A MOBILE-PHONE BASED WATER TESTING KIT APPLICATION

By
CS 14-15

DEPARTMENT OF COMPUTER SCIENCE
SCHOOL OF COMPUTING AND INFORMATICS TECHNOLOGY
COLLEGE OF COMPUTING AND INFORMATION SCIENCES

A Project Report Submitted to the School of Computing and Informatics Technology
For the Study Leading to a Project in Partial Fulfillment of the Requirements for the Award of the Degree of Bachelor of Science in Computer Science of Makerere University

Supervisor
MR.MWEBAZE ERNEST

Department of Information Technology
School of Computing and Informatics Technology
College of Computing & Information Sciences (CoCIS), Makerere University
E-mail:emwebaze@cis.mak.ac.ug, Tel: +256-41-540628,
Fax:+256-41-540620

June, 2014

Declaration

We group CS 14-15 do hereby declare that this Project Report is original and has not been published and/or submitted for any other degree award to any other University before.

	NAME	REGISTRATION NUMBER	SIGNATURE
1	LUGOBWA JOSHUA IVAN	11/U/9210/PS	
2	BABIRYE CLAIRE	11/U/11484/PS	
3	NANKUNDA PETER	11/U/11451/PS	
4	WASSWA NELSON	11/U/11706/EVE	

Date: ....

Approval

This Project Report has been submitted for Examination with the approval of the following supervisor.
Signed : ....

Date : ....

Mr. Ernest Mwebaze
Department of Information Technology,
SCHOOL OF COMPUTING AND INFORMATICS TECHNOLOGY

Dedication

We Group CS 14-15 members dedicate this work to our families and friends who have encouraged us, prayed for us and boosted us at all times they all deserve a vote of thanks and more so to the Lord who made everything possible and beautiful at the right timing. We would also like to thank our parents and supervisor for their support both moral and material.

Acknowledgment

First and foremost, we thank the almighty God for having enabled us to successfully work out this project. We extend our sincere gratitude to our beloved parents, guardians and great friends for the spiritual and financial support accorded to us throughout the course of this journey of education.

We acknowledge Mr.Mwebaze Ernest our project supervisor for his support, patience and guidance throughout the entire time. Special thanks go to the administration of the School of Computing and Informatics Technology for availing the necessary resources including, supervisor, internet services and the library to mention just a few.

Also special thanks go to the AI-DEV research group of Makerere University for their support during the code implementation phase and clarification on the image classification algorithms. We also appreciate and thank National Water and Sewerage Corporation International Resource Center, particularly Dr.Babu Mohammed for the initial technical guidance in regard to the project. In addition, we also extend our heartfelt gratitude of appreciation to Dr. Kiryowa Outhman for the invaluable support in sharing his knowledge with us through the questionnaire he answered with out any hesitation. We are also grateful to Omusono labs for helping us print the 3D kit.

We thank you all and may the Lord bless you and reward you all.

Contents

List of Tables

List of Figures

Acronyms

UN	United Nations
WHO	World Health Organisation
NUDF	Northern Uganda Development Foundation
NCBI	National Centre for Biotechnology Information
NTU	Nephelometric Turbidity Unit
FNUs	Formazine Nephelometric Units
SURF	Speeded Up Robust Features
HSV	Hue Saturation Value
KNN	K-Nearest Neighbor
SIFT	Scale Invariant Feature Transformation
GIMP	GNU Image Manipulation Program
3D	Three-Dimensional
OpenCV	Open Computer Vision
NGO	Non Goverment Organization
4D	Four-Dimensional
CAD	Computer Aided Design
CS	Creative Suite
LED	Light Emitting Diode
STL	Standard Tessellation Language
2D	Two-Dimensional
AI-DEV	Artificial Intelligence for Developing Countries

Abstract

Currently, a significant number of communities are faced with an unbearable cost of treating water-related diseases resulting from consumption of unsafe and dirty drinking water. This study is about the development of a mobile-phone based water testing kit application, an application running on an Android platform. The application assesses the quality of drinking water and the focus is primarily on the water used by people in their homesteads, delivered from public water sources like taps, boreholes and wells. The study will avail to the homestead community a cheap and affordable tool to determine whether the water deemed to be safe is indeed safe for drinking or not, this will save the people from the high costs they incur while treating preventable water-related diseases.

Group CS 14-15 set out to review the relevant literature, analyze the needs for the application, design and implement the intended mobile water testing kit application. Data collection techniques such as questionnaires and document review were applied to get the necessary information. The literature was reviewed and analyzed in order for the group to acquire a proper understanding of the problem and what has been done so far. Mobile phones were used to take the required images of the water samples. Python programming language was used to implement the prototype and Java for the final system. The whole application was based on an Android platform. Adobe Photoshop CS 6 was used to design the application splash screen. Image libraries like Matplotlib and Numpy were used for extensive image processing.

Machine learning classification algorithms like KNN, Naive Bayes and SVM were used to automatically classify the state of water; whether safe or unsafe. Image data needed to train the algorithm was obtained locally using rain water, spring water, boiled and bottled water etc. The application also integrated a cuboidal 3D kit modeled and printed to provide for a light source, a hole for the phone camera, a holder for the water to be tested, a toggle switch and pair of batteries. The kit was designed using SketchUp Pro 2013 and Cinema 4D. We also carried out unit testing to ascertain the functionality of the light source component, toggle switch and later integrated it into the whole system. Image classifier evaluation was also carried out for both Naive Bayes and KNN and the results showed that Naive Bayes performed better than KNN.

We conclude with possible recommendations to extend the application to a Windows platform specifically Windows 8 smart phones and also implement a component for the percentage accuracy confidence level in the Android application.

CHAPTER 1

1 Introduction

Water is essential to all forms of life. Water is also a vital resource for agriculture, manufacturing, transportation and many other human activities. Despite its importance, less effort has been taken to ensure its safety and proper use. This has led to the use of unsafe or contaminated water by a large number of communities resulting into illnesses and sickness from the contracted water borne diseases.

1.1 Background

Drinking clean and safe water is a basic need for all human beings therefore it is a right earned by every human being to conveniently access and avail oneself safe, clean and adequate drinking water. But nevertheless this remains to be a luxury to many of the world's poor population. Our everyday life depends on clean drinking water. We can survive without food for up to 2 months, but without water, we will die within 3 days. Basing on the following facts as fully explained by (Foundation, 2005) water related diseases are really disastrous:

- 80% of all illnesses result directly from waterborne pathogens.
- 5 million people, most of them children, die each year from diarrhoea. The primary cause of diarrhoea is contaminated drinking water.
- 2.5 billion incidents of illness are caused by contaminated water every year.
- 50% of hospitalizations result from waterborne diseases.
- The leading cause of death for children under the age of five is infection from waterborne diseases.
- At any one time, approximately one billion people suffer from diseases contracted by consuming contaminated water.
- 1.2 billion people do not have access to safe drinking water. The World Health Organization predicts that by 2025, this number will increase to more than 2 billion.

Inability to drink safe water not only has health implications, but also social and economic impacts on communities which could cripple the global economic growth by 2050, with the emerging economies hit hardest (Kaestle, 2009). Waterborne diseases can have a significant impact on the economy, locally as well as internationally. People who are infected by a waterborne disease are usually confronted with related costs and not seldom with a huge financial burden. This is especially the case in less developed countries. The financial losses are mostly caused by; costs for medical treatment, costs for transport, special food, and by the loss of manpower. According to (Wikipedia, 2014), many families must even sell their land to pay for treatment in a proper hospital. On average, a family spends about 10% of the monthly households income per person infected

Furthermore, the failure to drink clean water has highly paralyzed the education system around the globe with over 150 million school-aged children with waterborne parasites suffering from resulting anemia, stunted health, and other symptoms that cause absenteeism , drop-out, decreased thinking ability, and attention deficits (Robinson, 2013).

1.2 Problem Statement

Currently, a vast population drinks clear but unsafe water deeming it as safe for drinking and this has led to a high rate of mortality more especially among young children from preventable water-related diseases. This is because people do not have a quick and reliable way of testing the quality of water so that they can be able to tell whether it is indeed safe to drink or not. Therefore, we developed a simple, easily usable and accessible mobile water testing kit that quickly tests the quality of drinking water.

1.3 Main Objective

The project aimed at building a mobile-phone based water testing kit application that assesses the quality of drinking water by analyzing the image of the water sample and giving a percentage accuracy value on how safe the water is.

1.4 Specific Objectives

1. To review current and existing relevant literature on the status and quality of drinking water and current water testing applications.

2. To design a water testing kit that assess the quality of drinking water.

3. To implement the designed system.

4. To test and validate the system.

1.5 Scope

The study focused on potable (drinking) water used by people in their homesteads delivered from public water sources like taps, boreholes, and wells. The context of the application was limited to the urban and rural communities of Uganda.

1.6 Significance/Justification

Water being the most important basic need for people, its safety must always be an assured and trusted element before human consumption. Therefore the project is significant for the following reasons:

1. The application is cheaper and more affordable for the target communities than any of the presently used instruments particularly the equipment used in testing laboratories which are not even accessible to the people facing the real problem of unknowingly consuming unsafe water.

2. The application provides a faster way of assessing the quality of drinking water as compared to the current lengthy laboratory procedures that are being used.

3. In terms of accessibility, the application is installed on the users mobile phone thus making it very easy to access in case one needs to determine the quality of a given water sample he or she intends to drink.

4. The application requires less technical requirements when assessing the water quality as compared to water quality testing instruments like the turbidity meters.

5. In addition, research studies can be conducted from the result sets of data obtained from water samples taken from the various water sources in the different regions of the country, this can help address aspects that the application has not tackled which will provide a better preventive and defensive platform against preventable water-related diseases.

CHAPTER 2

2 Literature Review

This chapter explores the available and relevant literature on the subject of water contamination and drinking of dirty water in Uganda. It defines the key terms in the literature, presents detailed realistic background information on the current status of water quality in Uganda, the existing mobile water testing applications and the techniques and scientific principles that are currently used to assess water quality.

It also provides a clear picture as to why the proposed application is necessary for heightened safety standards of potable water.

2.1 Definition of Key terms

According to a published online article, (Mapping a healthier future water, 2014), safe drinking water is water that is free from disease causing organisms, toxic chemicals, color, smell, and unpleasant taste.

Still according to (Mapping a healthier future water, 2014), in Uganda, safe drinking water is defined as water from a tap and piped water system, borehole, protected well or spring, rain water, or gravity flow schemes. Open water sources including ponds, streams, rivers, lakes, swamps, water holes, unprotected springs, shallow wells and water trucks are considered unsafe.

In reference to the (Business dictionary, 2014), safe water is potable water free from harmful microorganisms and substances, even if it may have color, odor, or taste problem due to dissolved minerals.

Drinking water or potable water is water safe enough to be consumed by humans or used with low risk of immediate or long term harm (You, 2013). In most developed countries, the water supplied to households, commerce and industry meets drinking water standards, even though only a very small proportion is actually consumed or used in food preparation.

2.2 Current status of water quality in Uganda

From an online website (water.org, 2014), it is shown that the status of water in relation to safety and sanitation has improved over the past 10 years with access to safe water supplies throughout Uganda and improved sanitation being 65% and 48% respectively. Although the number of people with access to safe water and sanitation has improved, there are still many communities (both rural and urban) that rely on contaminated water sources such as streams and open wells.

This is due to the highly degrading status of water quality in Uganda especially in Kampala, the capital city of Uganda where flooding is frequent and severe in most suburbs (New Vision, 2010). Furthermore, most people do not have access to proper toilet facilities thus large amounts of faecal waste being discharged to the public water sources and this is likely to have major impacts on infectious disease burden and quality of life (Hutton et al, 2007).

Pollutant industrial activities including food-processing and textiles are also sources of mass and heavy contaminants of drinking water delivered from public water sources. A large proportion of the population, 40-70% according to (AFDB, 2006) lives in low-income informal settlements which has also downgraded the quality of water.

From the available data in relation to the quality and status of water, there are strong indicators that the water sources for human consumption are threatened by pollutants, faecal bacteriological contamination, eroded soil materials etc.

2.3 Case Studies

2.3.1 Case Study on suburbs of Kampala city

According to (Rukia Haruna, 2005), springs are a major source of water for domestic use in the sub-urban areas of Kampala city. Though spring water is considered to be aesthetically acceptable for domestic use, presence of poorly designed pit latrines, poor solid waste management as well as poor and inadequate spring protection, may lead to contamination of spring water with pathogenic bacteria.

A study was conducted by NCBI in joint association with African Health Sciences, Makerere Medical School in 2002. The objectives of the study were to examine the bacteriological quality of water from ten springs in Katwe and Kisenyi parishes of Kampala to identify and quantify risks for spring water contamination with faecal bacteria.

A cross-sectional sanitary risk assessment using a standardized format was carried out in ten randomly selected springs in the parishes of Katwe and Kisenyi parishes in Kampala. A total of 80 samples of water from these springs were collected from December 2001 to March 2002. The samples were analyzed for indicators of faecal contamination: total coliforms, faecal coliforms and faecal streptococci. Physical-chemical parameters were also measured.

The results showed that the aggregate qualitative sanitary risk scores ranged from medium to high. The total coliform counts in 90% of the samples exceeded the WHO guideline for drinking water. All the samples had faecal coliform counts above the WHO guideline. A strong correlation was observed between the median faecal coliform counts and the sanitary risk score. Sixty percent of the samples had nitrate levels above the WHO recommended limit. Conclusively, Water from the ten protected springs studied was unsuitable for drinking without treatment.

2.3.2 Case Study on Northern Uganda by Northern Uganda Development Foundation (NUDF)

NUDF, a Canadian-based humanitarian organization that works to improve the standard of living of rural people in Northern Uganda, financed and constructed 42 wells serving over 50,000 people in Oyam District, Northern Uganda. Out of concern for the health of well users, the NUDF undertook to test the water quality of these wells.

The study focused on field sampling and laboratory testing of drinking water quality from boreholes. The study tested for the biological, chemical and physical properties of the water including pH, colour, turbidity, hardness, fecal coliforms and E. coli, the organization determined if water met Ugandas national standards.

A preliminary study conducted in July 2008 in Kamdini Parish in Oyam District (Opio, 2012), showed that water samples collected from NUDF wells had good bacteriological and satisfactory physical characteristics commensurate with Uganda's potable water standards and could be used for domestic consumption.

2.4 Existing Water Testing Applications

2.4.1 Mercury Mobile Application

This is a mobile application that detects mercury levels in water. This application operates by taking a picture of the sheet which is dipped in the water with a digital camera, and it allows the user to learn the specific concentration of the mercury. The application uses open access GIMP (GNU Image Manipulation Program) program to see the color coordinates, which they then compared with the reference values to determine the mercury concentration, this is as stated by(Technologies, 2013).

This is more applicable to those who live or work downstream from industrial and mining sites such as gold mines and the coal-fired powered plants and may want to drink water.

2.4.2 H-2-0 Mobile Water Laboratory

The H-2-0 Mobile Water Laboratory is the combination of a Windows Phone 7 smartphone, a portable microscope, a software application that resides on the mobile phone, data storage on the Windows Azure Storage and the SQL Azure, the cloud-based storage offerings from Microsoft (Parrish, 2011).

The application helps emergency responders to collect, analyze and communicate water testing data so that safe water sources can be identified. This works best for the technical people like the lab technicians, microbial analysts among others.

2.5 Weaknesses in the existing water testing methods and applications

- Most of the methods are very lengthy and time consuming like field sampling and laboratory testing.
- Applications like the mercury mobile application test for a small scope in regards to the assessment of water quality as it only detects mercury in water yet there are other water contaminants and pollutants.
- Most techniques are complex in terms of technicality and thus require user expertise like the cross-sectional sanitary risk assessment used in the case study on Kampala suburbs.

2.6 Proposed System

The proposed system will be a water testing kit composed of a mobile application running on an android platform and a 3D model kit. The system will assess the quality of drinking water by use of an image processing technique. Here a machine learning algorithm embedded in the application will process and analyze the image of the water sample taken by the phone camera.

The 3D model kit will provide for a light source that propagates light rays through the water sample in the test tube which is also slotted into the kit. The set up will borrow an idea of the turbidity meter in terms of component arrangement. The result of the analysis will be a percentage accuracy value displayed on the screen of the phone on how safe the water is.

2.6.1 Image Processing

Image processing involves performing some operations on an image in order to get an enhanced form of the image or to extract some useful information from it. Image processing basically involves analyzing and manipulating the image. It is a rapidly growing technology today in the core research area within engineering and computer science disciplines.

Application of Image Processing technique(Image feature extraction in detection of cassava mosaic)

From the development study of an Automated Vision-Based Diagnosis of Cassava Mosaic Disease conducted by (Jennifer R. Aduwo, 2010), the image processing technique was applied to a large set of images taken from the crops, both the healthy and infected, then machine learning classification algorithms were used to automatically classify the state of health and infected plants.

For this particular application, three image processing techniques were employed to obtain representative feature data from the leaf images of the health plants and from those with cassava mosaic disease. One method was based on the colour distribution of the leaves while the other two are based on the shape (image gradient information) of the leaves.

For the first dataset, a normalized histogram of the hues of pixels was obtained, taken by converting the image to HSV colour space.

For the second, SURF (Speeded Up Robust Features), a scale and rotation invariant interest point detector and descriptor to obtain representative features was used.

For the third, SIFT (Scale Invariant Feature Transformation) was used to obtain shape features corresponding to a (4 * 4) grid of histograms around each key point location. All these three methods were used to understand how classification performance changes with the use of different features.

2.6.2 Turbidity

Turbidity is an optical characteristic or property of a liquid, which in general terms describes the clarity, or haziness of the liquid. Turbidity relates to the loss of transparency due to the effect of suspended particles such as sludge, limestone, yeast, colloidal material or microorganisms. Lack of turbidity results in clarity or clearness because, it is in fact the effect of these various suspended materials on light passing through a liquid. The phenomenon of turbidity is measured by optoelectronic turbidity meters.

Principle of operation of a turbidity meter

An artificial light source emits a known intensity of light through a sample. The suspended particles scatter or absorb the light. The scattered light is then recorded on a photo detector. The scattered light is measured at an angle of 90 degrees, as shown in figure 1. This measurement principle is known as nephelometry. A nephelometer is therefore a turbidity meter that measures scattered light at an angle of 90 degrees. The results are shown in NTU (Nephelometric Turbidity Unit). To obtain defined, reproducible results, turbidity meters are calibrated and adjusted using formazine solutions (reference standard). These meters display their results in FNUs (Formazine Nephelometric Units).

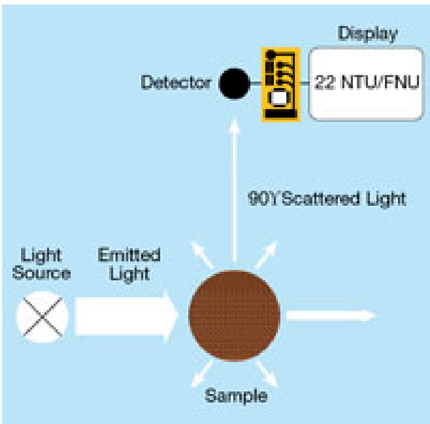

Figure 1: set up of a turbidity meter.

From the above existing water testing applications, we were able to draw a table of comparisons between them and our proposed application as shown in table 2.6.

Feature	Mercury Mobile Application	H-2-0 Mobile Water Laboratory	Proposed system
Scope	Small	Wide	Wide
Technicality	Simple	Complex	Simple
Usability	Easy	Not very easy	Easy

Table 2.6: Comparison of existing applications and the proposed system

2.7 Conclusion

According to the present data obtained from the different studies and surveys, the quality of drinking water from a bigger number of potable water sources in Uganda does not meet the required health safety standards of drinking water and it is unknowingly being used for drinking by a vast population in the country.

We propose a mobile water testing kit application that will assess the quality of drinking water as a way of ensuring that people drink safe water free from any water contaminants.

CHAPTER 3

3 Methodology

The methodology includes tools and techniques that were used in data collection, design and implementation of the system.

3.1 Data Collection Techniques

3.1.1 Questionnaire

A form of questions formulated by the research group was presented to Dr. Kiryowa Outhman, a pharmacist at Laborex Uganda Limited. The main aim was to gather technical knowledge in regard to the development and implementation of the application. Areas of concern were how turbidity of water is measured, the technicality involved, how important the turbidity element of water is when assessing the quality of water and understanding the turbidity principle well from which our application borrows an idea. This gave us vast knowledge on how well to build our application and achieve the set objectives.

3.1.2 Document Review

Document review involved reading the documentation done by several scholars, researchers and developers in our area of study. Here, the group mainly looked at written literature relating to the current status of water quality in Uganda, existing water testing applications, research areas where the image processing technique has been applied and turbidity as a core element in assessing the quality of water. This helped the group to acquire a better understanding of the project. The sources of information were online articles, research papers and academic websites.

3.1.3 Mobile phones

Mobile phones were used for image capture of the water samples to form the training and test data set.

3.2 System Design Tools

3.2.1 SketchUp Pro 2013 and Cinema 4D

These are modeling design softwares that were used to model and design the 3D kit.

3.2.2 Adobe Photoshop CS 6

Adobe Photoshop CS 6 is a graphics editing program that was used to design the splash screen of the application user interface.

3.3 System Implementation Tools

This section discusses the various tools, methods and software that that were used to implement the system

3.3.1 Python

Python was the programming language used during the prototyping phase of the application.

3.3.2 Numpy

This is a python library that was used to deal with the N-dimensional array objects.

3.3.3 Matplotlib

This is a plotting library for python that was used for plotting purposes like in image histograms.

3.3.4 OpenCV in Python

OpenCV was used because it supported the machine learning algorithms that were be implemented in the code application for example KNN,Naive Bayes.

3.3.5 Java and OpenCV in Java

Java was used in the code implementation of the final application on an android platform. OpenCV in Java was used to provide for the machine learning algorithms for image classification.

3.3.6 3D kit

This is a solid object that was printed using the 3D printing technology, as shown below in figure 2. It was modelled in a cuboidal form and it has provision for a light source, a hole for the phone camera, a holder for the water to be tested, a toggle switch and a pair of batteries.

Image of the kit

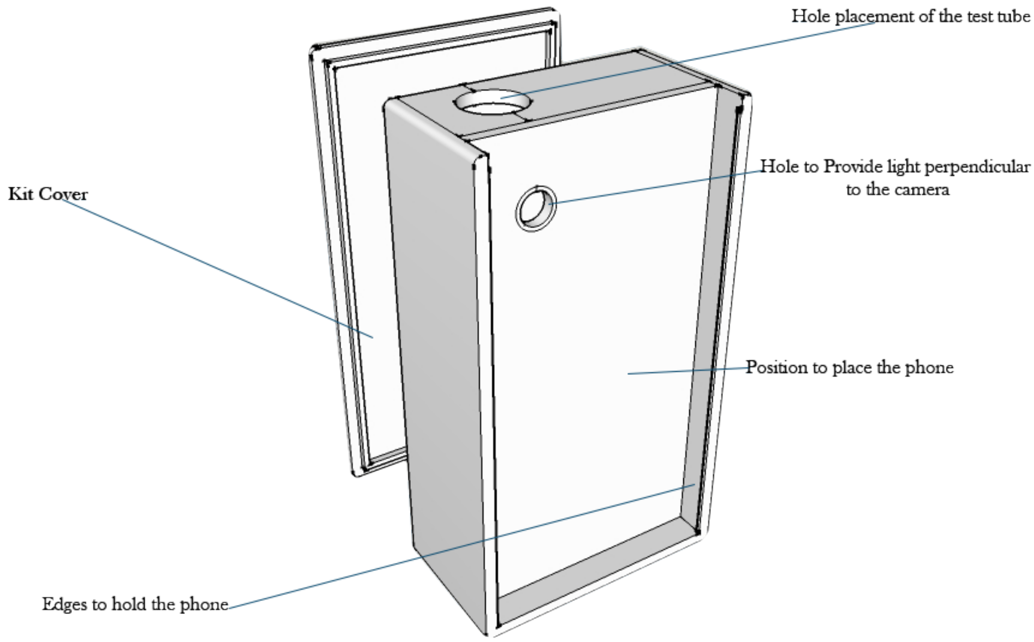

Figure 2: image of the modelled 3D kit

CHAPTER 4

4 System Design and Implementation

This section describes the processes, software and technologies that were used in designing and implementing the system software and hardware.

4.1 System Design

System design mainly entailed the design of the 3D kit and the splash screen of the application.

4.1.1 Design of the 3D kit

The kit was designed using SketchUp Pro 2013, a modeling program that enables the designer to build the model with custom outlooks and attributes. SketchUp Pro is like a pencil with superpowers, starts by drawing lines and shapes, push and pull surfaces to turn them into 3D forms. Also Cinema 4D, a modeling program, was used and it is where the modeled design was imported because it has a better rendering engine than SketchUp Pro 2013.

4.1.2 Design of the Splash Screen

Here Adobe Photoshop CS 6, a graphics editing program, was used to design the words that appear on the screen as illustrated below in figure 15

4.2 Algorithm Implementation

Under this phase, processes like image capture, image preprocessing, histogram extraction, feature extraction and machine learning were carried out.

4.2.1 Image Capture

Images were captured by use of a mobile phone camera through the designed interface of taking pictures as shown below in figure 16. The captured images were then labelled and categorized into clean and dirty image data sets and this formed our training data set. For the clean class, we included boiled water, spring water, rain water and bottled water as shown in figures 3, 4, 5 and 6 respectively. For the dirty class, we had very dirty/muddy water, slightly dirty water and mixed samples of the clean class with sugar, glucose, salt, sugar and glucose, sugar and salt and sugar, glucose and salt as illustrated down in figure 7 and 8 for image samples of boiled water with glucose and spring water with glucose respectively.

4.2.2 Image samples of the Clean class

Boiled water

Figure 3: image of the boiled water sample.

Spring water

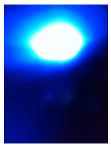

Figure 4: image of the spring water sample.

Rain water

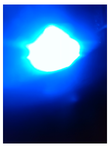

Figure 5: image of the rain water sample.

Bottled water

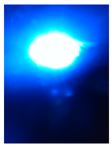

Figure 6: image of the bottled water sample.

4.2.3 Image samples of the Dirty class

Boiled water with glucose

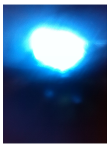

Figure 7: image of the boiled water with glucose.

Spring water with glucose

Figure 8: image of the spring water with glucose.

4.2.4 Image preprocessing

Image preprocessing was carried out to modify the image in a way that makes it more suitable for further analysis and extraction of its features. Here, the captured digital colored image was implicitly converted into a bitmap since Java deals with images in form of bitmaps, (although no implicit conversion occurred in python for the system prototype). A bitmap is an image stored as a series of pixels where each pixel is assigned a color and then arranged in a pattern to form the image.

The bitmap image was then converted into a mat array object in form a vector that stores the image components as an ordered sequence of numbers in rows and columns.

The image was finally transformed into a gray scale image, one having many shades of gray with pixel values ranging from 0 to 255. This was done because for a gray scale image, less information needs to be provided for each pixel.

4.2.5 Histogram Extraction

An image histogram is a graph or plot, which gives an overall idea about the intensity distribution of an image. It is a plot with pixel value, ranging from 0 to 255, for the 8-bit grayscale image on X-axis and corresponding number of pixels (frequency) in the image on Y-axis.

Here we plotted intensity histograms as shown in the figures 9, 10, 11 and 12 below. For each histogram, we used 50 bins, each bin containing a feature that can be extracted from the histogram. From the computed histograms, we were able to extract image features for the training dataset.

INTENSITY HISTOGRAMS FOR THE GRAY SCALE IMAGES

Boiled water

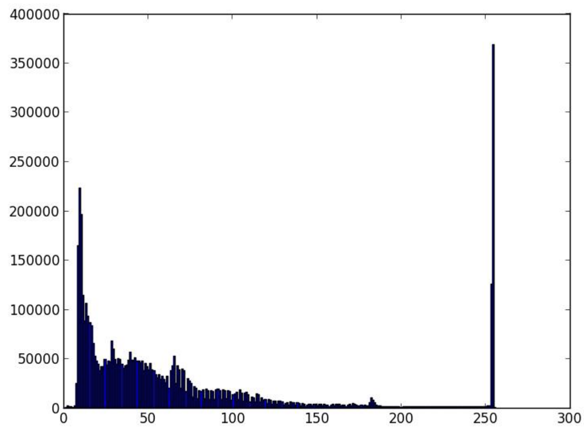

Figure 9: intensity histogram for boiled water.

Bottled with glucose and salt

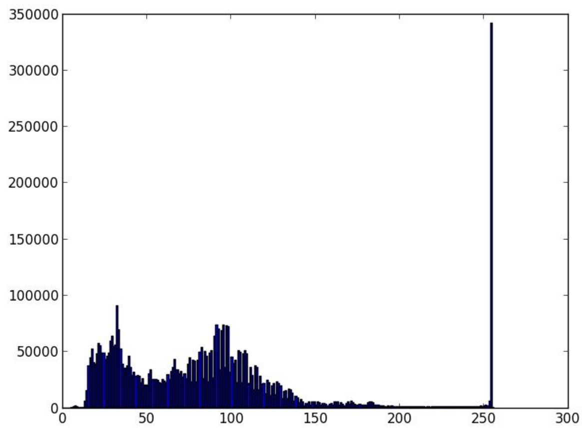

Figure 10: intensity histogram for bottled water with glucose and salt.

Very dirty water

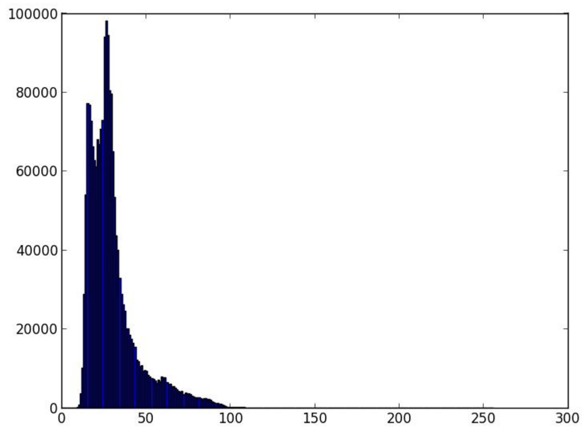

Figure 11: intensity histogram for very dirty water.

Spring water

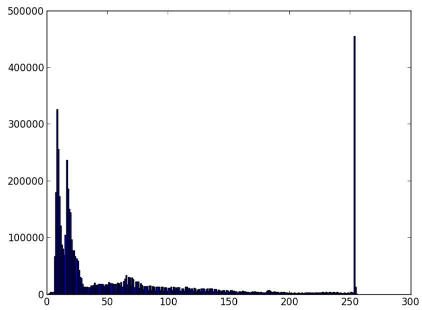

Figure 12: intensity histogram for spri
water.

Key
Y-axis = Number of pixels (Frequency)
X-axis = pixel values

4.2.6 Feature Extraction

Feature extraction is the process of transforming of input data into a set of features. This enabled us to generate a data set to be used for training. Here, we extracted the number of pixel values, the frequency count for each bin of the intensity histogram by performing an OpenCV function **calcHist**.

This provided us with a feature set of numbers representing the relevant values for our training data set.

4.2.7 Machine Learning

Machine learning is a field of artificial intelligence that provides computers or system applications with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can learn from data and change when exposed to new data. For this particular project, the ultimate goal was image classification of water samples into either the clean or dirty class.

Image classification involves analysing the numerical properties of the extracted image features and organise them into classes. This process involved two phases: training and testing of the image classifiers.

Under the training phase, the image classifiers were trained using the training data set, the clean and dirty classes. Here we trained each classifier on image samples from either classes so that it learns the different descriptive features in each sample. This aimed at perfecting the accuracy and reduce on the error rate for each classifier.

Under the testing phase, we ran the algorithms on test data to verify that each algorithm had been trained well and had actually learned. The result was then displayed on the phone screen as shown below in figure 17.

Machine learning algorithms that were trained and tested on include: Naive Bayes, a probabilistic classifier, KNN and SVM both deterministic image classifiers. Training and testing of the KNN and Naive Bayes image classifiers was done for the application prototype in Python. In Java, for the final application, only Naive Bayes and SVM were implemented.

KNN Algorithm

KNN is a classification algorithm used under supervised learning. The idea is to search for closest match of the test data in a feature space. Here, if a sample point has features similar to the ones of points of a particular class, then it belongs to that class. These points are known as nearest neighbors.

The algorithm also involves a parameter **k** that specifies the number of neighbors (neighboring points) used to classify one particular sample point. Finally, the assignment of a sample to a particular class is done by having the k neighbors considered to be legal. In this fashion, the class represented by the largest number of points among the neighbors ought to be the class that the sample belongs.

Steps of the KNN algorithm

- Determine the parameter **k**, number of nearest neighbors.
- Calculate the distance between the test sample and all the training samples.
- Determine the nearest neighbors based on the minimum distance of **k**.
- Gather the class or classes of the nearest neighbors.
- Use simple majority of the class of the nearest neighbors as the prediction of the test sample.

Naive Bayes Algorithm

Naive Bayes Classifier is a probabilistic classifier based on the Bayes theorem with independence assumptions between predictors. Bayes assigns a posterior probability to a class based on its prior probability and its likelihood given the training data.

Algorithm

Bayes theorem provides a way of calculating the posterior probability, P($c \mid x$), from P(c), P(x)and P($x \mid c$). Naive Bayes classifier assumes that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

$$\mathrm{P}(c \mid x) = \mathrm{P}(x \mid c)\mathrm{P(c)}/\mathrm{P(x)}$$

P($c \mid x$) is the posterior probability of class (target) given predictor (attribute).
P(c) is the prior probability of class.
P($x \mid c$) is the likelihood which is the probability of a predictor given class.
P(x) is the prior probability of a predictor.
Prior probabilities are based on previous experience.

SVM Algorithm

A support vector machine is a supervised machine learning algorithm used for data classification and estimating the relationships between variables. It is a supervised algorithm because there is an initial training phase involved where you feed the algorithm data that has already been classified (labeled). After this initial training phase is completed, future data sets given to the algorithm can be classified with no or minimal human intervention.

SVM supports both linear classification and non-linear classification using the kernel method. The kernel method can pull data points apart into 3-dimensional space instead of using a line as a separator. Kernel functions implicitly define a metric feature space for processing the input data. Refer to figure 13 for diagrammatic representation of the algorithm flow.

Diagrammatic representation of the algorithm flow

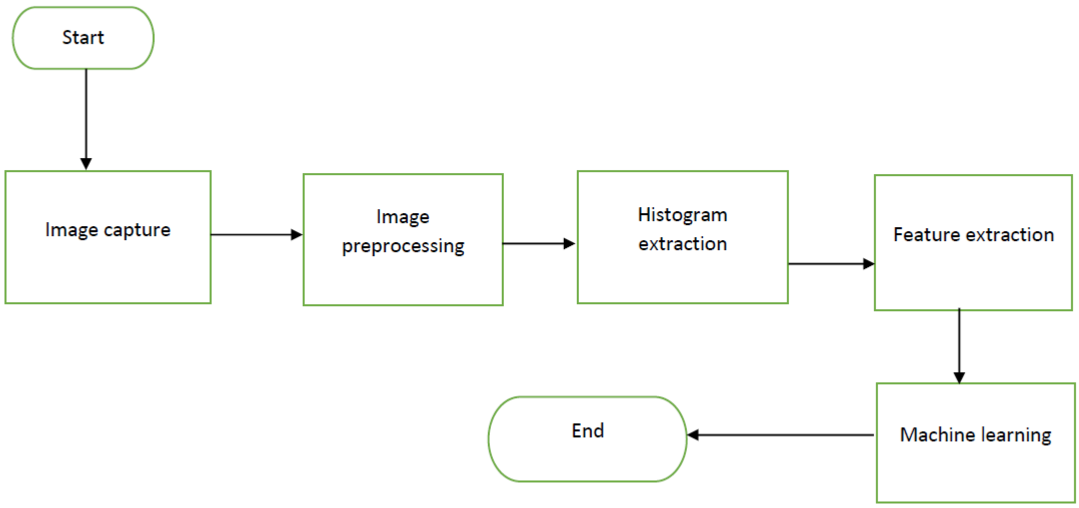

Figure 13: image showing the algorithm flow.

MENU SCREEN SHOWING THE APPLICATION ICON

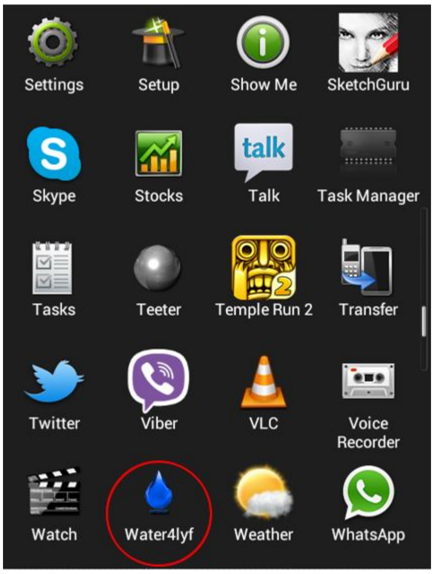

Figure 14: showing the circled application icon on the menu screen.

SCREEN SHOT FOR THE SPLASH SCREEN

Figure 15: the splash screen interface.

SCREEN SHOT FOR THE TAKE PICTURE INTERFACE

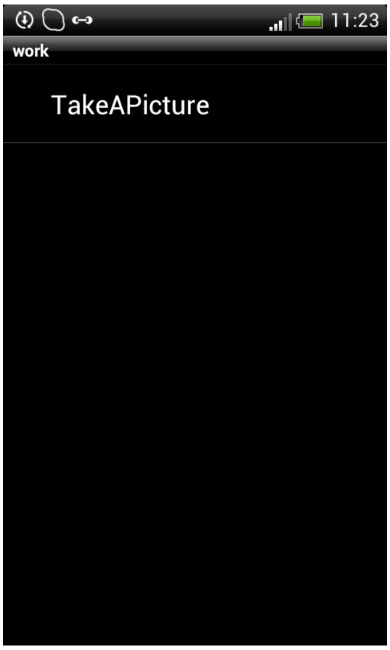

Figure 16: the take picture interface of the phone.

SCREEN SHOT FOR THE RESULTS INTERFACE

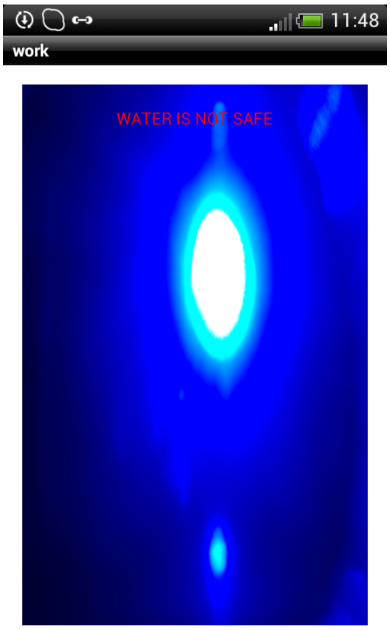

Figure 17: the results displayed on the phone screen.

4.3 System Hardware Implementation

4.3.1 3D Printing

This is the technology that was used to print the 3D kit. 3D printing or additive manufacturing is a process of making a three-dimensional solid object of virtually any shape from a digital model. 3D printing is achieved using an additive process, where successive layers of material are laid down in different shapes until the entire object is created. 3D printing is usually performed by a materials printer using digital technology.

How it works

- One designs a virtual object for what he or she intends to create. For our application, we designed the virtual object for the kit we wanted to create. The virtual object was made using a CAD(Computer Aided Design) file using 3D modeling programs, SketchUp Pro 2013 and Cinema 4D R15.
- To prepare the digital file created in a 3D modeling program for printing, the software slices the final model into hundreds or thousands of horizontal layers and exported as an STL also known as Standard Tessellation Language file

format. When this prepared file is uploaded in the 3D printer, the printer creates the object layer by layer.

- The 3D printer reads every slice (or 2D image) and proceeds to create the object blending each layer together with no sign of the layering visible, resulting in one three dimensional object. At this step, the 3D kit was the output.

4.3.2 Hardware Components in the 3D Kit

The printed kit was modelled to provide for the following components that were to be incorporated into it:

- **Battery**
  A 9 Volts battery also known as the transistor battery was used in the circuit to provide voltage within the circuit. It has a rectangular prism shape with rounded edges and a polarized snap connector at the top.

- **LED bulb**
  A LED bulb also known as a Light Emitting Diode bulb. It is a miniature light engine that runs on electrons. LED bulbs do not easily burn out or get hot as compared to other ordinary incandescent bulbs. They also produce more light yet they cost cheap to purchase and so these were used as the source of light.

- **Resistor**
  A resistor was used to reduce on the current flow and at the same time to lower voltage levels within the circuit. The current though a resistor is in direct proportion to the voltage across the resistor's terminals.

- **Toggle switch**
  A toggle switch is a switch that uses a toggle joint with a spring to open or close an electric circuit as an attached lever is pushed through a small arc. This was used to turn the LED bulb on and off.

- **Connecting Wires**
  These are electrical connectors that are made out of copper metals with electrical terminals of positive and negative that are used to complete the circuit. These were used to connect the batteries, toggle switch and the LED bulb. Below, figures 18 and 19 illustrate the components of the kit.

Images of the kit

Figure 18: image of the 3D kit showing the inner components.

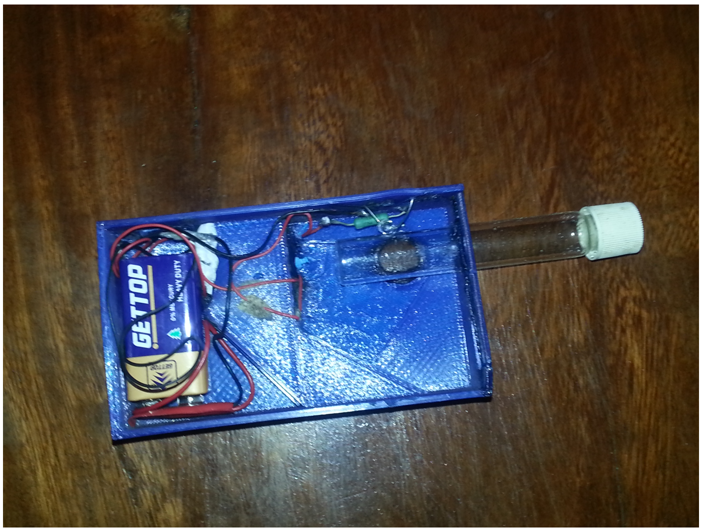

Figure 19: image of the 3D kit with the test tube.

CHAPTER 5

5 System Testing and Evaluation

Testing was done aiming at identifying and rectifying any errors in the design as well as determining whether the system in form of components and as a whole met the required specifications. Image classifier evaluation was also carried out in order to find out which classifier performed best.

5.1 System Hardware Testing

Here testing was done on the 3D kit components. The types of testing that were applied include:

- Unit testing: This was conducted to test and determine how efficiently each unit to be incorporated into the kit worked prior to system integration. Here we tested the batteries to ensure that they provided enough voltage for the LED lights to give the required light intensity plus we also tried different forms of LED lights as way of trying to find the most appropriate one for the required light intensity.

- Integration testing: This involved testing how well the system as a whole worked. Here we connected all the components to be incorporated into the kit; batteries, toggle switch, resistor and LED lights to ensure that they could work as expected when fit together.

- Usability testing: This was carried out to determine how convenient the system was for the user in terms of usability for example observing how users take images of the water samples, is it easy for them or hard.

5.2 Image Classifier Evaluation

Here we measured and assessed the performance of the classifiers that had been used. Performance was based on accuracy, sensitivity and specificity for each classifier, this was done with the aid of a confusion matrix as shown in figure 20 below. A confusion matrix contains information about actual and predicted classifications done by a classification system. The matrix holds values for the true positive, false positive, true negative and false negative as shown in the confusion matrices for KNN and Naive Bayes in the tables 5.2 and 5.2b below. These enabled us to calculate the accuracy, sensitivity and specificity.

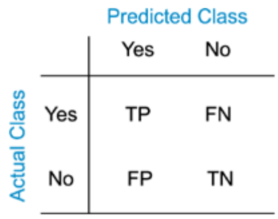

Figure 20: confusion matrix structure.

	Yes	No
Yes	5	3
No	3	29

Table 5.2: KNN confusion matrix

	Yes	No
Yes	0	8
No	0	32

Table 5.2: Naive Bayes confusion matrix

- True positives (TP) : This is the proportion of the clean class samples that were correctly identified as clean.
- False positives (FP): This is the proportion of dirty class samples that were incorrectly classified as clean.
- True negatives (TN): This is defined as the proportion of negative cases that were classified correctly, dirty class samples classified as dirty.
- False negatives (FN): This is the proportion of positive cases that were incorrectly classified as negative, clean class samples classified as dirty.

Accuracy

K-Nearest Neighbor (KNN)
Accuracy = (TP + TN)/n, where n is the number of test samples.
Accuracy = (5+29)/40 = 0.85
= 85%
Naive Bayes
Accuracy = (TP + TN)/n, where n is the number of test samples.
Accuracy = (0+32)/40 = 0.8 = 80%

Sensitivity

This is a statistical measure of how well a binary classification test correctly identifies a condition. It is the probability of correctly labeling members of the target class.

K-Nearest Neighbor (KNN)
Sensitivity = TP/(TP + FN)
Sensitivity = 5/(5+3) = 0.625
= 62.5%

Naive Bayes
Sensitivity = TP/(TP +FN)
Sensitivity = 0/(0+8) = 0
= 0%

Specificity

The specificity is a statistical measure of how well a binary classification test correctly identifies the negative cases. It uses the false positive rate or false alarm rate.

K-Nearest Neighbor (KNN)
False positive rate = FP/(TN + FP)
False positive rate = 3/(29 + 3) = 0.09375
Specificity = 1 - False positive rate
Specificity = 1 - 0.09375
= 0.90625

Naive Bayes
False positive rate = FP/(TN + FP)
False positive rate = 0/(32 + 0) = 0
Specificity = 1 - False positive rate
Specificity = 1 - 0
= 1

Performance of the classifiers was evaluated basing on accuracy, sensitivity and specificity. But, considering the type of application that had been developed, we chose specificity as the main evaluation factor because we preferred that always dirty or unsafe water samples be classified as unsafe for drinking. From the computed values of specificity, Naive Bayes emerged as a better classifier than KNN.

CHAPTER 6

6 Summary, Challenges, Conclusion, Recommendations and Future Work

This chapter summarizes the project and also covers the challenges faced, conclusion on the work done, the recommendations and the planned future work.

6.1 Summary

Currently, a significant number of people in different communities are suffering from water-borne diseases facing an unbearable and high cost in seeking proper medical care and attention, many dying with an enormous number being children. Evidently, this is very devastating and urgent attention must be given to all possible solutions to mitigate this grievous condition in our communities. Therefore our mobile water testing kit application aims at reducing on the alarming count of people dying from preventable water-borne diseases by empowering the general community that has been hit hard with an application that assess the quality of drinking water. The application will enable people to drink water that is safe for their lives. This will in turn reduce on the huge numbers of people dying due to the consumption of unsafe water.

6.2 Challenges

We faced a major challenge in obtaining a large and required training data set for the purpose of training of the image classifiers and this highly affected the performance of the classifier since we did not have a balanced training data set.

We also failed to implement the SVM algorithm in Python during prototyping and this because SVM required a library, libsvm that failed to work with numpy that we had already implemented.

We also failed to acquire a fine model of the 3D kit in time because 3D printing is not a very popular technology in Uganda and the use of the improperly modeled kit highly affected the quality of our images.

We were also not able to implement the percentage accuracy confidence level in regard to the safety of water in the Java application.

6.3 Conclusion

The study focused on developing and implementing a mobile water testing kit application with its major goal being quality assessment of drinking water. With all the multiple cases of deaths due to water-borne diseases, the application will play a major role in fighting this growing nemesis. The application will help eradicate a serious burden affecting a vast population in various communities.

6.4 Recommendations

Future researchers who intend to extend the functionality of the application should focus on mapping and plotting of areas with safe drinking and those with unsafe drinking water as this can be helpful for the various programs conducted at the regional and possibly country level by the various health institutions of the government and NGOs.

The government of Uganda should promote and carry out information technology training at the community level so that people dwelling in such communities can easily adapt to modern mobile technologies like the use of smart phones on which applications like the mobile water testing kit application run.

6.5 Future Work

We plan to extend the application to a Windows platform specifically Windows 8 smart phones and also implement a component for the percentage accuracy confidence level in the Android application.

7 References

Robinson, B. (2013). The dangers of dirty water. Retrieved from http://www.weathermatic.com/blog/dangers-dirty-water.

Rukia Haruna, F. E. (2005, March). The quality of water from protected springs in Katwe and Kisenyi parishes, Kampala city, Uganda. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1831893/.

Jennifer R. Aduwo, E. M. (2010). Automated Vision-Based Diagnosis of Cassava Mosaic Disease. Industrial Conference on Data Mining - Workshops, 114-122.

Foundation, K. M. (2005). Water Intervention Projects. Retrieved from http://www.kalonzomusyokafoundation.org/projects/projects.aspx?ProID=3.

Kaestle, J. (2009, October). Access to Safe Drinking Water and Its Impact on Global Economic ... Retrieved from http://faculty.washington.edu/categ/healthanddevgbf/wordpress /wp-content/uploads/2010/03/Access-to-Safe-Drinking-Water.pdf

Parrish, D. (2011, March 5). The H-2-O Mobile Water Testing Application. Retrieved from http://jparrishucf.wordpress.com/2011/03/05/183/

You, D. J. (2013, August). Special Issue on Drinking Water and Health. Retrieved from scirp: http://www.scirp.org/Journal/jwarp

The Water & Sanitation Crisis.(2014). Retrieved from http://water.org/country/uganda

Technologies, T. S. (2013, February 14). The Water Network, Use your cell phone to detect mercury levels in water. Retrieved from https://water.tallyfox.com/mosaic/text/use-your-cell-phone-detect-mercury-levels-water.

New Vision (2010) Floods hit Kampala City Suburbs. News report 9 May 2010. http://allafrica.com/stories/201005100171.html

Mapping a Healthier Future. (2014). Safe Drinking Water Coverage and Poverty (pp. 17 - 28). Retrieved from http://www.wri.org/sites/default/files/pdf/mapping_a_healthier_future_front.pdf.

Safe Water. (2014). Retrieved from
http://www.businessdictionary.com/definition/safe-water.html

Waterborne diseases: Socioeconomic impact (2014, May 6).
Retrieved from http://en.wikipedia.org/wiki/Waterborne_diseases.

Opio, C. (2012). Building Effective Drinking Water Management Policies in Rural Africa: Lessons learnt from Northern Uganda.
Retrieved from
http://www.africaportal.org/sites/default/files/Africa%20Initiative%20Policy%20Brief
%20No.%205_0.pdf

Hutton G, Haller L & Bartram J. (2007). Economic and health effects of increasing coverage of low cost household drinking-water supply and sanitation interventions to countries off-track to meet MDG target 10. Geneva, Switzerland, World Health Organization. http://www.irc.nl/page/38443.

AfDB (2006) Appraisal Report: Implementation of an Integrated Project of Water & Sanitation Services for the Urban Poor in Kagugube Parish, Kampala. Retrieved from http://www.africanwaterfacility.org/fileadmin/uploads/awf/projects-activities/APPRAISAL
%20NWSC%2021.DEC.2006_0.PDF.

8 Appendices

8.1 Appendix A

Questionnaire

Name of the respondent :______________________________

Title position :____

Place of Work :_______________________________________

QUESTIONS

1. What do you understand by the term safe water?

2. Quality is quite a general term; one can define it as a characteristic property that defines the apparent nature of something. What elements are considered when assessing the quality of water?

3. What is turbidity?

4. How efficient is turbidity as a sole factor used in assessing the quality of water?

5. What technicalities are involved in using turbidity as a metric for assessing water quality?

6. What is your opinion about having a water mobile testing application kit that borrows an idea of the principle of a turbidity meter?

7. How best do you think this system can be implemented?

THANKS A LOT!

8.2 Appendix B

QUESTIONAIRE.

Name of the interviewee: Kirtowa Sithuman

Title position: Pharmacist

Place of Work: Laborex (U) Ltd.

QUESTIONS

1. What do you understand by the term safe water?
   - Water free of impurities
2. Quality is quite a general term; one can define it as a characteristic property that define the apparent nature of something. What elements are considered when assessing the quality of water?
   - Colour; Clarity; taste; turbidity; density, Boiling pt, freezing pt.
3. What is turbidity?
   - degree of absorbance of light in a sample.
4. How efficient is turbidity as a sole factor used in assessing the quality of water?
   - Turbidity is efficient but absolute.
5. What technicalities are involved in using turbidity as a metric for assessing water quality?
   * photometric instrumentation.
6. What is your opinion about having a water mobile testing application kit that borrows an idea of the principle of a turbidity meter?
   - A great idea. It is efficient; cheap; reliable; data is real time (saves time removing)
7. How best do you think this system can be implemented?
   * The tool can be best be implemented through village health teams (District health teams) ~~...~~ and partnerships with NGO.

THANKS A LOT!

8.3 Appendix C

IMAGES OF THE KIT

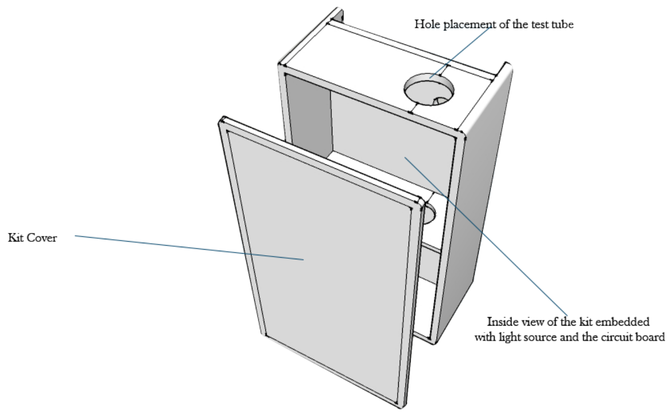

Figure 21: image showing the lateral view of the kit.

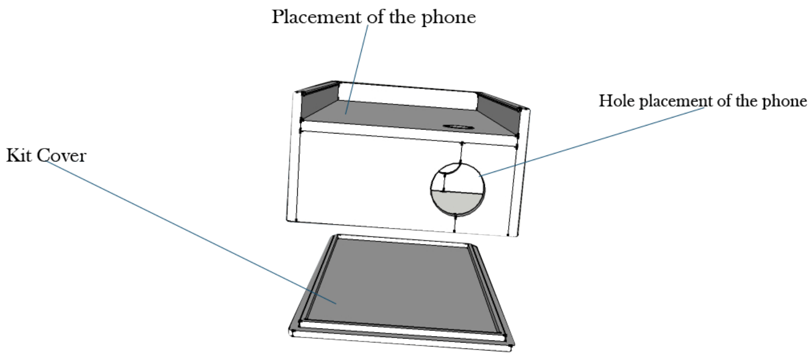

Figure 22: image showing the upper view of the kit.

8.4 Appendix D

Algorithm Code Samples

Loading of the image data set

```
# get images from the different folders

trainingDirtyImages = getImages(trainingDirtyDataPath)
trainingCleanImages = getImages(trainingCleanDataPath)
testDirtyImages = getImages(testDirtyDataPath)
testCleanImages = getImages(testCleanDataPath)
```

Attaching Class Labels

```
# define classes for the data, Clean - class 1, Dirty - class 0

trainingDirtyClasses = np.zeros((len(trainingDirtyData),1)).astype(np.float32)
trainingCleanClasses = np.ones((len(trainingCleanData),1)).astype(np.float32)
testDirtyClasses = np.zeros((len(testDirtyData),1)).astype(np.float32)
testCleanClasses = np.ones((len(testCleanData),1)).astype(np.float32)
```

Feature Extraction

```
# extract data from the images by calculating their histograms

trainingDirtyData = calculateHistograms(trainingDirtyImages, bins).astype(np.float32)
trainingCleanData = calculateHistograms(trainingCleanImages, bins).astype(np.float32)
testDirtyData = calculateHistograms(testDirtyImages, bins).astype(np.float32)
testCleanData = calculateHistograms(testCleanImages, bins).astype(np.float32)
```

Naive Bayes algorithm flow

```
def naiveBayes(trainingData, trainingClasses, testData, testClasses):
nbClassifier = cv2.NormalBayesClassifier()
nbClassifier.train(trainingData, trainingClasses)
ret, results = nbClassifier.predict(testData)
print results
return calculateAccuracy(testClasses, results)
```

KNN algorithm flow

```
def kNN(trainingData, trainingClasses, testData, testClasses):
kNNClassifier = cv2.KNearest()
kNNClassifier.train(trainingData, trainingClasses)
ret, results, neighbours, dist = kNNClassifier.findnearest(testData, 3)
print results
return calculateAccuracy(testClasses, results)
```