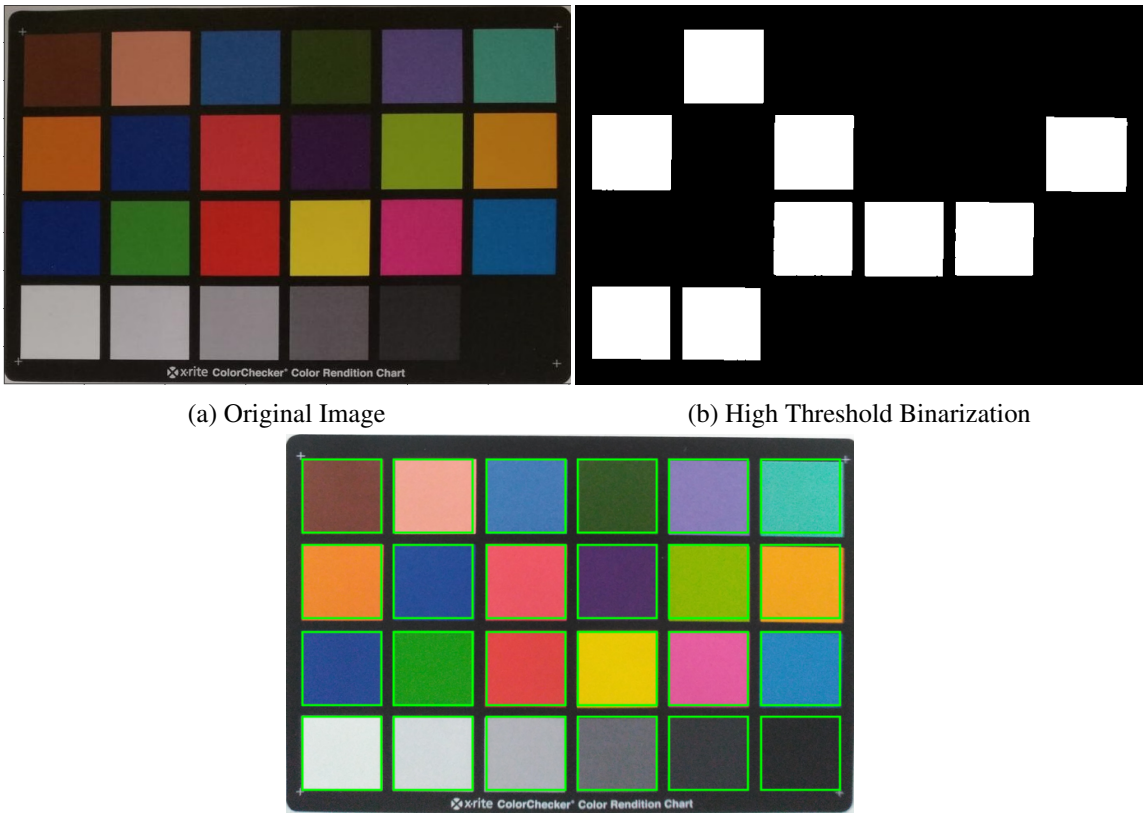
(a) Original Image

(b) Low Threshold Binarization

(c) Noise Removal

(d) Color Patches

Figure 4.6: Segmentation with a higher sensitivity adaptive threshold value



(a) Original Image (b) High Threshold Binarization

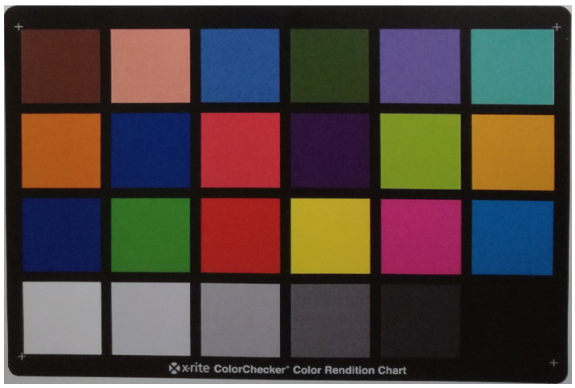
(c) Color Patches

Figure 4.7: Segmentation with a lower sensitivity adaptive threshold value

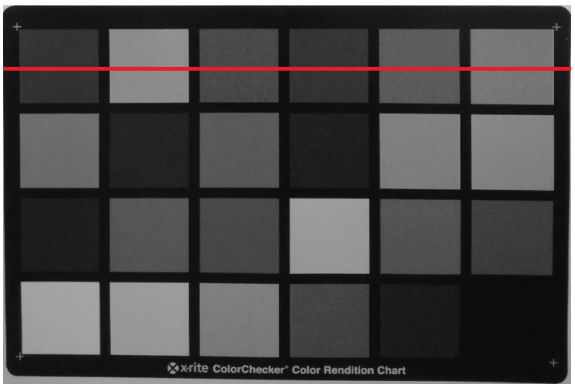
As shown in Figure 4.6, even using a highly sensitive local operator to perform adaptive thresholding, the darkest patches remained undetected, as just as before, their relative position had to be estimated. Figure 4.7 demonstrates that by using a lower sensitivity local operator to perform adaptive thresholding, this method can detect at least one color patch in each row and column and guarantees that it is possible to then estimate the position of all the remaining patches. Because there is at least one patch in every row and column, we have the information necessary to determine the position of the missing patches.

Method 3: Segmentation based on the intensity profile

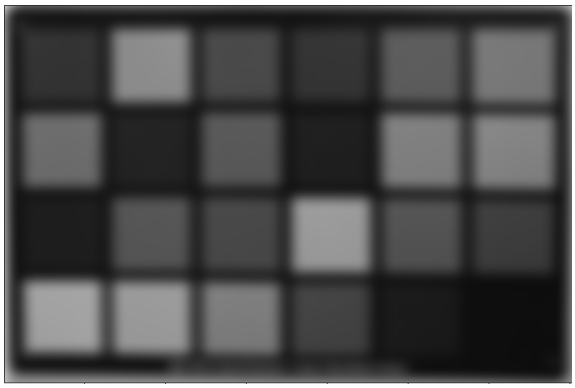
The final method proposed in this thesis to isolate the individual color patches uses the image intensity profile to detect sudden changes in intensity. Much like with the Canny method, these sudden intensity changes are classified as edges. The results for the intermediate stages of this segmentation algorithm are displayed in Figure 4.8.



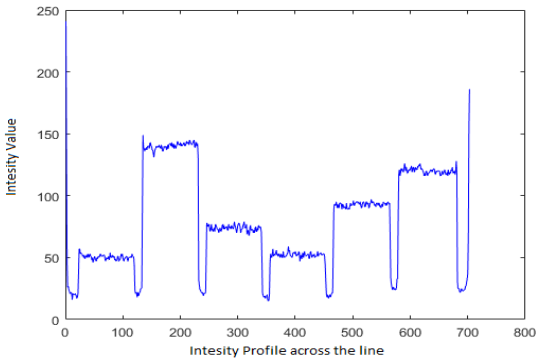
(a) Original Image



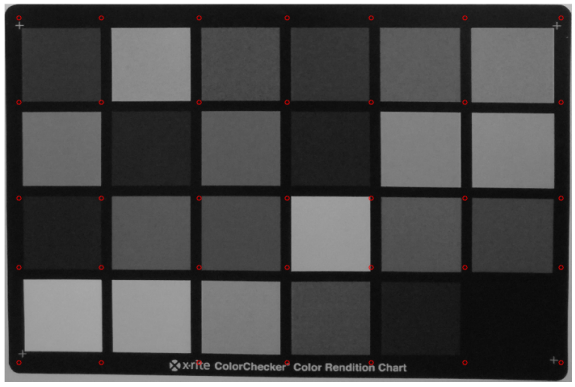
(b) Gray Image and Profile Line



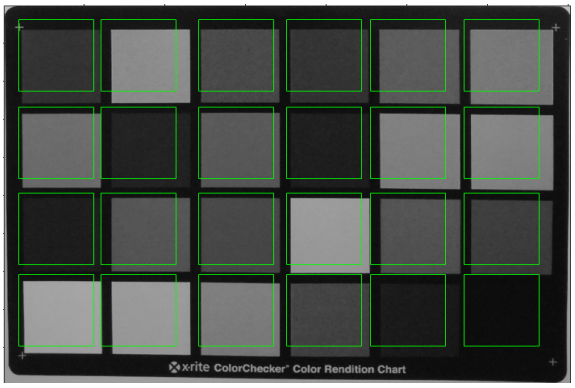
(c) Gray Image after Gaussian Filter



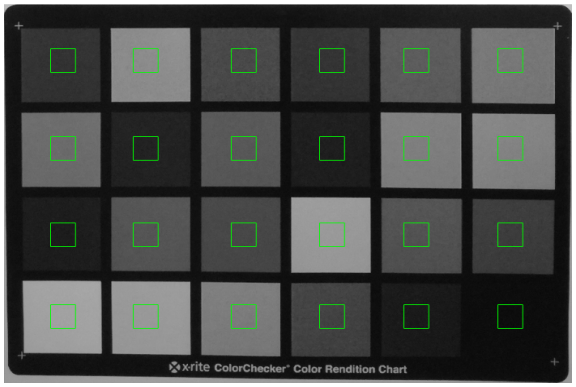
(d) Intensity Profile



(e) Position of the Intesity Profile Peaks



(f) Color Patch Detection



(g) Adjusted Color Patch Detection

Figure 4.8: Intesity Profile Based Segmentation

As shown in Figure 4.8, this method can easily detect the correct number of edges, but because of the Gaussian filter applied to smooth the image, it overshoots the position of the edges. Reducing the amount of smoothing performed by the Gaussian filter, did not help with this problem because it meant that some edges were misidentified. To solve this problem, the area to be segmented was shrunk (Figure 4.8 (g)). Although this causes the color patches to be under-segmented, this algorithm shows that it can correctly isolate at least a portion of each individual color patch.

Method Comparison

All the color checker segmentation methods described above were tested with the same 33 randomly selected images from the database, including 11 images from each of the different smartphone cameras used. To evaluate the quality of the methods, the results of the automatic segmentation were compared to the results of the manual segmentation of these images (ground truth). Shown in Table 4.2, are average results for each one of the methods.

Table 4.2: Results obtained from the evaluation of the different methods used to segment the color patches from the color checker.

	Precision			**Recall**		
Method	**Mean(%)**	**Std**	**Median(%)**	**Mean(%)**	**Std**	**Median(%)**
Edge-Based (Canny)	98.55	0.54	98.49	97.33	0.73	97.44
Higher Sensitivity Thresh.	93.23	1.89	92.74	97.35	5.16	99.15
Lower Sensitivity Thresh.	99.44	0.44	99.59	88.54	2.20	89.97
Intensity Profiling	80.66	3.59	80.61	79.47	2.90	79.86

Method	**F-score**
Edge-Based (Canny)	0.98
Higher Sensitivity Thresh.	0.95
Lower Sensitivity Thresh.	0.94
Intensity Profiling	0.80

As illustrated by Table 4.2, the method that performed the best was the edge-base segmentation method (Canny Method). An F-score of 0,98 is satisfactory, and it means that this method almost perfectly isolates each color patch. Besides having the highest F-score, this method also has one of the lowest deviations between images, meaning that it is very robust.

The adaptive thresholding methods also proved to be very good candidates for color checker segmentation. Due to the nature of the approach, the thresholding method that uses a higher sensitivity parameter has the highest deviation between images, meaning it is not as robust as the edge-based segmentation method.

Although, the adaptive thresholding method that uses a lower sensitivity parameter has the second lowest F-score, it also has the highest precision values, averaging an almost perfect precision score, which means that this method does not over-segment its elements and almost of the pixel it isolated contain relevant information. The problem with this method is its recall score, meaning that there is in fact some relevant pixels that the method fails to segment. The biggest

problem of this segmentation method is that it is not as robust as the previous two, since it estimates the position of the majority of the color patches, if there is some sort of the bend or fold in the color checker chart, this method is not flexible enough to adapt, causing it to perform a bad segmentation.

The low F-score for the intensity profile method was to be expected since this method purposefully under-segments all the color patches. With that being said, the precision and recall scores might not be the best metrics to judge the quality of this algorithm.

Based on these results, going forward, the algorithm that will be used to segment the color patch is the edge-based segmentation (Canny method) algorithm.

4.3 Color Constancy

This section of the dissertation presents the results generated by the algorithms proposed to guarantee color constancy throughout the images. It begins by evaluating the results of the white balance algorithm and then it evaluates the results of the two methods used to color correct the images. To judge the performance of the various color constancy algorithms, the color values for each patch of the color rendition chart after the transformations were compared to their reference values. Finally, a comparison between the algorithms is given and the best alternative to achieve color constancy is determined.

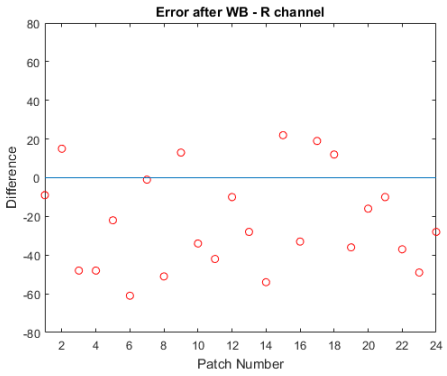
4.3.1 White Balance

Figure 4.9 displays the color checker before and after the application of the WB algorithm.

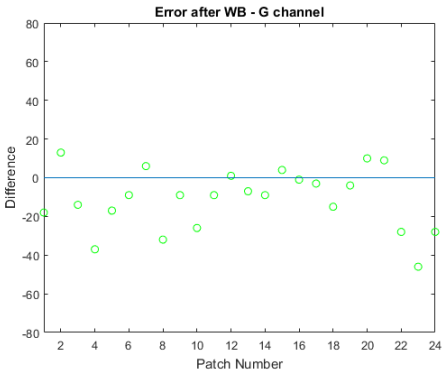
(a) Original Image (b) Corrected Image

Figure 4.9: White Balance Results: Before and After

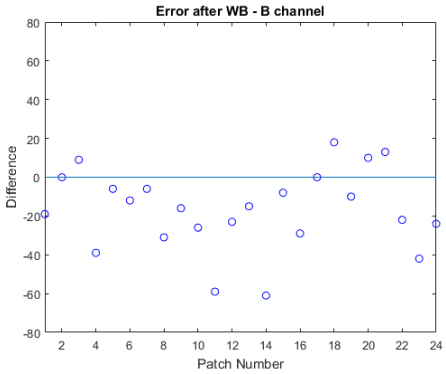
Presented in Figure 4.10 is a visual representation of the results obtained when using the white balance algorithm. These results were generated using 24 randomly selected images (8 images from each camera). They represent the average difference between the median value of the RGB components measured in each color patch and the color chart reference value. In Figure 4.11 this difference is given as a percentage relative to range of each RGB component.



(a) Red Channel

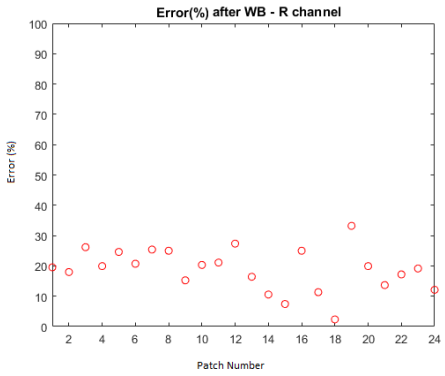


(b) Green Channel

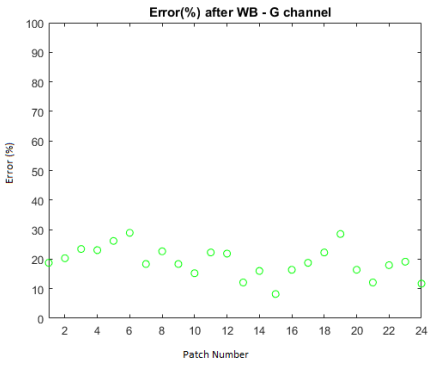


(c) Blue Channel

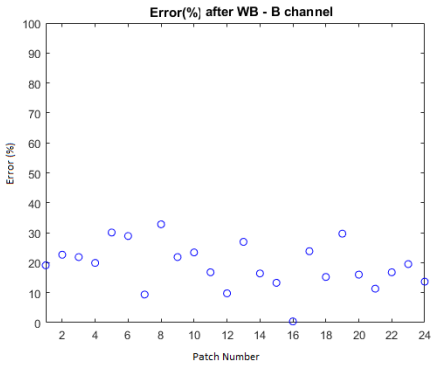
Figure 4.10: White Balance Results - Absolute Error



(a) Red Channel



(b) Green Channel



(c) Blue Channel

Figure 4.11: White Balance Results - Relative Error

Table 4.3 shows the total average relative error for each RGB component.

Table 4.3: Results from the White Balance algorithm

Red		**Green**		**Blue**	
Mean (%)	**Std**	**Mean (%)**	**Std**	**Mean (%)**	**Std**
18.81	6.98	19.12	5.2	19.16	7.65

As evidenced by Table 4.3, the performance of this color constancy algorithm was underwhelming. The difference between the RGB values measured in the corrected image and the color rendition chart reference values is too high. The focus of this white balance method is to remove unrealistic color casts, so that objects which appear white in person are rendered white in captured images. Unlike the other color correction algorithms described in this dissertation, it does no try to calibrate the image with a given calibration target. Therefore, these results were expected and are understandable.

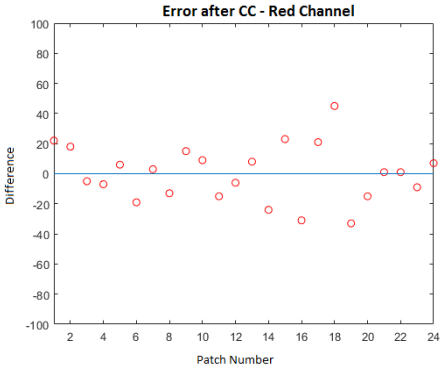
4.3.2 Color Correction (RGB-RGB)

Figure 4.12 displays the color checker before and after the application of the RGB-RGB color correction algorithm.

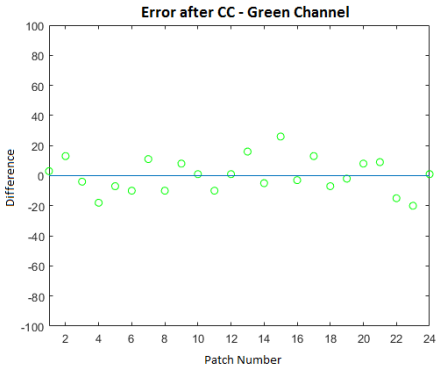
(a) Original Image (b) Corrected Image

Figure 4.12: White Balance Results: Before and After

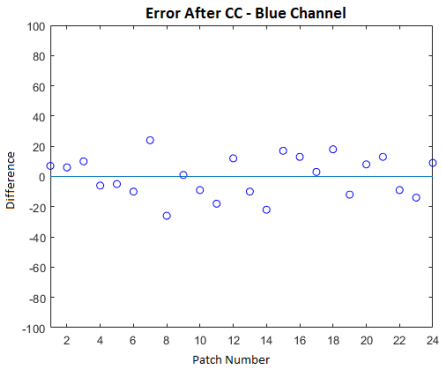
Presented in Figure 4.13 is a visual representation of the results obtained when using the RGB-RGB color correction algorithm. These results were also generated using 24 the same randomly selected images (8 images from each camera) as before. They represent the average difference between the median value of the RGB components measured in each color patch after correction and the color chart reference value. In Figure 4.14 this difference is given as a percentage relative to range of each RGB component.



(a) Red Channel

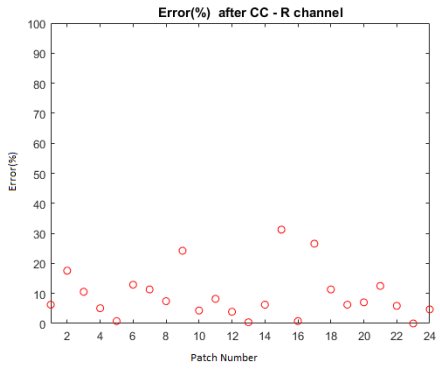


(b) Green Channel

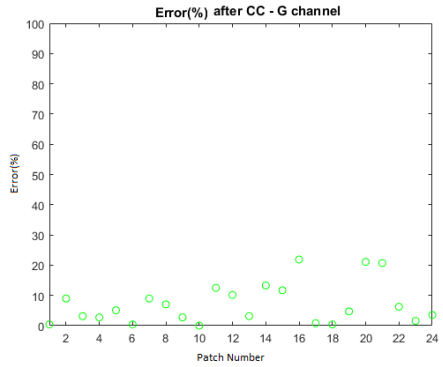


(c) Blue Channel

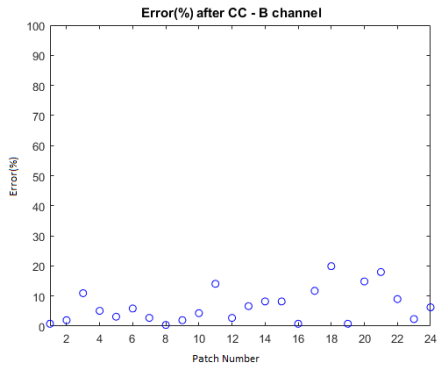
Figure 4.13: RGB-RGB Color Correction Results - Absolute Error



(a) Red Channel



(b) Green Channel



(c) Blue Channel

Figure 4.14: RGB-RGB Color Correction Results - Relative Error

Table 4.4 shows the total average relative error for each RGB component. By analyzing the

Table 4.4: Results from the RGB-RGB color correction algorithm

Red		**Green**		**Blue**	
Mean (%)	**Std**	**Mean (%)**	**Std**	**Mean (%)**	**Std**
9.39	8.22	7.13	6.77	6.69	5.66

results shown in Table 4.4, it is safe to assume that this color correction algorithm is a better alternative to white balance. There are a few color patches where the difference to the reference value are still too high, but overall, this method does a respectable job in matching the colors measured from the color chart to its reference values and is a suitable candidate to color correct the images in the database.

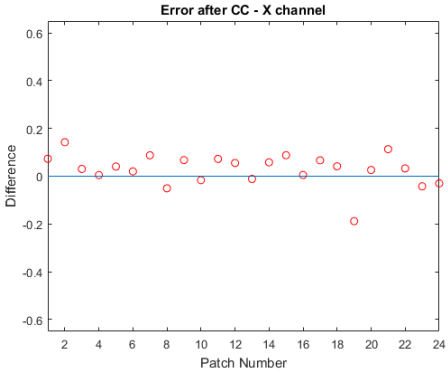
4.3.3 Color Correction (RGB-XYZ)

Figure 4.15 displays the color checker before and after the application of the RGB-XYZ color correction algorithm.

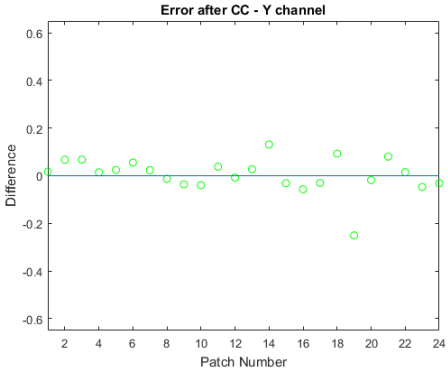
(a) Original Image (b) Corrected Image

Figure 4.15: White Balance Results: Before and After

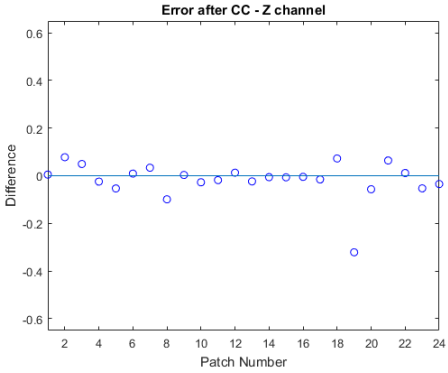
Presented in Figure 4.16 is a visual representation of the results obtained when using the RGB-XYZ color correction algorithm. Just as before, these results were generated using the same 24 randomly selected images (8 images from each camera. They represent the average difference between the median value of the RGB components measured in each color patch after correction and the color chart reference value. In Figure 4.17 this difference is given as a percentage relative to range of each XYZ component.



(a) X Channel

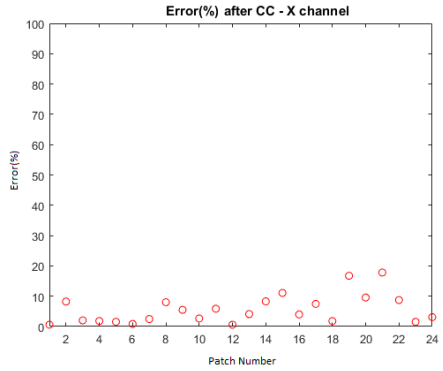


(b) Y Channel

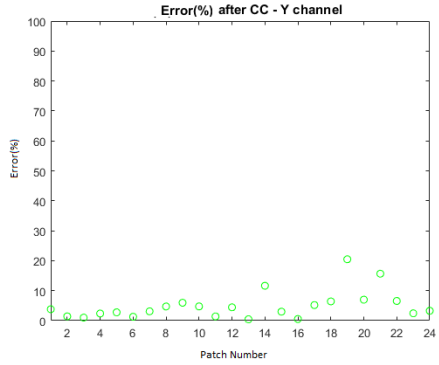


(c) Z Channel

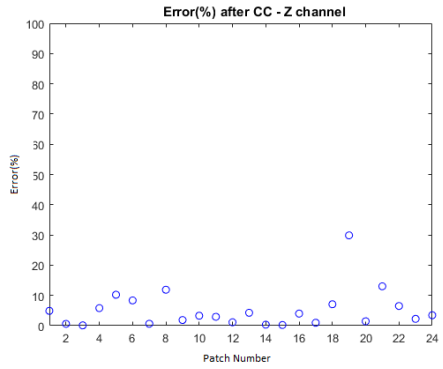
Figure 4.16: RGB-XYZ Color Correction Results - Absolute Error



(a) Red Channel



(b) Green Channel



(c) Blue Channel

Figure 4.17: RGB-XYZ Color Correction Results - Relative Error

Shown in Table 4.5 is the total average relative error for each XYZ component.

Table 4.5: Results from the RGB-XYZ color correction algorithm

X		**Y**		**Z**	
Mean (%)	**Std**	**Mean (%)**	**Std**	**Mean (%)**	**Std**
5.59	4.79	4.97	4.81	5.20	6.44

Table 4.5 proves that this color constancy algorithm is an extremely good option to color correct the images from the database. Compared to the RGB-RGB color correction algorithm, this approach does a much better job in matching the measurements from each individual color patch to their reference values. Going forward, this will be the method used to achieve color constancy among the database images.

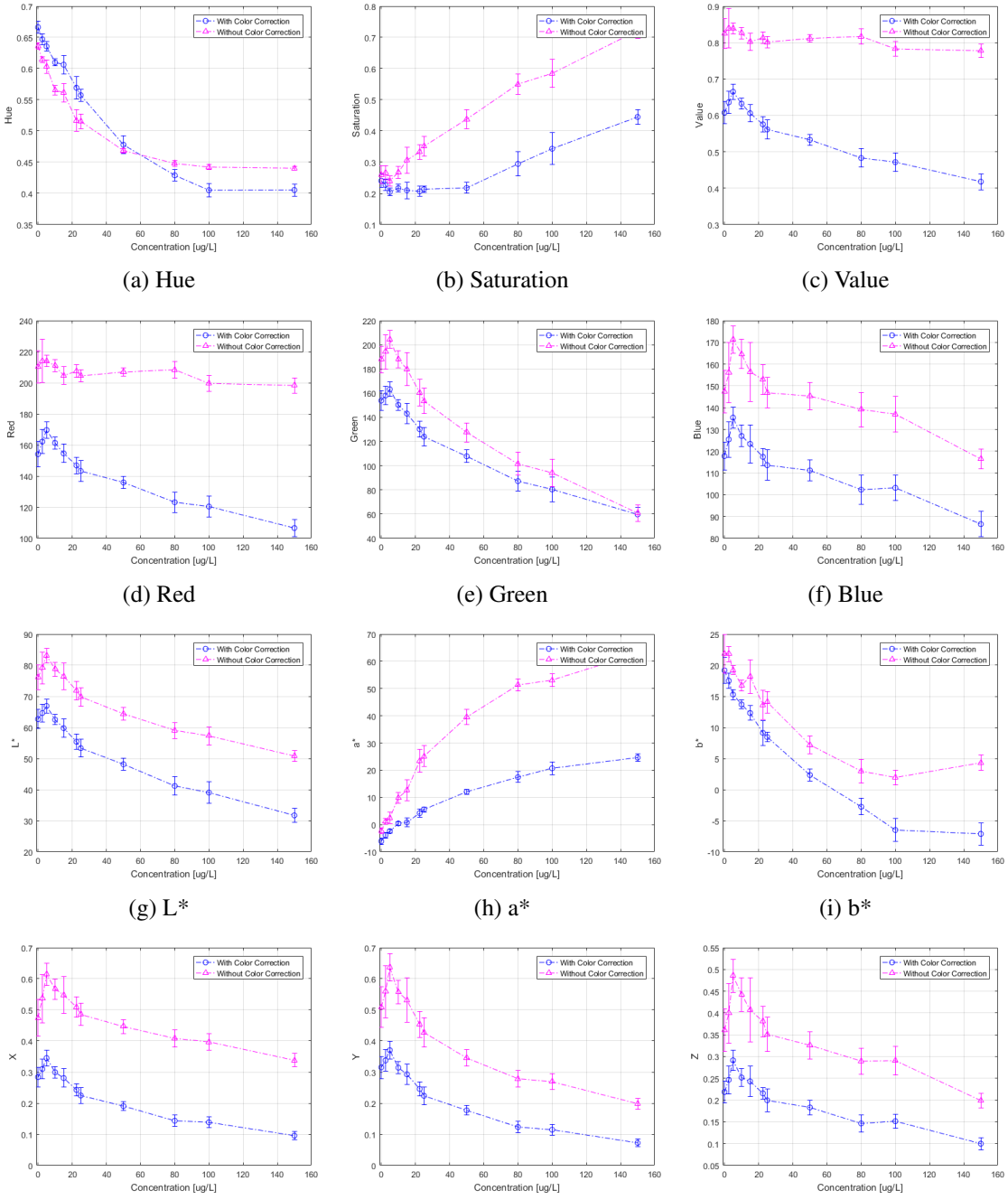
It is important to know, that when using color correction methods based on the color rendition chart reference values, that it is imperative to assure a homogeneous diffusion of the light source throughout the chart. Because the reference values of each color patch are the result of measurements done under uniform lighting, if the correct light diffusion cannot be assured, then the color correction method is invalid.

4.4 Color/Concentration Model

In this section, all the intermediary results used to create a model that describes the relationship between color and SA concentration are provided. These results had a considerable influence in the creation of the final model. After presenting those intermediary results, the function that best describes is this model is given.

4.4.1 Color Space Comparison

The first step taken to create a model that correlates the color information extracted from the samples and their SA concentration was to determine which color space was best suited to describe this relationship. The color spaces considered for this study were: RGB, HSV, CIELAB and CIEXYZ. The way each component reacts to an increase of SA concentration is shown in Figure 4.18. Also visible in Figure 4.18, is how the color correction algorithm influenced the color information extracted from the samples. The error bars present in the plots represent the standard deviation of the measurements.



(a) Hue (b) Saturation (c) Value

(d) Red (e) Green (f) Blue

(g) L* (h) a* (i) b*

(j) X (k) Y (l) Z

Figure 4.18: Color Space Comparison

For the different values of concentration used, the color parameter is the average value of the measurements from 6 different images (2 from each camera), with and without color correction, from the first set of image acquisitions (Training Database). After close examination of Figure 4.18, the color components chosen to model the relationship between color and SA concentration were the Hue (HSV) and a* (CIELAB), since they present a monotonous behavior, i.e. they are always increasing or decreasing, which benefits the curve fitting process.

4.4.2 Curve Fitting

Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points [76]. This process was applied to the datasets of the color components previously selected (hue and a*) using the RMATLAB2017a Curve Fitting software and the results are displayed in Figure 4.19 and 4.20.

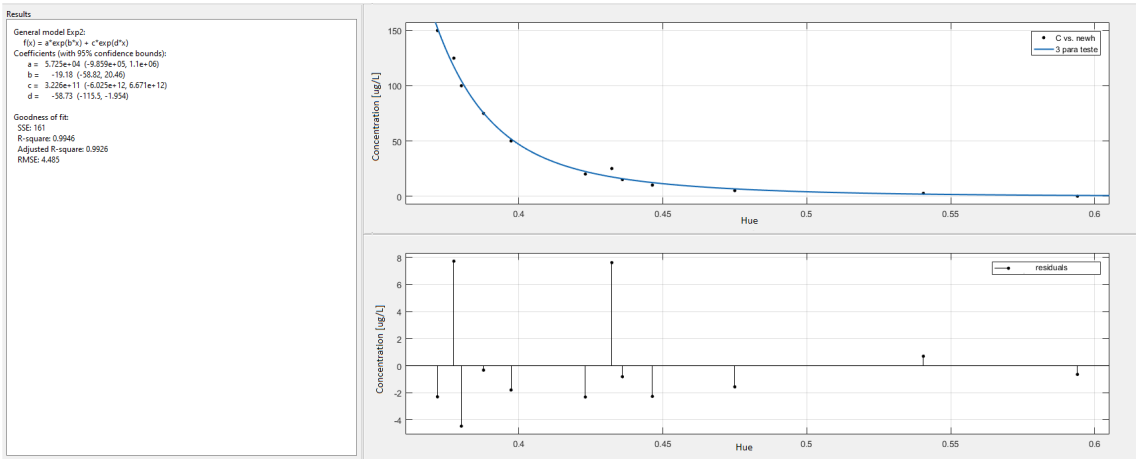


Figure 4.19: Curve Fitting using the Hue component

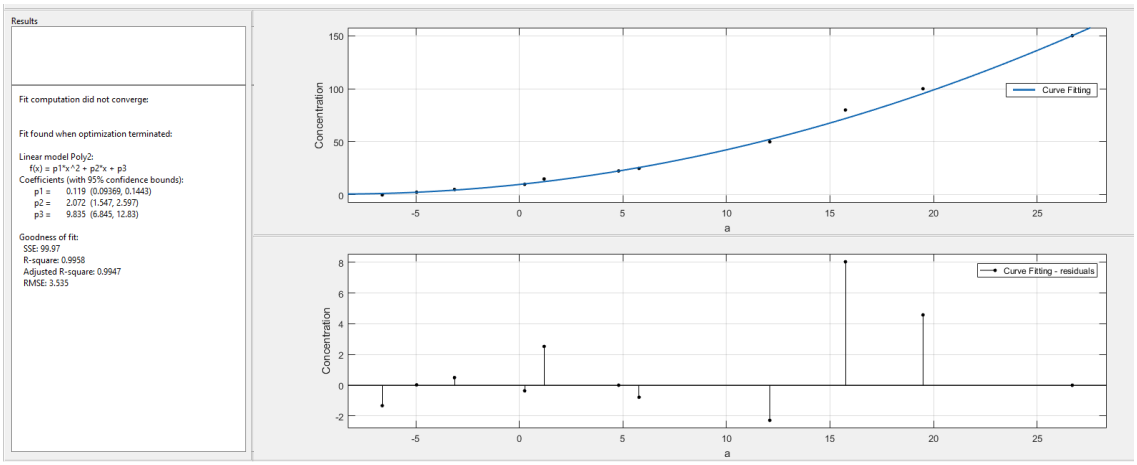


Figure 4.20: Curve Fitting using the a* component

As evidenced by Figure 4.19 and Figure 4.20, a second-degree polynomial function was used to fit a curve to the a* dataset and a second-degree exponencial function was used to fit a curve to the hue dataset. Looking at each of the curves residuals plots and their goodness of fit results it is evident that the functions generated by this software are extremely satisfactory and accurately describe both datasets used. Using the a* parameter produced the best goodness of fit results, therefore the curve fitted for this dataset is the one that will be used for the color/concentration model.

4.4.3 Model Validation

To validate the color/concentration model, images from the second set of image acquisitions (Testing Database) were used. Having prior knowledge about the SA concentration that is represented in each image allow us to test the accuracy of the model. Table 4.6 shows the different SA concentrations used for this validation.

Table 4.6: SMZ concentrations used in the database acquisition

SMZ Standard Concentration [μgL^{-1}]	**SMZ standard volume [mL]**	**SMZ concentration [μgL^{-1}]**	**Sample volume [mL]**
Blank	0	0.0	10
1000	0.025	2.5	10
1000	0.050	5.0	10
1000	0.100	10.0	10
1000	0.150	15.0	10
1000	0.200	20.0	10
1000	0.250	25.0	10
1000	0.500	50.0	10
1000	0.750	75.0	10
1000	1.000	100.0	10
1000	1.200	120.0	10
1000	1.500	150.0	10

For each of the concentrations displayed in Table 4.6, 9 images were used (3 from each camera). The results of this validation are presented in Figure 4.21, where the calculated value for the SA concentration is the average of the 9 images.

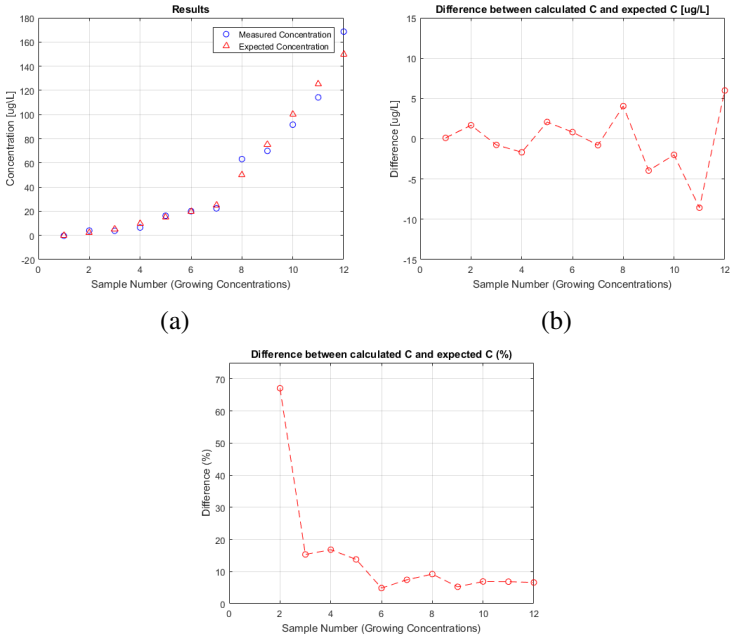
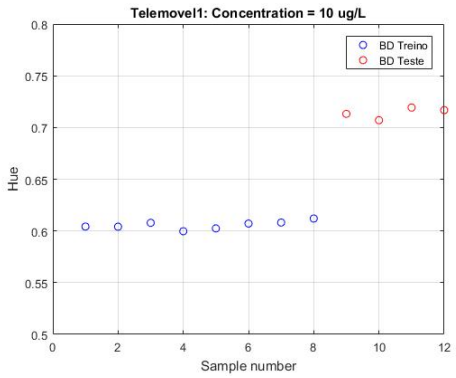


Figure 4.21: Model Validation (a): Concentration: Expected vs Calculated; (b): Absolute Error; (c): Relative Error

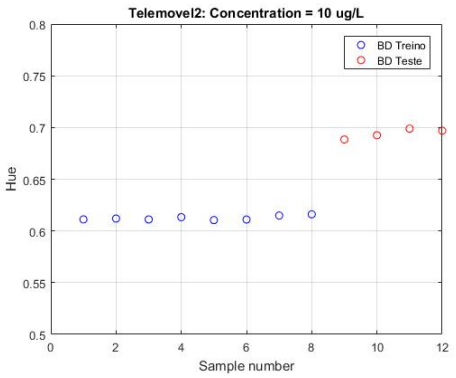
With an average error of 22.3% between the measured value of concentration, these results were not as good as was expected. The error between the measured value of concentration and the expected is too high for a correct determination of the SA concentration in a given sample. However, for higher concentrations of SA, the performance of our system is much better, with the average error dropping to approximately 9.02%.

To understand what went wrong in this stage of the work and to define what can be done to fix this problem, several tests were devised.

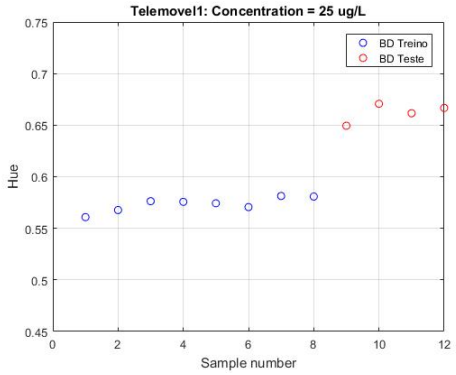
The first test consisted in a comparison of the color information extracted from samples with the same SA concentration between images from the two different databases (Training and Testing). The results of this test are displayed in Figure 4.22.



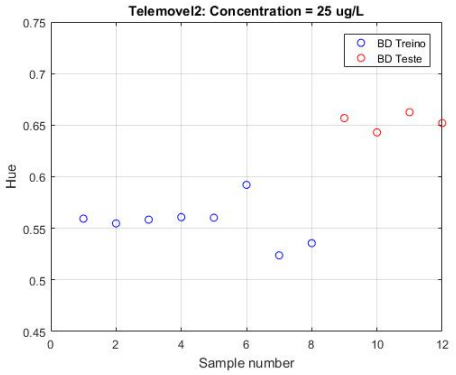
(a) Camera 1 :SA Concentration = 10μg L^{-1}



(b) Camera 2 :SA Concentration = 10μg L^{-1}



(c) Camera 1 :SA Concentration = 25μg L^{-1}

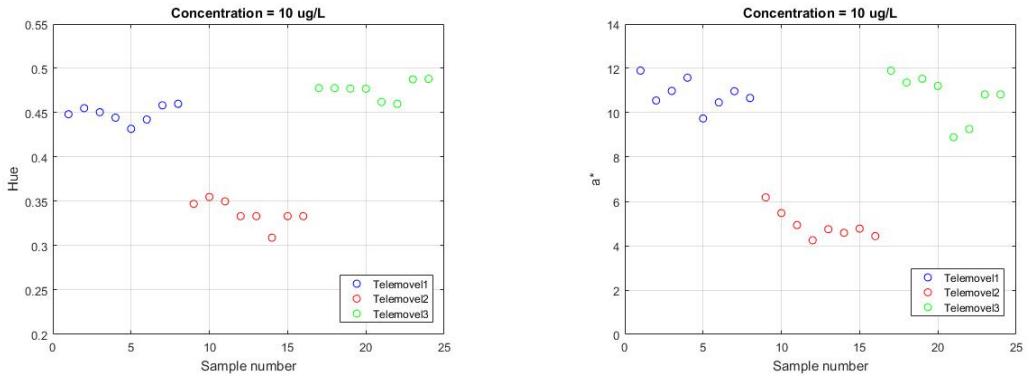


(d) Camera 2 :SA Concentration = 25μg L^{-1}

Figure 4.22: Image Database Comparison

These results clearly show that there is a noticeable problem between images from the different databases. The most probable cause for this effect is the quality degradation of color rendition chart. The two image databases were acquired with a three-month interval between them using the same print of the color rendition chart. During this time, some of the color of that print must have faded, meaning that the color correction algorithm did not work as intended for the images in the Testing Database. Another explanation could be that, since we used an unofficial print of the color chart instead of the real thing, the reference values used might not accurately represent the actual values for each color patch.

The next test compared the color measurements for samples with the same SA concentration and from the same database (Training Database) after color correction between the three different cameras used. The results of this test as shown in Figure 4.23.

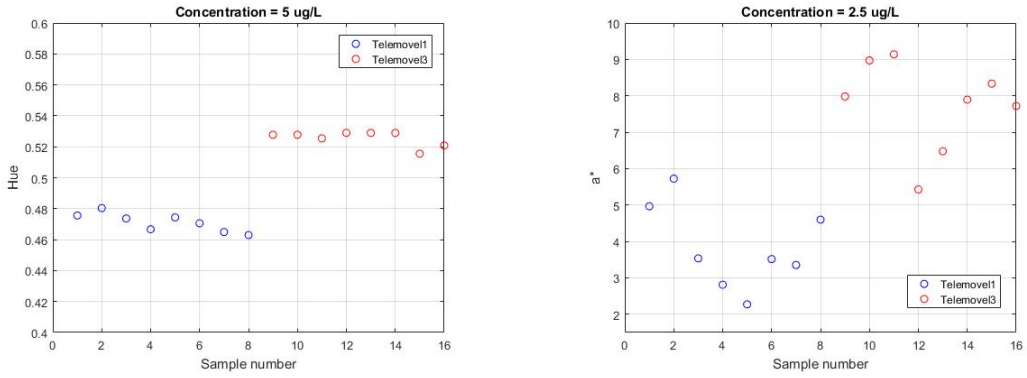


(a) Hue values for a SA Concentration = 10μg L^{-1} (b) a* values for a SA Concentration = 10μg L^{-1}

Figure 4.23: Smarthphone Camera Comparison

As evidence by Figure 4.23, there is clearly an under-performing camera - camera 2. As seen in the Table 3.1, the camera of this device (Vodafone Smart Mini) has a much lower resolution than the other two, which can explain this occurrence. Because the color/concentration model was created using the average values for the color measurements of the images from the three different cameras, this variance can negatively impact the overall quality of the model. As a result of this problem, the images acquired with camera 2 should no longer be used. This also means that, for the color correction algorithms to work properly, there is a required minimum camera resolution to acquire the images.

After noticing that problem, it was decided that it was necessary to analyze the color measurements from each individual sample. This analysis yielded another important observation: there is some variance between measurements of samples with the same SA concentration for images captured with the different cameras (see Figure 4.24).



(a) Hue values for a SA Concentration = 5μg L^{-1} (b) a* values for a SA Concentration = 2.5uμg L^{-1}

Figure 4.24: Color Measurements Analysis

Presented in Figure 4.24 is a comparison between the hue and a* values measured for the same SA concentration from 2 different cameras after color correction. This variance detected in the measurements of samples with the same SA concentration for images captured with the

different cameras means that the color correction algorithm used to process the images is not good enough to create a color/concentration model that is device-independent.

Using this newly acquired information, a new model was created using only the images collected with the same device - Huawei P9 Lite, only from the training database. To validate this new model, a cross-validation technique known as “leave-one-out” was used. This technique divides an entire dataset into mutually exclusive subsets [77]. One of these subsets is used to create the model, while the other is used to calculate its accuracy (see Figure 4.25). This process is repeated multiple times alternating the composition of the data subsets resulting in the creation of a well-rounded model.

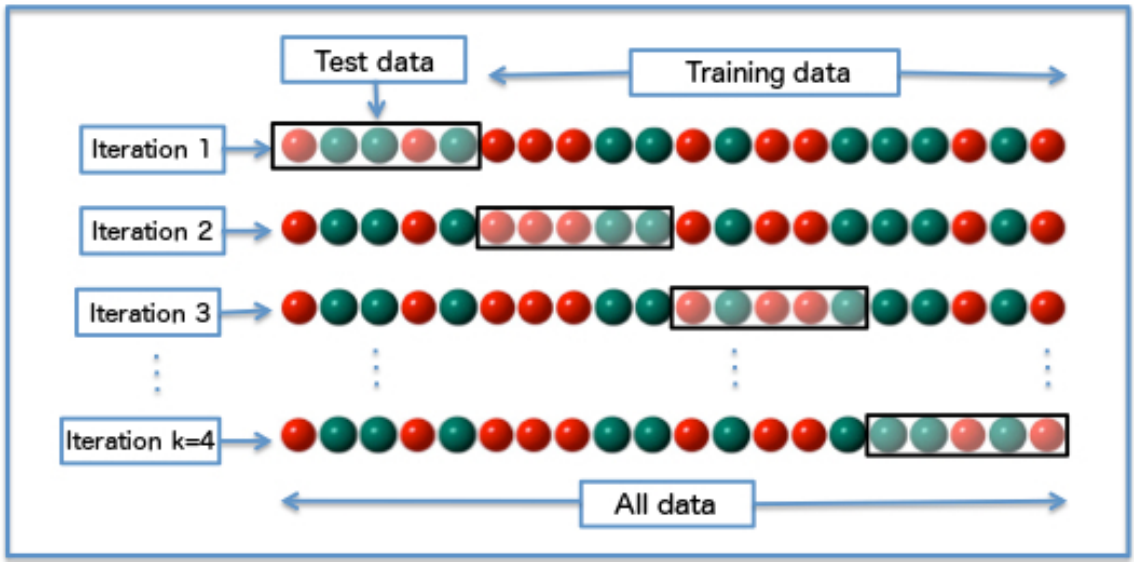
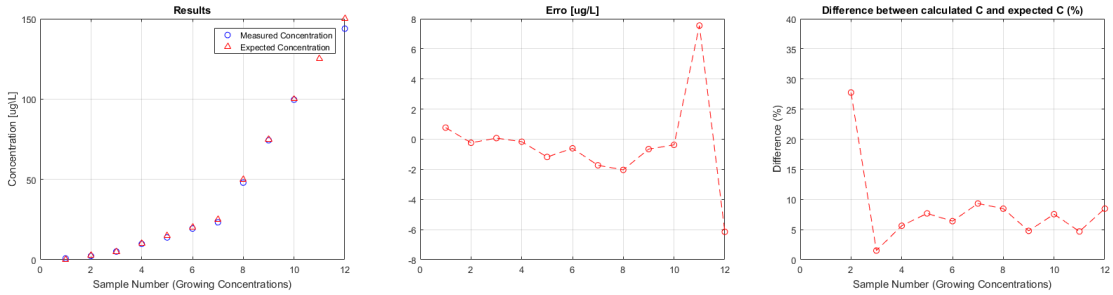


Figure 4.25: "Leave-one-out" cross validation method [6]

Presented in Figure 4.26 are the results obtained using this new approach.



(a) (b) (c)

Figure 4.26: New Model Validation (a): Concentration: Expected vs Calculated; (b): Absolute Error;(c): Relative Error

As evidenced by Figure 4.26, these results are much better than before. The average error between the estimated concentration in the sample and its actual concentration is less than 10%,

which is acceptable, for such an application. An error that low, means that, at the very least, this approach can be used as a robust screening method for SAs.

When compared with the rest, the error calculated for the lowest concentration used is alarmingly high. An explanation for this is given by the Horwitz function [7], proposed by William Horwitz, that states that as the analyte concentration decreases, the relative standard deviation (RSD) of said concentration increases (see Figure 4.27). This increase in the deviation leads to less accurate color readings from the lower concentrations of SA, because we cannot guarantee that exactly the same concentration is being used. Moreover, the lowest concentration tested is close to the limit of detection/quantification, where a largest relative uncertainty is expected.

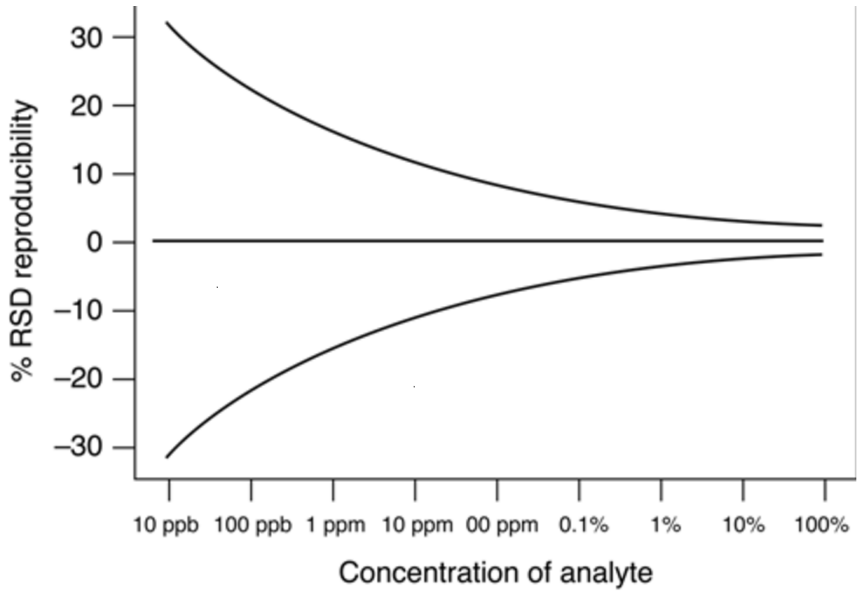


Figure 4.27: Horwitz Funtion also known as "Horwitz Trumpet" [7]

4.4.4 Limit of Detection

In analytical chemistry, the detection limit, lower limit of detection, or LOD (limit of detection), is the lowest quantity of a substance that can be distinguished from the absence of that substance (a blank value) within a stated confidence limit (generally 1%) [78] [79].

Figure 4.28 below illustrates the relationship between the blank, the limit of detection (LOD), and the limit of quantification (LOQ) by showing the probability density function for normally distributed measurements at the blank, at the LOD defined as three times standard deviation (σ) of the blank, and at the LOQ defined as 10σ of the blank.

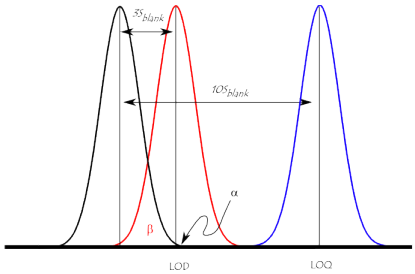
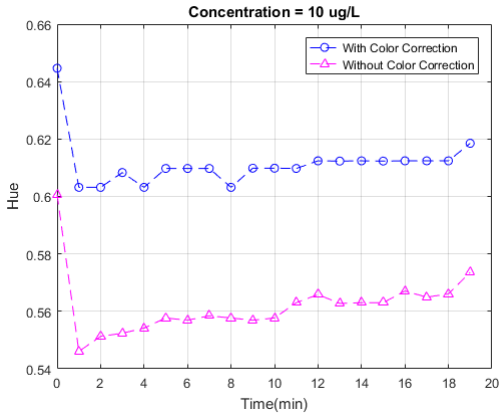


Figure 4.28: Illustration of the concept of LOD and LOQ [8]

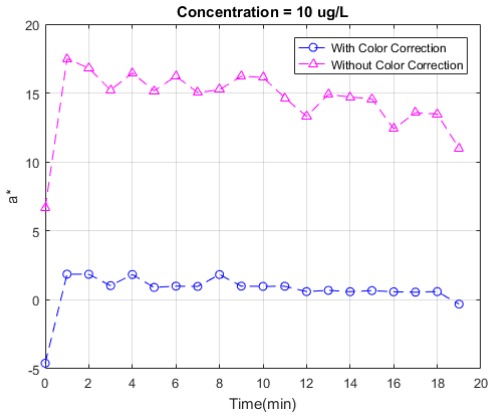
The limits of detection (LOD) for each color parameter (a* and Hue) were calculated as the concentrations that provided a signal to noise ratio (SNR) of 3. These values were 5 μg L^{-1} and 3.5 μg L^{-1} for color parameters Hue and a*, respectively. These values are very satisfactory, since it proves that our system has enough precision and sensitivity to detect and determine very low concentrations of SA.

4.4.5 Color Stability

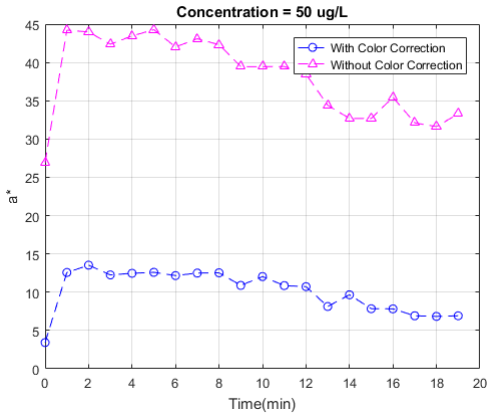
In this section, the results from the study of the color stability of the samples are presented. This study was done by using 3 different SA concentrations: 10, 50 and 100 μg L^{-1}. For each of the concentrations, starting from the moment the reagent was applied to the disk, an image was captured every minute for 20 minutes. The results of this study are shown in Figure 4.29.



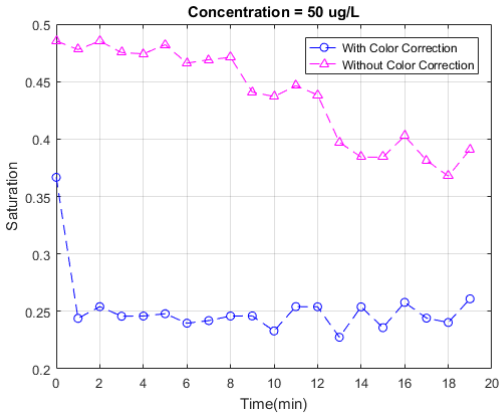
(a) Hue (C = 10 μg L^{-1})



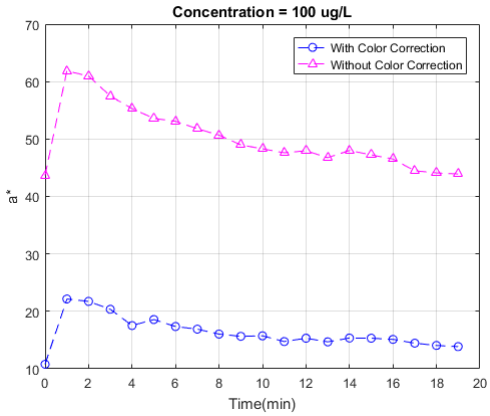
(b) a* (C = 10 μg L^{-1})



(c) a* (C = 50 μg L^{-1})



(d) Saturation (C = 50 μg L^{-1})



(e) a* (C = 100 μg L^{-1})

Figure 4.29: Results form the color stability study

As evidenced by Figure 4.29, the color of the different samples stabilizes after approximately 1 minute, and remains stable for, at least, 20 minutes after the application of the reagent. For the image database acquisition, the waiting time between the application of the reagent and the image acquisition was 90 seconds. Although it would have been better to wait a couple more seconds, this waiting time is sufficient for the color of the sample to stabilize.

4.4.6 Summary

In this section, a brief summary of the results is given. Firstly, the contents of the image database were described. Judging by the total number of different images present in the image database, it is safe to assume that, more than enough images were at our disposal in order to achieve very robust results.

After that, we evaluated the results of the image processing algorithms. Here, we concluded, that the technique used for the sample segmentation performs well. The results for all four color checker segmentation methods were compared, and while all of them performed reasonably well, the edge-based segmentation (Canny) method outperformed the rest.

In the next part of this chapter, we compared the results from the different color constancy algorithms proposed in this thesis. It was concluded that a simple WB algorithm is not sufficient to eliminate illumination-based variance in different images, as it was the color constancy algorithm that performed the worst. Although we thought that the other color correction methods performed well, we later proved that their performance also does not eliminate inconsistencies between images acquired with different devices. Because of this, a device-independent color/concentration model could not be determined. However, a very accurate, albeit device-dependent model was achieved.

After some additional tests, it was concluded that the probable reason for the underperformace of the color correction algorithms was the lack of an official color rendition chart.

The final model, created using only images acquired with the same device, was determined after exploring different probable causes for the shortcomings of the first model. Using this new model, we found that our system is capable of determining SA concentrations with an average error inferiror to 10%.

Lastly, some additional tests on the LOD and color stability were done. Our system's LOD was calculated as 3.5 μg L^{-1}. The final results show that color stability of the chemical reaction is achieved approximately 1 minute after the application of the reagent, and is maintained for, at least, 19 minutes after.

Chapter 5

Android Application

In order to determine SA concentrations in point-of-care tests, the creation of an Android application based on the results presented in the previous chapters is proposed. In this chapter, a description of the system architecture idealized for the application is presented. Also in this chapter, the system functional requirements are listed and the preliminary mock-up for this software is described in detail.

5.1 Functional Requirements

The first stage in the development of new Android application is to establish its functional requirements. These functional requirements are listed below:

- The interface of the application must be user-friendly, so as to no alienate less experienced users;
- The software must allow the user to make as many measurements as desired;
- The application must be able to store results of previous readings;
- The user must be able to analyze images from the target device memory storage.

The final product must accommodate every requirement.

5.2 System Architecture

Displayed in Figure 5.1 is the proposed system architecture for the Android application.

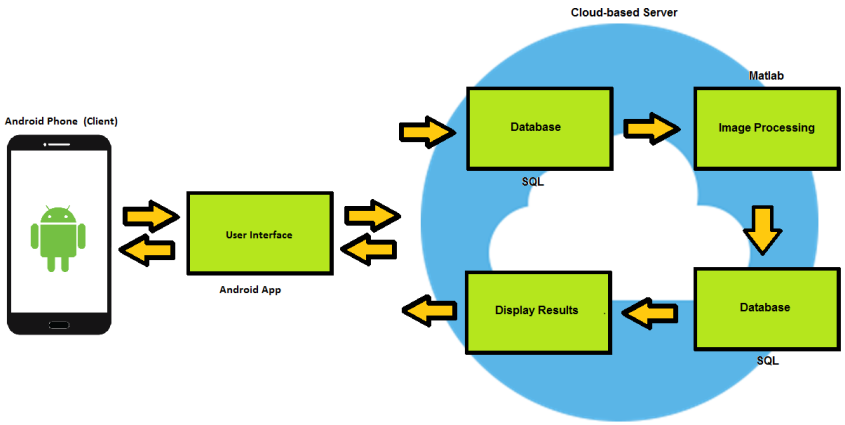


Figure 5.1: System Architecture of the Android application

The system architecture idealized for the Android application is simple, easy to implement and easy to understand. The application will connect the user of an Android phone, also known as the client, with a cloud-based server, where the image processing algorithms necessary to determine SA concentration are stored.

The user interface will then allow the user to upload an image directly from the device memory storage to the server, where it will be processed. Once this image processing step is complete, the results of this digital image colorimetry technique are sent to back to the client, where the user will have to option to view the results or store them in an online database, for future use.

5.3 Mock-up

In this section, a mock-up of the Android application is presented to give a better idea of how the finished product will look like. Displayed in following figures, are the different screens and features proposed for the final product. Figure 5.2 shows the Home screen and the Main Menu.

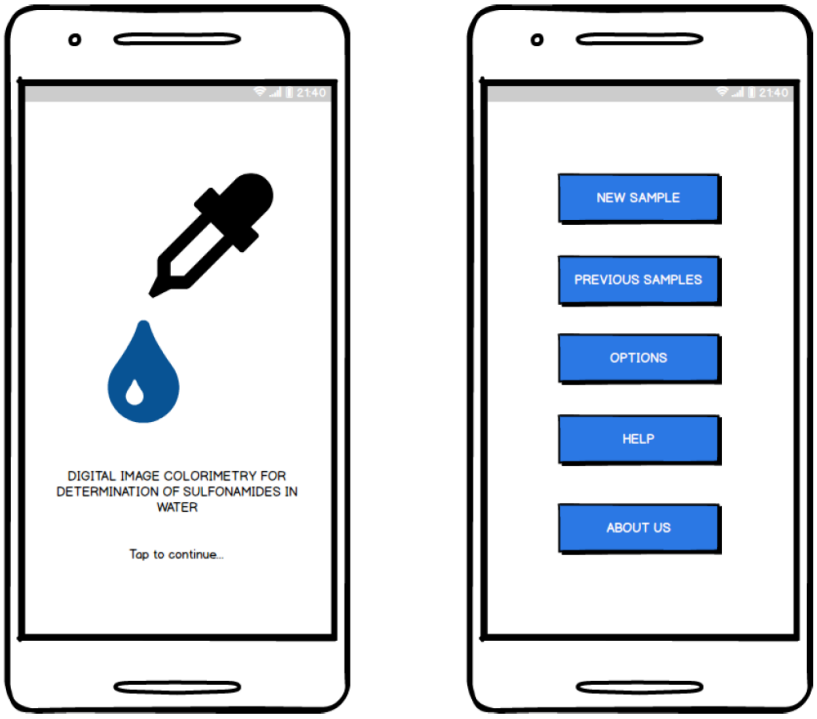


Figure 5.2: Home Screen and Main Menu

The Home screen will be the landing page, therefore this will be first screen the user sees. It contains the name and logo of the application. The Main Menu displays the all the actions the user can perform. A typical interaction for determining the SA concentration is depicted in Figure 5.3.

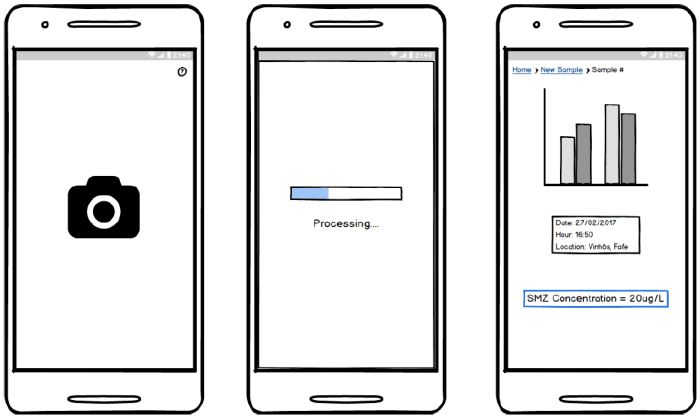


Figure 5.3: The 3 step process for SA determination

To get the SA concentration values of a given sample, the user must perform 3 easy steps: Upload the image intended for analysis, wait for the application to finish processing the image and finally, obtain the results. Unless the user chooses otherwise, this software automatically stores the results for every sample it analyzes. Along with the SA concentration present in the sample, the time, data and approximate location at the time of the analysis. These results are available for user consultation as displayed by Figure 5.4.

Figure 5.4: Additional Functionality: Consult past results

While consulting the stored results, the user will have the option to upload some results to a different platform. There will also be a option that allows the user to clear some, or all, the stored

results. Finally, Figure 5.5, displays some additional features of the software, such as: the About Us screen, the Frequently Asked Question (FAQ) section and an Options menu.

Figure 5.5: About Us, FAQ and Options Menu

Through the About Us screen, the user has access to information about the software developers, such as, motivations and the mission statement. This screen will also provide useful links and the developers contact information. As the name states, in the FAQ screen, the most frequently asked questions will be displayed. This is intended to allow the user to troubleshoot some problems he/she might encounter while using the software. Finally, through the Options Menu, the user can change some settings, such as the time/date format and the preferred unit of measurement to display concentration values.

5.4 Initial Prototype

Based on the system architecture and mock-ups presented earlier, a initial working prototype of the Android application was created. In this section, this prototype will be described in detail.

As soon as the user starts the application, the Home Screen will be displayed. Present in the Home Screen are two buttons: 'Upload Image' and 'Get Results' (Figure 5.6a). By pressing the 'Upload Image' button, the software will access the device's memory storage and prompt the user to select an image to be processed (Figure 5.6b). Once the image is selected, a preview of the selected image will be displayed in the Home Screen and the user will be asked name the image (Figure 5.6c).

(a) Home Screen (b) Select Image from Gallery (c) Image Preview

Figure 5.6: Initial Prototype: Upload an Image

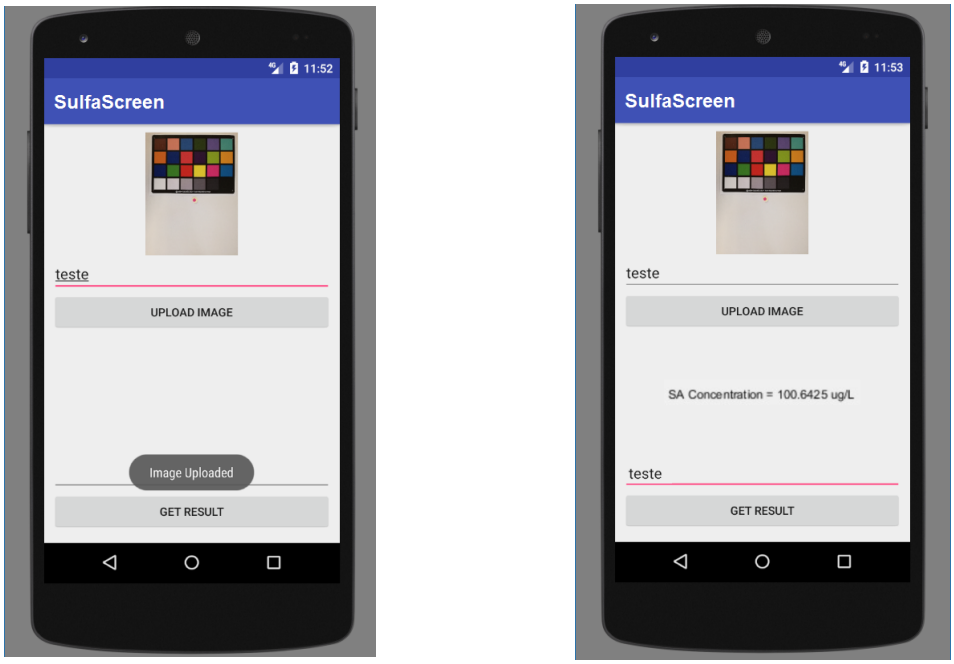
If the user wishes to change the image to be processed, he can simply press the image preview, which will instruct the application to, once again, open the device's memory storage, allowing the user to pick a different image.

Pressing the 'Upload Image' button after an image has been selected will start the upload process. This process starts by converting the selected image into a string using Base64 encoding. This string is then sent to an online server, via HTTP, where it will be decoded using a PHP script. The image is then stored into the server's image database, created using Structured Query Language (SQL).

When a new image is added to this database, the MatLab script is triggered to run the colorimetric analysis described in the earlier chapters. The result of this analysis is stored in a different database along with the name previously provided by the user.

Once this process is completed, a notification will appear on the application's Home Screen, informing the user that the image was uploaded successfully and the results are ready to be displayed (Figure 5.7a).

In order to view these results, the user must simply enter the corresponding image name and press the 'Get Results' button. This action will prompt the application to search the online database for the result that matches the name that was provided. This result is displayed in the Home Screen, informing the user of the SA concentration value calculated from the sample present in the image (Figure 5.7b).



(a) Successful Upload Notification

(b) Display Results

Figure 5.7: Initial Prototype: Displaying Results

The entire process from uploading the image to getting the results, takes, on average, 25 seconds, provided that there is a strong internet connection.

Since this is an initial prototype, only the minimum requirements are implemented. Additional features will be implemented at the later stages of the software's development cycle. However, the performance of this prototype was satisfactory, which validates the system architecture proposed.

5.5 Summary

Due to time constraints, the Android application proposed in the early chapters of this dissertation, is still in the early stages of its development. However, the functional requirements of the software have been established and a preliminary system architecture has been proposed. To give a broad idea of the final product, a mock-up has also been provided. Finally, a working initial prototype of the application was presented.

Chapter 6

Conclusions and Future Work

In this chapter, the final conclusions of this thesis, as well as, some future work suggestions for this work are highlighted.

6.1 Conclusions

At this point of the dissertation is important to look back and evaluate the work done and discuss some important results.

According to the literature review presented in chapter 2, the method proposed in this thesis to determine SA concentrations is very unique. This method is based in digital image colorimetry, meaning it uses automatic image processing algorithms to analyze the color characteristics of a given sample such as, image segmentation and color correction.

While the results for the various image segmentation methods studied in this thesis were satisfactory, the same cannot be said about the color correction methods. Guaranteeing color constancy throughout the images captured by different devices was a crucial step in the development of a device-independent algorithm for SA determination. With that being said, it is necessary to improve the performance of said algorithms.

Even though a device-independent algorithm was not achieved, a device-dependent alternative was successfully created. This alternative can be used to determine the SA concentration of a given sample with an average error inferior to 10% and a LOD of 3.5 μg L^{-1}, which is very satisfactory for such an application.

Since, it is believed that this is a novel approach to determine SA concentrations in water, I hope that the work presented in this dissertation will motivate further research in this field.

6.2 Future Work

Here are some future work suggestions, that we believe would improve the overall results of the work presented in this thesis.

The first, and more obvious suggestion is to re-evaluate the performance of the color correction methods, by acquiring an official industry standard color rendition chart. By using a printed version of the color rendition chart the results of the color correction process would always be questionable.

The second suggestion is to develop an artificial neural network to automatically establish the relationship between the color of a sample and its concentration. Although the software used to relate these two parameters was satisfactory, we believe that by using a neural network, the overall system would be much more robust and produce better results.

Further testing of the system robustness is also suggested. For example, it should be interesting to determine the system's response to temperature variations.

To make this system ready for point-of-care tests, we suggest improving the image processing algorithms. Using this system outside a controlled environment presents a greater number of obstacles, and it is important to assure that this system still presents accurate results even under harsher conditions.

Some general improvements to the user interface of the Android application are recommended. Finally, the creation of a similar application compatible with the Apple operating system (OS) is suggested.

References

[1] Vladimir V. Apyari Veronika V. Tolmacheva Yury A. Zolotov Stanislava G. Dmitrienko, Elena V. Kochuk. Recent advances in sample preparation techniques and methods of sulfonamides detection – a review. *Analytica Chimica Acta*, August 2014.

[2] W.P. Jencks. Catalysis in chemistry and enzymology. *New York: McGraw-Hill*, 1969.

[3] Gusakova N.N. Eremenko S.N. Chernova, R.K. and S.Yu. Doronin. *Ser. Khim. Khim. Tekhnol., vol. 39, no. 6*, pages 30–34, 1996.

[4] N. Ahuja E. Ozdalga, A. Ozdalga. The smartphone in medicine: a review of current and potential use among physicians and students. *J. Med. Internet Res. 14*, 2012.

[5] E. Dickey C. Pishko L.F. Capitan-Vallvey, in: C.A. Grimes. Encyclopedia of sensors, 1st ed. *The Pennsylvania State University, Pennsylvania, USA*, 2005.

[6] Cross Validation Methods. https://commons.wikimedia.org/w/index.php?curid=51562781. *Acessed June 20, 2017*.

[7] Thompson M. The amazing horwitz function. *AMC Technical Brief No.17, Royal Society of Chemistry*.

[8] Limit of Detection. https://commons.wikimedia.org/wiki/file:lod.png. *Acessed June 20, 2017*.

[9] Antibiotic Resistance Threats. https://www.cdc.gov/drugresistance/pdf/ar-threats-2013-508.pdf. *Acessed June 24, 2017*.

[10] H. Song Y.L. Pan X. Wang X.T. Zhao, Q.B. Lin. *Agric. Food Chem. 59*, pages 9800–9805, 2011.

[11] J. Wang Y. Liu R. Wang W. Cao X. Shi, J. Liu. *Anal. Lett. 43*, pages 2246–2256, 2010.

[12] Z. Li R. Li J. R. Liu, P. He. *Chromatogr. Sci. 49*, pages 640–646, 2011.

[13] Y. Yang M.U. Farooq, P. Su. *Chromatographia 69*, pages 1107–1111, 2009.

[14] I. Varenina G. Scortichini L. Annunziata M. Brstilo N. Rudan N. Bilandžic, B.S. Kolanovic. *Food Control 22*, pages 1941–1948, 2011.

[15] Y.H. Chung K.G. Lee H.H. Chung, J.B. Lee. *Food Chem.113*, pages 297–301, 2009.

[16] S. Moretti B. Puppini G. Saluti L. Persic R. Galarini, F. Diana. *Food Control 35*, pages 300–310, 2014.

[17] E.Y. Frag El-Dien, G.G. Mohamed. *Chem. Pap. 63*, pages 646–653, 2009.

[18] S.G. Dmitrienko E.V. Klokova. *Mosc. Univ. Chem. Bull. 63*, pages 284–287, 2008.

[19] J.L.F.C. Lima L. Pezza H.R. Pezza T.A. Catelani, I.V. Tóth. *Talanta 121*, pages 281–287, 2014.

[20] B.U. Minbaev. Shiffovy osnovaniya (schiff bases). *Alma-Ata: Nauka*, 1989.

[21] I.M. Korenman. *Moscow: Khimiya*, 1970.

[22] Chernova R.K. Cavvin, S.B. and S.N. Shtykov. Surface-active substances. *Moscow: Nauka*, 1991.

[23] Yatzimirsky A.K. Osipov A.P. Martinek, K. and J.V Beresin. *Tetrahedron, vol.29, no.5*, pages 960–963, 1973.

[24] Mityakina M.G. Yatsimirskaya N.T. Krivova, S.B. and A.K. Osipov. *Vestn. Mosk. Gos. Univ., Ser. 2: Khim., vol 32, no. 4*, pages 365–367, 1991.

[25] Krivova S.B. Yatsimirskaya N.T. Sokolovskaya, E.M. and V.N. Rybalkov. *VINITI, Moscow, no. 4070-V90*, 1990.

[26] Yatzimirskaya N.T. Yatzimirsky, A.K. and S. Kashina. *Anal. Chem., vol. 66*, pages 2230–2232, 1994.

[27] A.J. Ricco M.X. Tan D.E. Williams V. Gubala, L.F. Harris. Point of care diagnostics: status and future. *Anal. Chem. 84*, pages 487–515, 2011.

[28] A. Rasooly K.E. Herold. Lab-on-a-chip technology (vol. 2): Biomolecular separation and analysis. *Caister Academic Press, Norwich, UK*, 2009.

[29] J. Mettakoonpitak C.S. Henry D.M. Cate, J.A. Adkins. Recent developments in paper-based microfluidic devices. *Anal. Chem. 87*, pages 19–41, 2014.

[30] V.G. Amelin Yu Zolotov, V.M. Ivanov. Chemical test methods of analysis. *1st ed., Elsevier, Amsterdam*, 2002.

[31] U. Sikora S. Padmanabhan I. Navruz A. Ozcan O. Mudanyali, S. Dimitrov. Integrated rapid-diagnostic-test reader platform on a cellphone. *Lab Chip 12*, pages 2678–2686, 2012.

[32] A. Ozcan. Mobile phones democratize and cultivate next-generation imaging, diagnostics and measurement tools. *Lab Chip 14*, pages 3187–3194, 2014.

[33] D.W. Sun T. Brosnan. Improving quality inspection of food products by computer vision. a review. *J. Food Eng. 61*, pages 3–16, 2004.

[34] W.S. Stiles G. Wyszecki. Color science: Concepts and methods, quantitative data and formulae. *Wiley Classics Library, Denver, USA*, 2000.

[35] D. Pascale. A review of rgb color spaces from xyz to rgb. *The Babel- Color Company*, 2002.

[36] O.M. Medvedeva S.A. Badakova L.N. Pyatkova Y. Shishkin, S.G. Dmitrienko. Use of a scanner and digital image-processing software for the quantification of absorbed substances. *J. Anal. Chem. 59*, 2004.

[37] C.S. Kim J. Park, W. Hong. Color intensity method for hydrogel oxygen sensor array. *IEEE Sens. J 10*, pages 1855–1862, 2010.

[38] X.D. Wang M. Landthaler P. Babilas O.S. Wolfbeis R.J. Meier, S. Schreml. Simultaneous photographing of oxygen and ph in vivo using sensor films. *Angew. Chem. Int. Ed. 50*, pages 10893–10896, 2011.

[39] N. Lopez-Ruiz A. Rivadeneyra Torres L.F. Capitan-Vallvey A.J. Palma A. Martinez-Olmos, J. Fernandez-Salmeron. Screen printed flexible radiofrequency identification tag for oxygen monitoring. *Anal. Chem. 85*, 2013.

[40] K. Suzuki Y. Suzuki. Optical sensors for ions and protein based on digital color analysis, frontier in chemical sensors. *Springer Ser. Chem. Sensors Biosens. 3*, pages 343–365, 2005.

[41] S.D. Cotton. Colour, colour spaces and the human visual system. *School of Computer Science University of Birmingham, Birmingham UK*, 1996.

[42] J. Berlien. Color classification with the tcs230 identifying and sorting colors by hue. *Texas Advanced Optoelectronic Solutions*, 2004.

[43] I. Orbe-Paya L.F. Capitan-Vallvey K. Cantrell, M.M. Erenas. Use of the hue parameter of the hue, saturation, value color space as a quantitative analytical parameter for bitonal optical sensors. *Anal. Chem. 82*, pages 531–542, 2010.

[44] R.C. Campos-D. Gannerman S. Paciornik, A.V. Yallouz. Scanner image analysis in the quantification of mercury using spot-tests. *J. Braz. Chem. Soc. 17*, pages 156–161, 2006.

[45] E.D. Marinetto-C.A. Abad I. de Orbe-Paya A.J. Palma-L.F. Capitan-Vallvey A. Garcia, M.M. Erenas. Mobile phone platform as portable chemical analyzer. *Sens. Actuators B 156*, pages 350–359, 2011.

[46] D.K. Griffin-N. Kelley-Loughnane I. Papautsky-J.A. Hagen R.C. Murdock, L. Shen. Optimization of a paper-based elisa for a human performance biomarker. *Anal. Chem. 85*, 2013.

[47] K.S. Suslickn B.A. Suslick, L. Feng. Discrimination of complex mixtures by a colorimetric sensor array: coffee aromas. *Anal. Chem. 82*, pages 2067–2073, 2010.

[48] A. Martinez-Olmos M.P. Cuellar-M.d.C. Pegalajar A.J. Palma I.d. Orbe-Paya L.F. Capitan-Vallvey S. Capel-Cuevas, N. Lopez-Ruiz. A compact optical instrument with artificial neural network for ph determination. *Sensors (Basel) 12*, pages 6746–6763, 2012.

[49] M.d.C. Pegalajar I. de Orbe-Paya-L.F. Capitan-Vallvey S. Capel-Cuevas, M.P. Cuellar. An expert system for full ph range prediction using a disposable optical sensor array. *IEEE Sens. J 12*, pages 1197–1206, 2012.

[50] A. Salinas-Castillo M.C. Pegalajar-J. Vukovic L.F. Capitan-Vallvey M. Ariza-Avidad, M.P. Cuellar. Feasibility of the use of disposable optical tongue based on neural networks for heavy metal identification and determination. *Anal. Chim. Acta 783*, pages 56–64, 2013.

[51] N.N. Gusakovam S.Y. Doronin, R.K. Chernova. Condensation of p-dimethylaminocinnamaldehyde with aniline and substituted anilines in micellar media. *Russ. J. Gen. Chem. 75*, pages 261–267, 2005.

[52] S.G. Dmitrienko E.V. Klokova. Spectrophotometric determination of sulfanilamides by a condensation reaction with p-dimethylaminocinnamaldehyde. *Mosc. Univ. Chem. Bull. 63*, pages 284–287, 2008.

[53] J.L. Lima L. Pezza H.R. Pezza T.A. Catelani, I.V. Toth. A simple and rapid screening method for sulfonamides in honey using a flow injection system coupled to a liquid waveguide capillary cell. *Talanta 121*, pages 281–287, 2014.

[54] V.V. Tolmacheva V.V. Apyari Y.A. Zolotov S.G. Dmitrienko, E.V. Kochuk. Determination of the total content of some sulfonamides in milk using solid-phase extraction coupled with off-line derivatization and spectrophotometric detection. *Food Chem. 188*, pages 51–56, 2015.

[55] Huawei P9 Lite. https://www.pickaboo.com/huawei-p9-lite.html. *Acessed June 20, 2017*.

[56] Vodafone Smart Ultra 6. http://www.gsmarena.com/vodafonesmartultra6-pictures-7313.php. *Acessed June 20, 2017*.

[57] Vodafone Smart Mini. http://www.fnac.pt/vodafone-smart-mini-black-telemovel-telemovel-smartphone-vodafone/a717641. *Acessed June 20, 2017*.

[58] Rafael C. Gonzalez; Richard E. Woods. Digital image processing. *Prentice Hall*, pages 1–3, 2008.

[59] Linda G. Shapiro and George C. Stockman. Computer vision. *New Jersey, Prentice-Hall*, pages 279–325, 2001.

[60] Lauren Barghout and Lawrence W. Lee. Perceptual information processing system. *Paravue Inc. U.S. Patent Application 10/618,543, filed July 11, 2003.*

[61] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. Sys., Man., Cyber. 9 (1)*, pages 62–66, 1979.

[62] Richard O. Duda and Peter E. Hart. Use of the hough transformation to detect lines and curves in pictures. 1971.

[63] Jeppe Jensen. Hough transform for straight lines.

[64] J Canny. Aa computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6)*, pages 679–698, 1986.

[65] T Moeslund. Canny edge detection. *Lecture from December 3, 2014*.

[66] G. Biernson. A feedback-control model of human vision. *Proceedings of the IEEE, vol. 54, no. 6*, pages 858–872, 1966.

[67] C. Cusano S. Bianco, G. Ciocca and R. Schettini. Improving color constancy using indoor/outdoor image classification. *Image Processing, IEEE Transactions on, vol. 17, no. 12*, pages 2381–2392, 2008.

[68] A. W. Morales Y. Kim, J.-S. Lee and S.-J. Ko. A video camera system with enhanced zoom tracking and auto white balance. *Consumer Electronics, IEEE Transactions on, vol. 48, no. 3*, pages 428–434, 2002.

[69] T. Gevers A. Gijsenij and J. van de Weijer. Computational color constancy: Survey and experiments. *Image Processing, IEEE Transactions on, vol. 20, no. 9*, pages 2475–2489, 2011.

[70] E. H. Land. The retinex theory of color vision. *Scientific American, vol. 237, no. 6*, pages 108–128, 1977.

[71] M. Ebner. Algorithms for color constancy under uniform illumination. *John Wiley Sons, Ltd*, pages 103–141, 2007.

[72] C.-L. Chen and S.-H. Lin. Intelligent color temperature estimation using fuzzy neural network with application to automatic white balance. *Expert Systems with Applications, vol. 38, no. 6*, pages 7718–7728, 2011.

[73] Tali Treibitz Bei Xiao Umut A. Gurkan Justine J. Allen Utkan Demirci Akkaynak, Derya and Roger T. Hanlon. Use of commercial off-the-shelf digital cameras for scientific data acquisition and scene-specific color calibration. *Journal of the Optical Society of America A 31, no. 2 (January 20, 2014)*.

[74] David M W Powers. Evaluation: From precision, recall and f-measure to roc, informedness, markedness correlation. *Journal of Machine Learning Technologies. 2 (1)*, pages 37–67.

[75] Evaluating Goodness of fit. https://www.mathworks.com/help/curvefit/evaluating-goodness-of-fit.html. *Acessed June 20, 2017*.

[76] Sandra Lach Arlinghaus. Phb practical handbook of curve fitting. *CRC Press*, 1994.

[77] R. KOHAVI. A study of cross-validation and bootstrap for accuracy estimation and model selection. *International joint Conference on artificial intelligence. [S.l.: s.n.], v.14*, pages 1137–1145, 1995.

[78] IUPAC. Compendium of chemical terminology, 2nd ed. (the "gold book"). 1997.

[79] Warren B.; et al. MacDougall, Daniel; Crummett. Guidelines for data acquisition and data quality evaluation in environmental chemistry. *Analytical Chemistry, 52*, pages 2242–2249, 1980.