# Food Image Processing for a Semi-Automatic Nutrient Intake Monitoring System (NIMS)

By

**Gregorio Villalobos**

A thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the degree of

Master's in Computer Science

Ottawa-Carleton Institute for Computer Science
School of Information Technology and Engineering
The University of Ottawa

Ottawa, Ontario, Canada

# Abstract

According to the World Health Organization (WHO) statistics, obesity has reached epidemic proportions, with over 1.5 billion adults suffering from overweight reported in 2008. As obesity and overweight are major risks to human health, obesity treatment has been the focus of a large number of recent studies. Obesity treatment requires constant monitoring of the patient's diet. The smart technologies of today's intelligent environment can be used for the development of appropriate monitoring systems for obesity treatment. In this thesis, we propose a smart system that takes advantage of smartphones to build a platform for monitoring the caloric intake of obesity patients. The patient uses the built-in camera of the smartphone to take a picture of any food that he/she wants to eat. The system then processes the images to detect the type of food and portion size, and uses the information to estimate the number of calories in the food choice. The use of gradient calculation and color rasterization to increase the capacity of the system to recognize the edges of the objects is applied inside the application to signalize the contours of each portion and perform the analysis. As part of our application, we use the thumb of the patient inside the picture as a measurement pattern, to perform the thumb recognition, a skin color algorithm is applied to locate the thumb and measure its size in pixels; the size extracted is then used to translate the food portions into real life size over the entire image. Finally with the portions separated and the computed sizes, a set of nutritional facts are applied in order to generate the amount of calories present in the picture.

# Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisors; to Professor Nicolas D. Georganas for opening the door for me to this great institution, and accepting me into his research group, and to Professor Shervin Shirmohammadi, for assigning me this exciting project that is intended to help persons and health systems. Without their support, the completion of this thesis would not have been possible. I am thankful for their efforts to help me in spite of their busy schedules.

I had the opportunity to work as part of a great team during the course of my thesis. The persons involved in the same project gave me a better perspective and new knowledge. I would like to thank all my colleagues in the DISCOVER lab, especially Dr. Behnoosh Hariri, for all her guidance and help; PhD candidate Mauricio Vines, for all his support in the completion of the courses; Rana Almaghrabi, for all her support in the nutrition calculations; and Parisa Pouladzadeh for all her support, good ideas and prompt answers during the work.

Finally, I would like to thank my family for the invaluable support they gave me, for all the understanding received by my son Andres Villalobos, and to my wife Ana Carballo, for always being there for me during the entire process of my studies.

# Table of Contents

# List of Figures

# List of Tables

# Glossary of Terms

| | |
|---|---|
| 2D | Two Dimensional Space |
| 3D | Three Dimensional Space |
| ACID | Affine Cell Image Decomposition |
| BLOB | Binary Large Object |
| BMI | Body Mass Index |
| CIE | International Commission on Illumination |
| CRT | Cathode Ray Tube |
| DEXA | Dual Energy X-ray Absorptiometry |
| DLW | Doubly Labelled Water |
| FFQ | Food Frequency Questionnaire |
| HLS | Hue, Lightness and Saturation |
| HSI | Hue, Saturation and Intensity |
| HSV | Hue, Saturation and Value |
| L-a-b Model | Lightness and distribution dimensional colour distribution over a and b |
| LCD | Liquid Crystal Display |
| MRI | Magnetic Resonance Imaging |
| NIMS | Nutrient Intake Monitoring System |
| PDE | Partial Differential Equation |
| RFID | Radio Frequency Identification |
| RGB | Red, Green and Blue Colour Space |

| | |
|---|---|
| RGBA | Red, Green, Blue and Alpha Colour Space |
| ROI | Region of Interest |
| SDK | Software Development Kit |
| SVM | Support Vector machine |
| WHO | World Health Organization |
| YCrCb | Digital Video signal for Luminance and dimensional colour distribution over Cr and Cb |
| YIQ | Analog Video signal for Luminance and dimensional colour distribution over I and Q |
| YUV | Analog Video signal for Luminance and dimensional colour distribution over U and V |

# Chapter 1 Introduction

## 1.1 Motivation

During the past few decades, we have witnessed an increasing problem of overweight and obesity in the population of many countries around the world. [1] The incidence of obesity has doubled globally since 1980. In 2008, the World Health Organization (WHO) reported over 1.5 billion adults as overweight and more in-depth study indicated over 200 million men and 300 million women were obese. The problem is increasing steadily, and is also now affecting children, with a reported 43 million children under the age of 5 being overweight. Overweight and obesity is the fifth leading risk globally for death, which means that, per year, more than 2.5 million adults die as a result of being overweight or obese.

The World Health Organization (WHO) defines obesity as an accumulation of an abnormal excess of fat [2] in the body of the person, which represents a threat to the well being of the person. The measurement used by WHO is the body mass index (BMI) because of its acceptance and the easy calculations involved. Based on this method, the definition from the WHO states that the division of the person's weight (in Kilograms) between the square of his/her height (in meters) will produce a rough calculation of the BMI related to this specific person. A more technical definition for obesity is given by G. A. Bray [3], who states that "Obesity is a disease whose pathology lies in the increased size and number of fat cells. An anatomic classification of obesity from which a pathologic classification arises is based on the number of adipocytes, on the regional distribution of body fat, or on the characteristics of localized fat deposits". In order to perform a classification using the

WHO approach, and with the number of BMI in hand, we can state that if the BMI number is between 25 and 29, the person is considered overweight and if the BMI number is 30 or higher, the person is considered obese.

Table 1: Classification of Overweight and Obesity

| | BMI (kg/m2) | Obesity Class | Men < 102 cm<br>Women < 88 cm | > 102 cm<br>> 88 cm |
|---|---|---|---|---|
| Underweight | < 18.5 | - | - | - |
| Normal | 18.5-24.9 | - | - | - |
| Overweight | 25.0-29.9 | - | Increased | High |
| Obesity 1 | 30.0-34.9 | 1 | High | Very High |
| Obesity 2 | 35.0-39.9 | 2 | Very High | Very High |
| Extreme | ≥ 40.0 | 3 | Extremely High | Extremely High |

Table 1 shows the complete detail of the BMI analysis for the complete categorization of the person analyzed. The classification is done over the entire spectrum of BMI possible values obtained from the calculations.

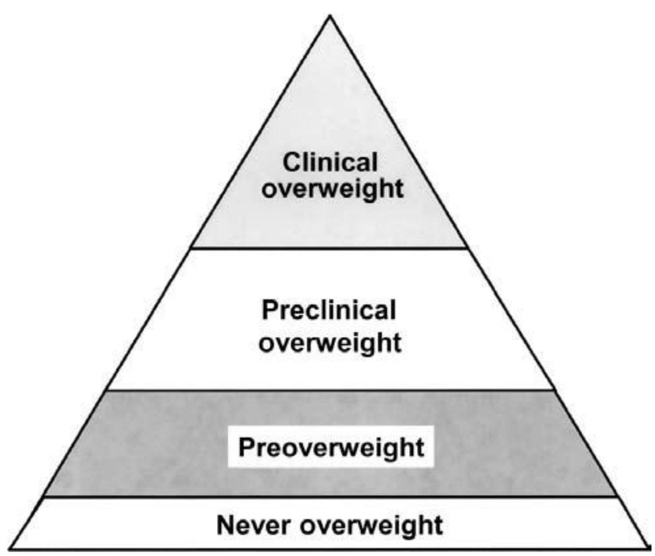



Figure 1: Pyramid of overweight

Figure 1 shows the pyramid of overweight classification. The image shows a number of interesting facts; first is that the persons who suffer pre-overweight are prone to become overweight very easily. Two classifications of overweight are also identified: preclinical and clinical overweight, where this separation takes into account whether the overweight patient has another clinical indication such as diabetes, hypertension, or other condition related to overweight.

The increasing number of obesity cases throughout the world is alarming, and serious health consequences are related to obesity and overweight. Diseases like type 2 diabetes, colon cancer, breast cancer, heart disease, stroke and musculoskeletal damage are just a few of the health conditions associated with overweight. The health care systems of nations like Canada are naturally concerned with this disease as it produces a high index of expenses in treatments, medication and hospitalizations. Figure 2 shows the top 30 nations in the world that have an Adult population that suffers from some sort of overweight.

The treatment of the overweight and obesity can involve several approaches that consider technology as a tool to aid the dietician and the patient in the process. With this in mind, the digital environment is being developed to perform this type of task and systems are regularly designed to increase the accuracy of monitoring calorie consumption.

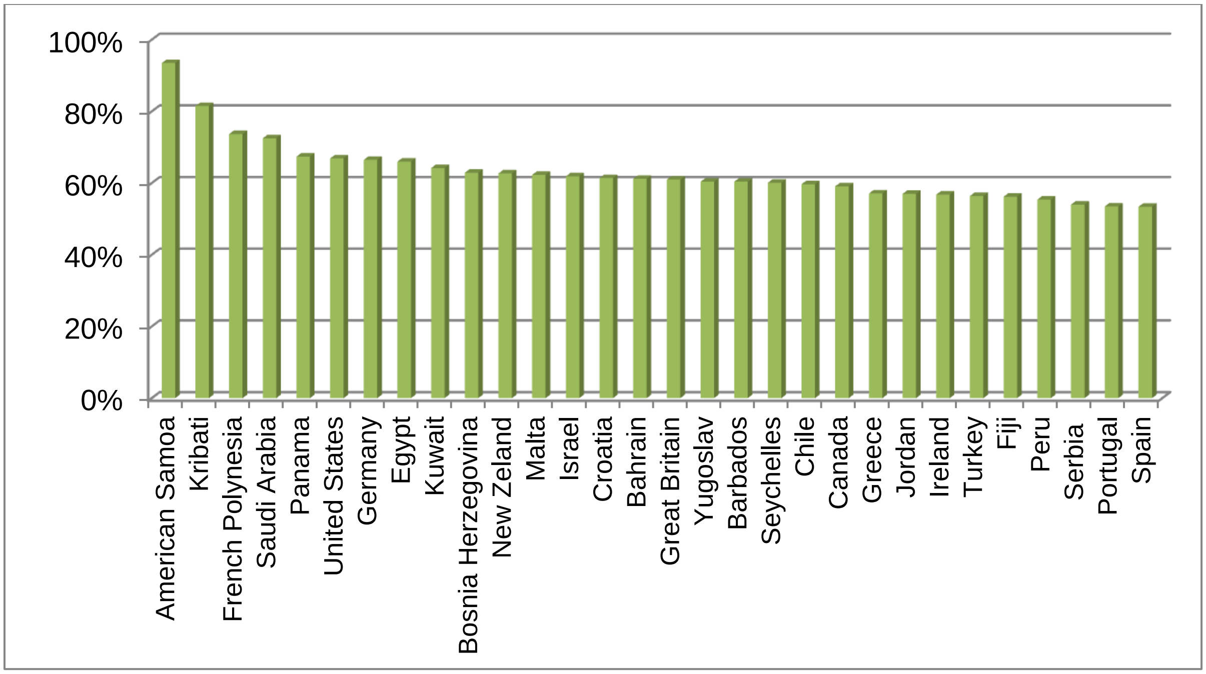



Figure 2: Chart of top 30 Nations with overweight problems

The stated health problems and the need for an application that can properly aid the dietician in a timely fashion are the factors that motivated the research performed and presented in this thesis. The solutions that currently exist for performing this type of task are slow, unreliable, prone to a high level of errors and specific to certain environments and equipment. In addition, they are often impractical and discourage the patient from performing the necessary periodic measurement of calorie intake, which increases the number of underreported meals and the lack of interest of patients in the use of this type of technology in their own treatments [4] [5] [6].

In the image processing area [7], the image segmentation concept is defined as the partition of an image into regions or objects. This process is considered to be a mid-level processing related to the complexity of the task involved. A digital image is composed of a finite set of elements, each with a particular location defined by $(x, y)$ and a specific value given by the intensity. These elements are regularly called picture elements or pixels. Performing image segmentation requires an understanding of the pixels as units of the image and their behaviour as part of regions of objects, inside of the image. In this case, the similarity between pixels and their locations are two characteristics used for classification and segmentation of the objects and regions present within an images. This capability for information extraction can be a powerful tool for use in the treatment of overweight.

## 1.2 Problem Statement

The idea presented in this thesis is the creation of a "*Nutrient Intake Monitoring System (NIMS)*" based on a process of image segmentation. The tasks involved in the development process are difficult and we need to state each problem, one by one, so that we can split the initial problem into several small tasks that can be completed to finally come up with an integral solution. In image analysis, we face problems such as:

- Quality of the picture to be analyzed
- Capacity of the user to take clear pictures
- The detail extracted cannot be too high in the segmentation process
- Illumination present in the environment while the user is taking the images

We can find some applications available in the mobile apps market related to this topic, but the majority of these are regular calorie counters, meal planners, or recipe managers and practically all of these applications require a high commitment and interaction by the user, since all of the information must be added manually into the application.

With these limitations, our intent was to produce a solution to this main problem:

- Produce a “*Nutrient Intake Monitoring System*” that will enhance and facilitate overweight and obesity medical treatment.

As aggregate value to this research, other conditions could be addressed as secondary goals:

- Based on the image capturing conditions, what is the best path to approach segmentation inside the food images?
- Can this application be extrapolated to more users and uses, in addition to medical treatment?

## 1.3 Proposed Solution

The application is intended to be intuitive and easy to use for the final user, who can be either the patient or the dietician. The application should enable the user to make use of it while moving from one location to another and the usability of the application cannot be attached to extreme conditions, or needs. The application must help and make suggestions to the user, inside a friendly environment, via an inviting interface that will cause the patient to enjoy the use of the calorie monitoring system. This will increase the number of recorded meals and will give dieticians more data to analyze for performing their regular tasks.

Figure 3 shows a diagram of the application. The tasks in green are the sections proposed as part of this thesis; the sections in orange are the sections developed by other team members.

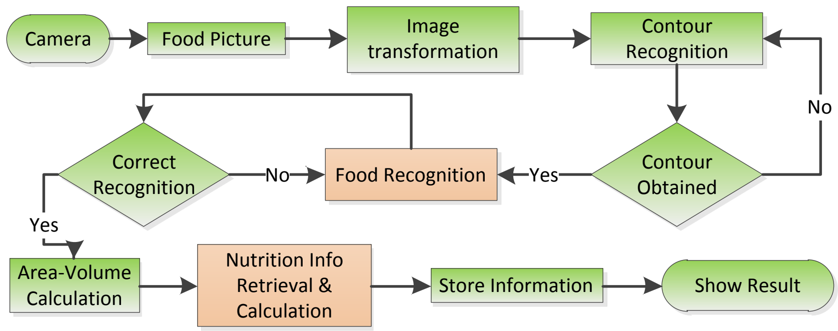



Figure 3: Proposed solution for the NIMS

Each of the functionalities proposed for the application are key conditions for the success of the system construction. The following is the set of aspects analyzed in depth to obtain a better understanding of each aspect.

- First of all, it is important to state that accurate food recognition and calorie evaluation is not possible because even nutritionists cannot tell the number of calories by looking at the food, and even they will have to "guess", in order to generate the required information. With this in mind, our intention was to develop an application that would aid the dietician in maintaining a calorie intake record for obesity treatment. In the process, the application would produce information as close as possible to reality, taking into account the initial condition.
- Picture taking and use: any type of hardware selected must be able to take and use pictures. This is because our approach is focused on image processing by segmentation and image analysis. This feature is a must because the pictures of the food are the main source for proceeding with the rest of our work.
- Mobility: to address this characteristic, our intention was to work over mobile applications for Smartphones. In this case, we can consider 3 different environments that will produce a similar effect: the Software Development Kit (SDK) for IPhone, Android and BlackBerry. Any of these

solutions will work correctly for our purposes, due to the similar capabilities of the hardware inside these three different technologies. Any of these artefacts can take pictures, run applications, connect to internet and apply all of these features in an easy-to-use interface.

- Use of a simple Measurement Pattern: once the picture is captured, at some point in the processing we must use a measurement pattern inside of the image taken by the user. With this, we can scale the portions from the image size into a real life size, in order to proceed with the measurement and calorie calculation per portion. For the measurement pattern, we need to have something simple: an object practical enough to carry around and not be aware of its existence until it is needed. In this case, we propose something as simple as the thumb of the patient. Initially, an image of the thumb will be captured and stored with its measurements. With this pattern, we can then perform the scale and calculations needed. As an alternative option, the user could use a coin instead of the thumb; this will add an extra degree of freedom in the use of the application.
- Image processing: one of the ways we can perform a calorie intake measurement for the patient is to let the user take pictures of the food before and after each meal, to capture the exact type of food inside of each image, and then do a simple image subtraction to remove from the calorie consideration any leftovers of the food. This is one of the most difficult sections of our work, but it is the core process of the application, and a good image segmentation and analysis will produce a successful project and application.
- Use context to increase success: our application cannot predict or measure some special conditions, such as poorly illuminated food, object occlusion, complexity in the image composition, types of food and types of ingredients used to prepare the food. To overcome most of these limitations, we will consider the context of the user in order to perform better

food recognition and prediction. Basic concepts from contexts like user nationality, user location, time of the day, user schedule and user routine will be considered to make the predictions inside of the application.

## 1.4 Contributions

- This is the first semi-automatic expert-out-of-the-loop application for the analysis of food images to aid patients suffering from overweight or obesity. In this case, the expert-out-of-the-loop concept will enable the users/patients to obtain an analysis result of their food intake from the application that will try to simulate the calculation procedure performed by their dieticians.
- The system will reduce under-reporting of food intake by the patients, and will encourage the user to take advantage of the application with a simple interface and a semi-automatic process of image analysis.
- A practical method is designed and developed for scaling and performing the measurements inside the application. This will allow the users to take pictures in any desired place using only their smartphones and thumbs.
- A food specific image database is proposed to recognize food portions and possibly recognize the overall food. Example: The database will contain simple training images with only one type of food in it, and we can use those images as patterns to perform the recognition.
- A method is designed and developed to convert a food image to a food volume, using the top and the side images of the food to reproduce an approximation for the food volume.
- Proof-of-concept implementation and performance evaluation.

## 1.5 Research Publications

The peer reviewed publications related to the field of Food Intake Analysis Based on Image Processing are given below.

## Conference Publications

Gregorio Villalobos, Rana Almaghrabi, Behnoosh Hariri, Shervin Shirmohammadi, "A Personal Assistive System for Nutrient Intake Monitoring," *International ACM Workshop On Ubiquitous Meta User Interfaces*, (Accepted for publication) Nov 2011. [8]

Rana Almaghrabi, Gregorio Villalobos, Parisa Pouladzadeh, Shervin Shirmohammadi, "A Novel Method for Measuring Nutrition Intake Based on Food Image," *IEEE International Instrumentation and Measurement Technology Conference*, (submitted) January 2012.

Gregorio Villalobos, Parisa Pouladzadeh, Rana Almaghrabi, Shervin Shirmohammadi, "An Image Processing Approach to Calorie Measurement for Obesity Healthcare Applications," *IEEE International Symposium on Medical Measurements and Applications*, (submitted) January 2012.

## 1.6 Thesis organization

The thesis is organized as follows:

Chapter 2: Background and literature review.

Chapter 3: Related Work.

Chapter 4: Proposed Methodology.

Chapter 5: Proposed System.

Chapter 6: Prototype Implementation.

Chapter 7: Experimental Results and Evaluation

Chapter 8: Conclusions and Future Work

# Chapter 2 Background

The background analysis includes three major areas to discuss and bring into consideration: Obesity Analysis, Image Processing and Decision Making based on Statistical Methods. The high relevance of these topics for our project forces us to give a meticulous explanation of all three in sections 2.1, 2.2 and 2.3.

## 2.1 Obesity and Measurement Methods

Obesity is considered a life threatening condition, and the number of diseases related to this condition is considerably high. Overweight can be a precursor of illness like diabetes, colon cancer (among other types of cancer), cardiovascular disease, hypertension, stroke and atherosclerosis, to mention just a few.

In the measurement and clinical classification of the adipose tissue and the distribution of the fat over the body, we need to consider the types of cells present in the human body, the type of obesity the patient needs to deal with, and the location of the fat cells in the different parts of the patient's body.

Once the condition of obesity is detected in a patient, a pattern of increased size and number of fat cells is correlated to the condition. By analyzing the size of fat cells, to obtain the average size per patient, and knowing the amount of fat present in the body, a simple estimation can give us the number of fat cells present within the body. Note that even for the same patient, the size of the fat cells will be different for different body parts, and with this taken into account, the average number of fat cells can be calculated. This concept will be more completely explained in the section on measurement methods for body adipose tissue.

The obesity condition has been evaluated and classified for more than 2 decades. These analyses have helped identify conditions such as Hypertrophic Obesity, which is an enlargement of the fat cells that produces a truncal fat accumulation; this will produce an android fat distribution, where the patient will develop the typical apple shape. Another identified condition  is hyper-cellular obesity, which is characterized by regular development during the childhood of the patient. In normal cases, due to the manifestation of the obesity at an early age in the patient, the obesity tends to be severe for the patients. The fat deposits can be characterized by lipodistrophy, single and multiple lipomas and liposarcomas. Lipodistrophy is a condition where the patient loses the fat tissue in one region of the body or in multiple areas. The patient cannot assimilate insulin, or the absorption of it is minimal, which results in a decrease in adipose tissue. Lipomas are a specific encapsulation of fat cells in different areas of the body; they can be present as single or multiple encapsulations, and their size can vary between 1 and 15 cm. In contrast, liposarcomas are normally very rarely found in the patient’s body, but if found, they are regularly located in the lower extremities.

In terms of obesity, an anatomical classification that analyzes the number of adipocytes present in the human body can be used to define this condition. Different types of approaches can also back up the diagnosis: the following are the most common methods used for measurement. Every approach is reviewed in detail in order to understand how each procedure works and the advantages and disadvantages of each of the methods.

### 2.1.1 Body Mass Index

Initially, we have a very simple calculation for obtaining the Body Mass Index (BMI) that is routinely used in health organizations such as the World Health Organization. In this analysis, we have two formulas to apply:

1. $BMI = \frac{Weight\ in\ Kg}{Height\ in\ m^2}$
2. $BMI = \frac{Weight\ in\ pounds * 703}{Height\ in\ inches^2}$

These formulas can give us only a coarse approximation because several specific conditions can apply to each different person. Nevertheless, this method is used regularly by international health organizations to measure indices of obesity over populations and other indicators.

### 2.1.2 Dual Energy X-ray Absorptiometry

In cases where a more thorough analysis and understanding of the real condition of a patient is needed, options of analysis like the Dual Energy X-ray Absorptiometry (DEXA) are available. Even though this procedure was designed to measure the bone mass and density, it is proven to work very accurately for the measurement of the body fat, generating information about the amount of fat present in the entire patient, and also per region of the body. The equipment can distinguish fat mass, free fat mass and bone structures with one scan. The measurement is performed by evaluating the differential attenuation of two X-ray beams with different intensities to obtain the final results. The radiation to the patient is very low, which means that this is a safe procedure, but the equipment is high in cost, and well trained experts are needed to operate the equipment and perform the analyses.

### 2.1.3 Densitometry Analysis Based on Bioelectrical Impedance

Bioelectrical impedance is based on the conductivity of the human body, the resistance opposed by the body to move this electrical current and the inverse proportion of the total water present inside the body related with the resistance. The test is performed by locating an electrode in the right arm and in the right leg; the electrical current is the moved from one electrode to the other, and the measurements are performed. With the raw data produced by the equipment, a set of equations that include age, height, weight and sex of the patient are applied to evaluate densitometry with lean body mass, and the corresponding adipose cells present, in the total equation.

This procedure is simple, easily accepted by the patient and highly accurate for performing body fat measurement. The equipment is portable, so it can be performed at different locations, which increases the use of this type of technique. Different types of equipment are available in the market to perform this test, with variations as to where the electrodes are located and used.

### 2.1.4 Skinfold techniques

This technique is simple and does not require a highly trained professional; it is also a non-invasive procedure, which means that it can be performed on the patient every day with no risk or secondary effects involved. The technique has proved effective to some extent, but high inter-observer variation occurs, even when the skinfold thickness is related to the body fat. The differences between subjects of these studies are even higher when the analysis is performed on obese patients.

In this procedure, the expert pinches the skin of the patient, producing a skinfold to measure the subcutaneous fat layer, strictly underneath the skin of the patient. A special caliper is then used to measure the skinfold. The measurements are then translated using predefined tables to give an approximate estimation of the relationship between the skinfold, the amount of fat present and the relationship with the entire body of the patient.

Even with all the limitations of this method, it has seen many reliable applications, because of its ease of use, not only in medical treatments, but in everyday analysis like the initial and monthly tests performed by personal trainers in gymnasiums and sport centers.

### 2.1.5 Circumferences

This type of measurement focuses on the fat distribution over different body parts; its accuracy is an element of concern of this method, even when it is less prone to produce inter-observer errors in a set of obese patients.

The measurement of the circumference is performed over the upper and lower arm, the upper and lower leg and over the abdominal and gluteal area. These measurements can evaluate different parts of the body, and the fat accumulation in these specific parts. An important analysis is performed with this method; namely, comparison of the Hip to Waist Ratio, given the high risk related with abdominal obesity compared with gluteo-femoral obesity.

Now that we have defined the basic concepts for overweight and obesity, we need to focus on the technical sections of our project The following section of this chapter covers the basic terms and definitions of the vast area that is image processing and image analysis.

## 2.2 Image Processing Background

Digital Image processing is the computer vision area where the input of the application is an image or a set of images, and the result of the process inside the application will be a new image, a set of images or a collection of characteristics extracted from within the image or images. This area of computer vision routinely requires a large number of tests and experiments to produce acceptable results. Each image contains a finite set of elements called image elements, picture elements or pixels. These elements have a specific value, usually referred to as intensity, and a location in a spatial plane, defined by x and y. A two dimensional function $f(x, y)$ can be defined using these coordinates.

### 2.2.1 Colour Spaces

The digital images contain picture elements already defined in the previous section as pixels; these pixels contain finite values (three or four values) generally based in a mathematical abstraction model to represent the colour required by that specific unit inside the picture in order to create the complete image. The models used to define the internal structure of the image are also called colour spaces, which define the specific type of information stored per pixel and the correct way to display these values in the visual representation of the image. Here we have basic models such as RGB and RGBA: these models are simply a saturation value assigned for each channel; in this case the red, green and blue for the first model, and an extra value of transparency called A, in the second. These basic standards are generally applied inside computer applications and computer vision analysis. Other colour encoding schemes are also recognized, such as the ones used for the transmission of television signals, like the YUV and YIQ, but these models are beyond the scope of the present research.

The International Commission of Illumination (CIE) has defined a specific diagram to define the colour chromaticity defined in the plot, which is the wavelength of light in nanometres. This type of plot could help us to understand, in an easy way, how to work with the different values for each colour channel red, green and blue, and to send these values to analog and digital displays like CRTs and LCDs. Because this representation is a well-accepted standard, many instruments of colour measurement, colour display and colour capture use this definition to work with the wavelengths of each colour in order to perform the corresponding analysis and representation in a two-dimensional plot.

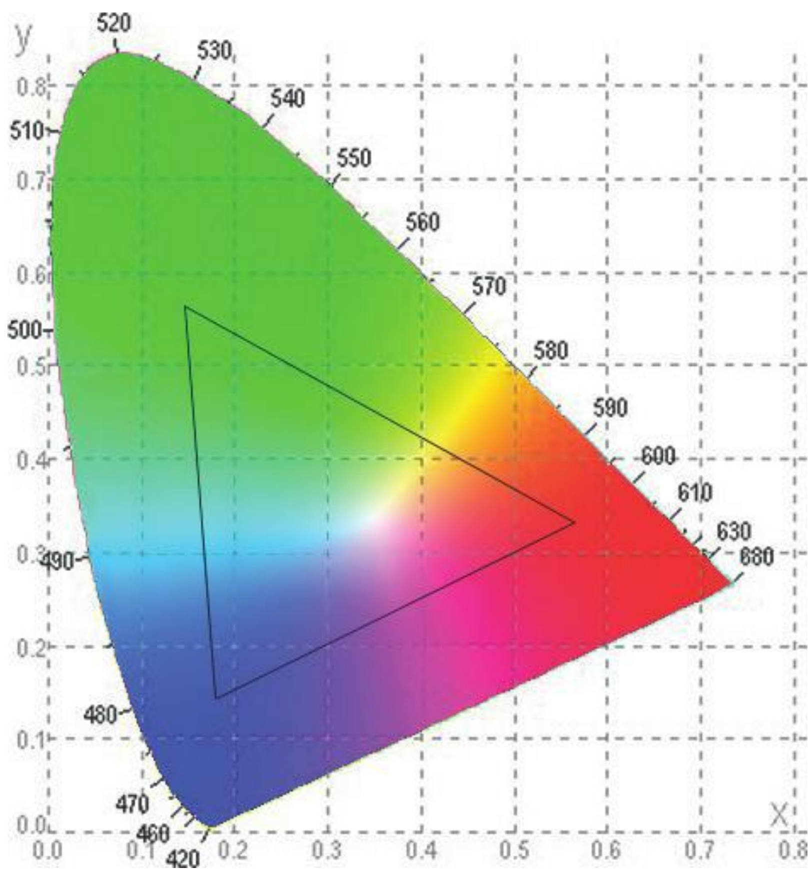



Figure 4: The CIE Chromaticity Diagram

Figure 4 The CIE chromaticity diagram. The scale of each axis is in nanometres, based on the wavelengths of the colours moving in the RGB scheme.

One colour coordinate system is a model known with 3 different names that refers to a similar approach of the same colour system: the model is variously known as HSI (Hue, Saturation and Intensity), HSV (Hue, Saturation and Value) or HLS (Hue Lightness and Saturation). The model is used for noise reduction on the image, the understanding of colour as a single value and the application of segmentation based on the specific characteristics of a defined surface of an object present on the image. This colour space is also highly related to how human vision works, and also how physical characteristics of colour considerations from many different specimens can be represented. Other colour spaces are represented geometrically and they are to some extent close or similar to the HSI colour space. An example is the L-a-b model, where L is defined by the luminance, or the amount of gray in the available scale, the value a is defined by the distance between the red and the green (it goes from +a to –a), respectively, and b is defined from +b to -b and it goes from yellow to blue. A specific relationship exists between the colour spaces HSI and RGB, and combinations of both allows translation of a simple conversion of the RGB channels into a specific new colour space that is highly dependent on the primary information generated by the RBG colour space. This is known as YCrCb; the intensity falls into the characterization of Y, while the hue and saturation are kept in the chrominance space corresponding to Cr and Cb. The transformation to this colour space relies on the analysis of the perception of different characteristics present inside the image.

Figure 5 shows a bi-conic representation of the Hue, Saturation and Intensity colour space. As shown in this image, the distance between black and white defines the intensity value, saturation is defined by the amount of colour used and hue is the colour itself. With the value of hue, we can have a unique numerical representation for each colour on the space representation.

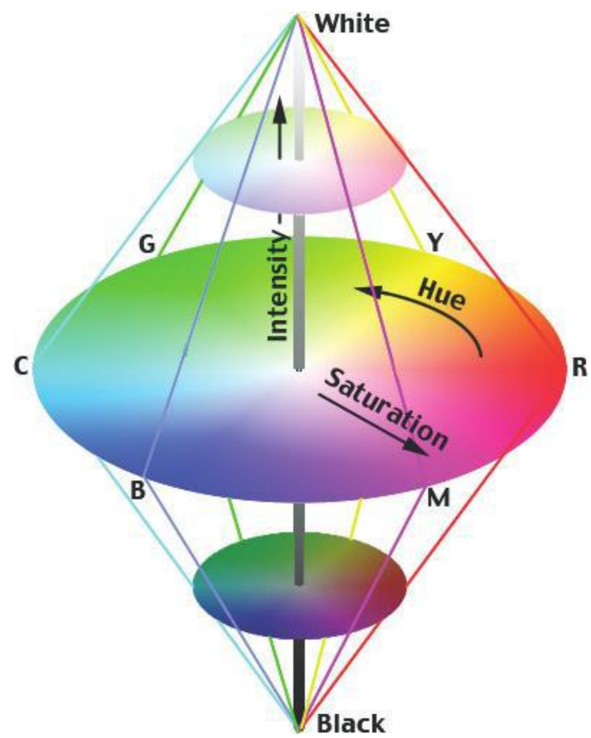



Figure 5: Hue-Saturation-Intensity Space bi-conic representation.

### 2.2.2 Conversion between RGB and HSI

The conversion from RGB colour space to HSI colour space is an important method often used to obtain the equivalent image in a different colour space, in order to perform different types of analysis and processes over the information stored inside the image. To move from RGB to HSI, we can use the following matrix multiplication to obtain the first set of values, V1 and V2, needed to obtain the final values of the transformation for HSI in each unit of the image:

$$\begin{bmatrix} I \\ V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

With the value of illumination I and the intermediate values V1 and V2 generated, we can proceed then to evaluate the final values for H and S:

$$H = arctan \frac{V_1}{V_2}$$

$$S = (V_1^2 V_2^2)^{1/2}$$

We must also, at this point, define the set of formulas needed to perform the inverse transformation, in this case to move from HSI colour space to RBG colour space. Initially, we must find the values for V1 and V2:

$$V_1 = S \cos H$$

$$V_2 = S \sin H$$

With these values defined, we can then perform the matrix transformation, as follows:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & \frac{-\sqrt{6}}{6} & \frac{\sqrt{6}}{2} \\ 1 & \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{2} \\ 1 & \frac{\sqrt{6}}{3} & 0 \end{bmatrix} \begin{bmatrix} I \\ V_1 \\ V_2 \end{bmatrix}$$

The conversion is not an exact translation, due to the color management in one channel in the HSI scheme, compared with a more accurate color system such as RGB, this condition and the trigonometric functions applied to the calculations produce differences in the final calculations of the conversion from RGB to HSI and from HIS to RGB, due to this conditions the matrices used to perform the conversions are not symmetric matrices.

### 2.2.3 Pyramid Application in Image Processing [9] [10]

Convolution is a mathematical approach for combining two signals with a finite set of numbers to produce a third signal. This mathematical approach is used in many fields, such as probability and statistics. Convolution is used in image processing to implement specific transformations to a set of pixel values using linear combinations, to produce new values for the pixels. In the case of computer vision, one input of the convolution is an image treated as a set of values and evaluated as a matrix, while the other input is a smaller matrix with the values to perform the linear transformation over the first matrix; this second matrix is also known as a kernel. The kernel must be applied over the entire first matrix—in this case the image—taking into account only the positions where the kernel fits entirely inside the boundaries of the image. The convolution of the image is simply applied, pixel by pixel, to obtain the new values of the different positions of the initial matrix that defines the image structure. If we understand the image matrix with M rows and N columns, and the kernel matrix to be composed by m rows and n columns, the final size of the result will be a matrix of $\mathrm{size\,M - m + 1}$ rows and $\mathrm{N - n + 1}$ columns. With these concepts defined, we can express the convolution as:

$$O(i,j) = \sum_{k}^{m} \sum_{l}^{n} I(i + k - 1, j + l - 1) K(k,l)$$

The application of pyramids as a generalization of the Burt algorithm is one of the tools we have at hand to perform the elimination of degradations inside the digital image, as well as to accentuate important characteristics that can be useful indicators inside the image for further analysis. The basic idea proposed by Burt et al. [10] is that, in the same scene, once the image is captured and based on the distance used to create the digital image, we can obtain or miss different types of characteristics and conditions attached to every object in the scene. Furthermore, the only way to achieve a complete analysis is to perform it at different levels of the scale at the same time. To perform this type of analysis at different levels simultaneously, we need to create a set of images of the same scene called a pyramid. For the application of the pyramid, it is more efficient to generate a downsizing convolution of the image analyzed instead of a regular convolution transformation, because the final results are highly similar and the time of application is always higher in the convolution application. The proposed idea is to generate a data structure that supports efficient scaled convolution, using the task over a set of images in reduced scales from the original image. To obtain the different versions of the original image sub-sampled, the image is filtered and sub-sampled by a factor of 2 every time, and the amount of sub-sampling application is the level of the pyramid built. With this definition, we can move from $G_0$ to $G_1$. The pyramid is obtained iteratively, with $0 < l < N$:

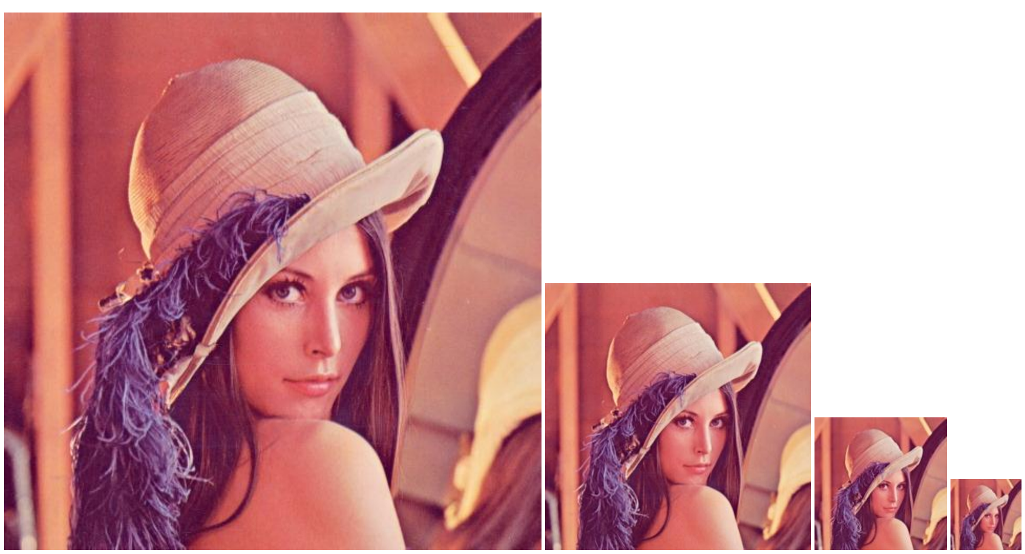
$$G_l(i,j) = \sum_m \sum_n w(m,n)\, G_{l-1}(2i+m, 2j+n)$$

Generating the pyramid using this process is fairly similar to applying Gaussian weighting functions to generate the convolution of the image. This is the main reason why this pyramid application can be also called a Gaussian Pyramid. Once the analysis has been performed over the image sub-sampled, performing a reverse procedure is important to return to the original image size. The expansion process can be obtained by the following formula:

$$G_{l,k}(i,j) = 4\sum_{m}\sum_{n} G_{l,k-1}\left(\frac{2i+m}{2},\frac{2j+n}{2}\right)$$

Since the reduction of the image is in a scale of 2, verifying the size of the image in width and height is important, in order to verify if both values are multiples of 2. In cases where the division cannot be performed, an extra transformation to resize the image must be completed to be able to apply this analysis.

Figure 6 shows a set of images where a sub-sampling has been applied to obtain the next level of the processed pyramid.



a) Original Image b) 1st Pyramid Level c) 2nd Pyramid Level d) 3rd Pyramid Level

Figure 6: Set of images of a 4th level pyramid

### 2.2.4 Boundary Definition

Once the images have been modified with different techniques to increase the accentuation of specific characteristics important for classification of the objects inside the image, we can concentrate on the definition of the boundaries between objects, and between the object and the background. This can help us define region boundaries; each region can become, at the end of the image analysis, part of the foreground in the sets of objects of interest, or part of the background. To perform the boundary definition, we can rely on histogram analysis to accentuate the changes in brightness and signalize the important regions that fall into this characterization. Another method is to apply a convolution with a Laplacian or a Gaussian kernel, and to generate similar results to help us demarcate the boundaries needed for each region of the image. For the procedure of colour rasterization and smoothing the textures over the entire image to produce similar effects to that of a regular convolution, we have at hand the Burt's pyramid as a faster solution for generating the boundary accentuation needed. In these cases, overcoming several difficulties such as illumination, image quality and similarities between adjacent regions becomes important. The use of active contours has been applied with success for these conditions, allowing identification of blurry boundaries over the images needed [11] [12].

### 2.2.5 Active Contour Extraction

An active contour is also known as a snake because of its behaviour over the different iterations of processing. It is a curve defined by the specific characteristics of the object of interest inside the image, and this type of curve will adapt itself following an energy minimization function [11]. For the energy minimization function, we need to understand the contour as a controlled spline under the manipulation of image and external forces. The function can be expressed as:

$$E_{snake}^{*} = \int_{0}^{1} E_{snake}\left(v(s)\right)ds = \int_{0}^{1} E_{int}\left(v(s)\right) + E_{image}\left(v(s)\right) + E_{con}(v(s))\, ds$$

The first integral of the equation is defined by the parameterization of the initial snake, given by $v(s) = (x(s), y(s))$. In the $E_{int}$ section, the internal image energy spline is considered; meanwhile, $E_{image}$ brings into the equation the energy force related to the image itself. Finally, the $E_{con}$ helps us to consider the external forces that will also affect the active contour.

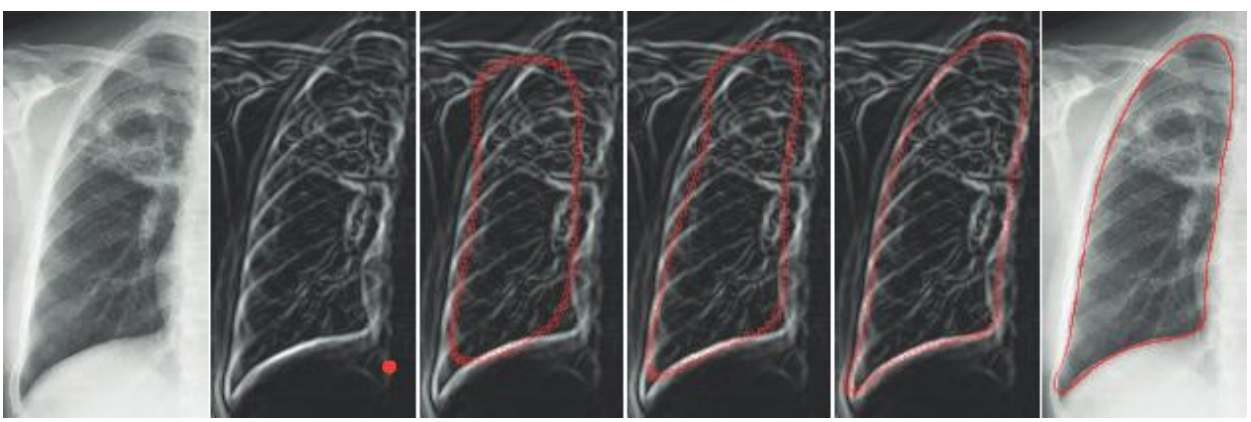
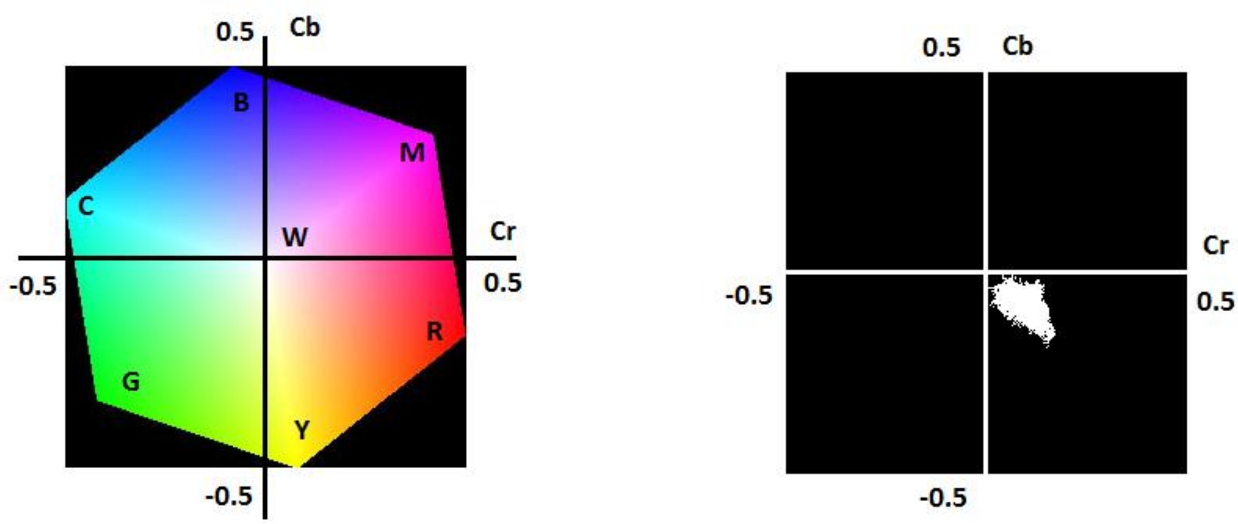


Figure 7: Images with an Active Contour example

Figure 7 depicts an image showing different sections of a contour calculation; the image processed is from a lung captured by x-ray. The initial image shows the original image, while images 2, 3, 4 and 5 show the same processed image and how the active contour is modified over the image, over and over, in every step of the application. The final image is the combination of the initial image and the final active contour retrieved from the computational calculations in image 5.

## 2.2.6 Skin Recognition

For skin and face detection investigations [13] [14], physical experiments indicate that one of the most common ways of perceiving colours can be characterized by 3 attributes: Hue, Saturation and Intensity. The easy separation and recognition product of the use of YCbCr colour space allows this configuration and transformation to be applied to the image, in order to characterize the sections of the picture that may contain skin pixels. Human skin falls into a very specific and special category of colours; this set of possible colours that skin represents is to some extent unique, if we compare its localization in the colour representation plot with respect to the normal colour present in common objects. Figure 8 shows the distribution of the common RBG colours over the Chrominance Cb and Cr in image a. inside the image b. are is the specific location of the possible skin tones found inside the images processed with this colour space. Based on the fact that Cb and Cr consider hue and saturation, skin tones are relatively consistent and no matter what skin tone is present, in the chrominance plane this is just a small area to analyze in order to obtain a good approximation.




a) RGB distribution over YCbCr    b) Human skin Distribution over YCbCr

Figure 8: Images of CbCr colours and skin location

Once the area of focus has been defined, a characterizer must be applied in order to define the pixels corresponding to a possible skin tone and the pixels that do not correspond to a skin tone. For this, the Bayesian Decision Rule for Minimum Cost can be applied, as follows:

$$R_0(X) = C_{00}p(w_0|X) + C_{10}p(w_1|X)$$

$$R_1(X) = C_{01}p(w_0|X) + C_{11}p(w_1|X)$$

$$R_0(X) < R_1(X) \ \rightarrow X \in w_0$$

$$R_0(X) > R_1(X) \ \rightarrow X \in w_1$$

For these formulas, the term $w_0$ corresponds to the pixels with no skin colour, and the term $w_1$ corresponds to the pixels with skin colour. The term $p(w_i|X)$ corresponds to the probability of being in a specific class, based on the sample $X$. $C_{00}$ and $C_{11}$ are the terms related to the cost coefficient of the correct classification, and $C_{10}$ and $C_{10}$ are the costs of misclassifications. Based on the given values, $R_i(X)$ is the cost of classification the unknown sample into the class $i$. These formulas will help us classify the pixels, following a simple rule of minimum cost decision making, as follows:

$$C_{10}p(w_1|X) < \ C_{01}p(w_0|X) \ \rightarrow \ X \in w_0$$

$$C_{10}p(w_1|X) > \ C_{01}p(w_0|X) \ \rightarrow \ X \in w_1$$

This kind of analysis becomes a practical problem with a possible solution, thanks to the characteristics of the colour space used to analyze the information.

## 2.3 Support vector machine (SVM)

A support vector machine (SVM) is a computer algorithm that learns by example to assign labels to objects. The SVM is successfully used for pattern recognition in various fields. The method is a very useful classifier since it determines a hyperplane that separates classes with the largest margin between the vectors of the two classes [15]. The problem here is how obtain a linear separation. SVM can solve the problem by using a kernel that shifts the feature space into a higher dimension. The linearly separable hyperplane in the higher dimensional space gives a non-linear decision boundary in the original feature space.

The sample set is:

$$\{(x_i, y_i)|i = 1,2, \dots, k\},$$

where the values of $x_i, (i = 1,2, \dots, k)$ are input values and $y_i$ , $(i = 1,2, \dots k)$ are class values of the method for the support vector machine. After apply kernel function, the corresponding classification is:

$$f(x) = sgn(\sum_i a_i^* y_i + b) \quad (1)$$

SVM could also be used in nonlinear regression analysis. Now, we replace the $(x_i, x)$ by $\Phi(x_i)$.

$$K(x_i, x_j) = \Phi(x_i). \Phi(x_j) \quad (2)$$

so nonlinear regression will be:

$$f(x) = w. \Phi(x_i) + b = \sum_i (a_i - a_i^*) y_i K(x_i. x) + b) \quad (3)$$

In this formula, if:

$$\sum_i (a_i - a_i^*) = 0,$$

then

$$0 \le a_i \le C , a_i^* = 0.$$

For the equation, $x_i$ is called standard support vector and b could be calculated from constraint condition. The most commonly used functions are:

1. Linear function

$$K(x_i . x) = x_i^T . x \quad (4)$$

2. Polynomial function

$$K(x_i . x) = (ɣ x_i^T . x + r)^d \quad (5)$$

where ɣ > 0

3. RBF (Radial Basis Function)

$$K(x_i . x) = \exp(-ɣ \| x - x_i \|)^2 \quad (6)$$

4. Two layer nerve function

$$K(x_i . x) = \tanh(ɣ\, x_i^T . x + r) \quad (7)$$

ɣwhere $r$ and $T$ are kernel parameters.

SVM requires that each data instance is represented as a vector of real numbers. Hence, if categorical attributes exist, we first have to convert them into numeric data. Actually, scaling before applying SVM is very important [15] [16].

# Chapter 3 Related Work

## 3.1 Food Intake Measurement

A determination of the impact of mobile telephones in young people and how this technology has an important role in their regular communications has been presented by Lenhart et al. [17]. The analysis confirmed how the social interaction based on mobile communication is an important tool, and that it inclusion of this type of equipment in our efforts could be useful for tackling certain types of procedures that include young people, for and for increasing the use of the application and the capability of final analysis. Children and adolescents are willing to learn and to interact with new technologies; this is an important fact that could make the difference in the project itself. This research takes on a different meaning when we consider the main problems found in conventional dietary assessment methods, where the patient finds the procedure difficult, a burden, and lacking incentives or creative sections that would increase the number of food reports and the accuracy compared with the real food intake of the patient [18] [19]. In this research, the methods analyzed to perform the measurement of daily food consumption have been common methods used by dieticians, such as the food record, the the24-hour dietary recall (24HR), and a food frequency questionnaire (FFQ) with external validation by doubly labelled water (DLW) and urinary nitrogen. For these methods, the data are logged manually, and this common factor introduces a high number of errors into the overall process. Another drawback from these methods is the difficulty of making the patient commit to the treatment. Our project will develop a semi-automatic application, with a simple but functional design, that will encourage the patient to interact with the application. The target hardware is mobile devices, such as smartphones. In this case, we intend to make use of the new technologies to increase the chances of use, and to log the complete set of meals consumed by the patient.

The main disadvantage of manual methods is the delay and inaccuracy of reporting the eaten food. This condition arises from several factors, such as age, gender, education, credibility and obesity. Harnack, et al. found significant underreporting of large food portions when food models showing recommended serving sizes were used as visual aids for respondents [20]. The report problem will be increased with our proposed system, due to the special characteristics that will make use of semi-automatic measurements and recognition to calculate the amount of calories present in the food. With this, the numbers of interactions with the user are minimized, reducing the human error in the process.

Zhu et al. [21] has demonstrated the difficulty of applying the previous assessment methods for measurement of the daily food intake in adolescents aged 11-14 years. They reported several drawbacks, but the two most important were the resistance developed by the adolescent to be under the parent's control and being under medical monitoring, even when the process is controlled by the physician. Our application will allow the adolescents to take control of a simple and fun application available inside one of their favourite gadgets, thereby increasing the patient interaction with the food log process and increasing the overall treatment process. S. Franka et al. [22] proposed analysis of hunger and gender as key aspects of calorie consumption. They analyzed a set of patients using Magnetic Resonance Imaging (MRI) to evaluate them in conditions of hunger and lack of it, and how the food in its different conditions can be stored as fat inside the bodies in different ways, based on the initial conditions of the intake. In this case, the imaging section was obtained by magnetic resonance and the analysis performed was to understand the behaviour in the energy storage developed in the body of the patients during the time the study was conducted.

Other studies have analyzed the gender, age and perception in order to predict the calorie intake [23] [24], but these were highly related with the sample used to perform the analysis.

The problem of overweight is growing across the entire globe, and the incidence of obesity and overweight in all age ranges, genders and countries is a palpable problem that all nations need to solve or reduce, due to its high cost in related treatments and the number of diseases linked to the obesity. In Canada, the proportion of obesity based on age distribution has increased from 10% during 1970, to more than double in 2003. This number has been reported to have grown to 23% in the assessment performed in 2003; in addition, the study also evaluated the incidence of obesity related diseases in the adult population, and that relationship increased by 138% for men between 1970 to 2003 [25].

## 3.2 Image-Based Measurement

The eagerness of users to explore high end gadgets such as smartphones has become the cornerstone for introducing these modern technologies in the construction of solutions to monitor food intake, and to help improve the way that users and nutritionists focus their efforts in obesity treatment. The interaction of food analysis based on imaging processing, with the help of high end technologies, has been proposed by Harnack et al. [20]. In this research, the image analysis is focused on large food portions. M Sun et al. [26] proposed a similar approach, where the idea was to take pictures of the food, and a calibration card located inside the picture was used as a measurement pattern to calculate the size of the food portions. The food is manually identified with the help of nutritional information retrieved from a database, the calories are calculated for each picture, and finally, the complete set of information is stored in a different database in the research facility. In this case, based on the known size of the calibration card, the portions can be translated into real life sizes, and the calculations are closely related with the real caloric content of each food. To overcome related shortcomings, our application takes advantage of a new measurement pattern definition, where the thumb of the patient will play this role. The simplicity of this option will enable the patient, in any place and any condition, to perform a food analysis in a practical manner. Y. Saeki et al. [27] proposed an approach that performed the complete measurement in a research facility, with special equipment that allowed the system to obtain images from the food with ideal conditions of resolution and illumination in a highly controlled environment that will extract the food present in multiple dishes on a tray. The approaches already explained must meet several conditions to perform the complete analysis, and this kind of requirement transforms the simple task of taking the picture of the food into a complicated step in the overall process. We intend, instead, to produce an application that is simple and portable, thereby increasing the number of reports per day and simplifying the task for the patient and the dietician involved.

Martin et al. [28] proposed a system where the user captures the images, again with a calibration card, and then the image is sent to a research centre for analysis. This will, of course, come with its own shortcoming of offline data processing. In contrast, our system is intended to perform the analysis in the same place where the patient is located. Using a different approach, P. Chi et al. [29] introduced the idea of a smart kitchen, designed entirely to aid the normal person to be aware of the calories associated with the food being prepared inside this facility. A camera overhead captures the images in the preparation process; meanwhile, weighting sensors work in the other section of the measuring control, in the counter and the stove. The environment is an augmented reality, where the normal tasks performed inside the kitchen are enhanced with different sensors, the information is processed by the application, and the user can obtain immediate responses from the application. In a similar approach, M.S. Westerterp-Plantenga [30] designed a food intake monitoring system where the complete set of analysis contains a special table where the plate is located while the patient is eating, the table constantly measures the weight of the food, and the information is quantified in order to measure how much is consumed from the original portion, the speed of consumption by the patient in the process, the lag between bites and the weight of each bite, in order to measure the differences between these. All of this information is compiled in a research laboratory, and the patient must be controlled over several hours. The user typically does not like this kind of control and attached environments, and is prone to forget or will prefer not to use this kind of measurement procedure, as weighing systems are not available everywhere and are not easy to carry around. Chang et al. [31] proposed a dining table equipped with different sensors located within the table, with Radio Frequency Identifiers (RFID) tags that identify the food, and sensors that weigh the amount of the portion. However, this requires the user to eat only at that table, so these environments are impractical, especially if we intend to apply this kind of measurement in common situations, such as at food courts or restaurants.

In our proposed system, we aim to use smartphones as monitoring tools as they are widely accessible and easy to use. We will take advantage of image processing for analysis. However, we will try to improve the quality of the processing compared to that indicated by existing literature and to alleviate the need for a calibration component.

## 3.3 Image Segmentation in Other Disciplines

For the image segmentation process in other disciplines, the approach of analyzing the active contours without edges has been proposed by T. Chan et al. [32]. Here, the analysis is applied by an energy minimization function in order to define the contours and boundaries of the objects present in the picture. Using this method, they focused their efforts on the definition of shapes and contours based on the conditions and distributions of the particles present inside the image. With a different approach, K. Kass et al. [11] presented the idea of active contour models. They worked over the idea to extract important information from an energy minimization function from the splines that was highly related with the image forces, and the results were attracted to the important features, such as edges. This application has some shortcomings due to the need for user interaction, and the definition of a starting point for the contour fairly close to the real contour. This is because the initial idea was to generate a system to aid the user to define the contours—a definition performed by the user and assisted by the application—not an autonomous image analysis application.

As an extension to the snakes work proposed by Kass, L. Cohen [33] introduced the use of active contour models and balloons. In this case, the curve created initially to start the process is inflated by an initial force. With this, the contour will behave like a balloon and the curve will then move over the image, passing over weak edges, and become fixed to strong edges. This enables the curve to start in a random initial position, and not be attached to the proximity of the real contour, as proposed by Kass. Along the same line of research, L. Cohen et al. [34] extended their previous work to operate in 2-D and 3-D images. In this case, the balloon model is deformed as a 3-D deformable surface, using the internal and external forces to attract the surface of the balloon towards the edges. This is achieved by defining the 3-D surface as a consecution of 2-D plane curves. The application has been confirmed as reliable, stable, and with a fast performance.

More modifications have been applied to the initial model; in this case, the definitions of the contour as a snake keep evolving. In these cases, we can find the topology of adaptive snakes, as proposed by T. McInerney et al. [35], where they analyzed a set of medical images to extract the contour of the different types of human tissues extracted by medical equipment, and defined the shape of each tissue based on the use of topology adaptive snakes or T-snakes. They introduced the concept of Affine Cell Image Decomposition (ACID); the contribution to the snake itself is given by the ability of the snake to adapt with topological flexibility, increasing the degree of automation without sacrificing the good performance shown in previous work. The use of gradients, combined with snakes, and the behaviour of the shapes, has been analyzed by C. Xu et al. [36]. They proposed the use of an external force called gradient vector flow, and the snake is created by the force balance condition, instead of the normal variation formulation used in the regular snakes. R. Liu et al. [37] proposed the use of mesh segmentation via spectral embedding and contour analysis, where the 3D mesh is analyzed and bisected, and then a projection of the sub-mesh is spectrally projected into the plane to perform the contour analysis and extraction. The contour extraction from digital images has many uses, like the one proposed by A. Jalba el al. [38] where, for the automatic extraction from an image of diatoms and its proper identification, the model applied a multi-scale mathematical morphology analysis to define and extract the contour. The most important data were then extracted based on an unsupervised cluster analysis. On the other hand, K. Siddiqi et al. [39] proposed the use of active contours to extract and define the shapes of the object present in the image evaluated, using the gradient flow over the image from a weighted area obtained from an image weighting depending factor. The information is evaluated with a partial differential equation (PDE), which helps to extract the shapes, with advantages over the regular procedures evaluated by the authors. In a similar direction of the PDE application, Alvarez et al. [40] proposed a curve evolution to apply to the snake, where a binary morphological set used in a (PDE) can make the snake adapt to the contours. This proposed work appeared to be faster and simpler, the final approach worked stably and the function did not need to be re-installed.

In the colour segmentation area, we have the work of J Angulo et al. [41], who have proposed a colour segmentation by ordering merges, where the initial analysis performed over the images is a separation of sections by the average colour relation between the adjacent pixels, and the colours are rasterized by the application of a pyramid of watershed to finally perform a simple analysis of the colour gradients, and produce the segmentation based on the chromatic-achromatic partitions and the saturation component.

In the sections of shape definition, recognition and extraction, the work of R. Malladi et al. [42] shows a new shape modelling scheme; this new perception of the scheme can be applied to complex and arbitrary shapes, including shapes with significant protrusions, and no previous definition of the characteristics of the shape has been performed, in order to start the process. They located a hyper-surface that will move in the desired direction, based on the gradient fields with constant speed that will be change based on the curvature of the surface.

# Chapter 4 Proposed Methodology

In this chapter, we are going to discuss the methodology used to develop our solution. First we must define our main objective, and the possible limitations that could be faced in the process, based on the structure of the application, by the user and by the environment itself. Once we identify all of the initial conditions and possible pitfalls, we can discuss the structure proposed for our application. This will include how we need to prepare the images to be processed, what set of preconditions will be needed, how is the user expected to capture the images to be analyzed, how to introduce the proposed simple measurement pattern included in our solution, how to make the internal procedures transparent for the user who only will receive the set of results in the review and correction section, and finally how this information will be transformed into the set of calories related to the meal captured in the photo, and how all of this information will be stored at the end of the process.

## 4.1 Objective

Extract from the images the portions of food in order to calculate their sizes and volumes and to translate these into the calorie measurement for the meal.

## 4.2 Limitations

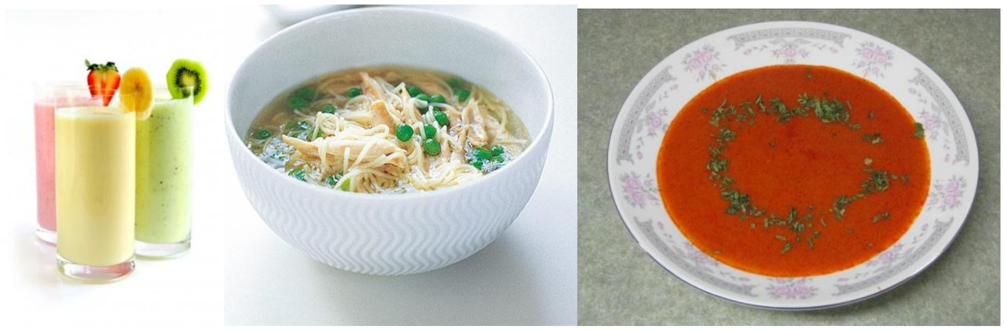
Numerous conditions can lead us to difficulties regarding the correct extraction of the images; these conditions are closely related to different areas that can be classified as limitations of the environment, limitations introduced by the patient or user of the application and the limitations corresponding to the image processing and analysis.

### 4.2.1 Limitations of the Environment

The environment is where the application must be able to run; in this case, we need to consider all of the initial conditions attached to this environment. First of all, we need to understand that even when the applications are designed to be portable, the capacity of mobile devices to store and process applications is to some extent limited. This is true even though the technological advances of smartphones and handhelds has taken important leaps towards equipment with high amounts of memory and storage area and accelerated processor speed; these kinds of devices are still limited by their size and the nanotechnology available. Some specific differences also exist between different models of smartphones, even within the same brand. For example, the camera resolution could be between 3 and 8 megapixels, and this kind of difference could impact the application performance and final results of the product proposed here.

Another important problem in the environment is related with the type of food to be consumed by the patient. For several types of food, even if we perform a correct segmentation, the food cannot be recognized, analyzed or included in the initial considerations. These types of conditions cannot be introduced into the final results to produce a good calculation of the calorie consumption by the patient. The following are the types of food conditions that cannot be evaluated:

- Liquid foods: in this condition, we can consider food items such as soups, fortified smoothies, gravies, baby food, and all types of food with a texture similar to a liquid. In the case of soups, for example, the image will capture only the upper surface of the soup, but we cannot extract the ingredients, unless they are shown inside the picture of the food. In cases of fortified smoothies, the mixture of fruit and the possible content of protein or other special fortifying ingredients will be unknown to the application. Figure 9 shows a set of examples of the liquid foods considered inside this characterization.



a) Power Juice Image b) Noodle Soup Image c) Tomato Soup Image

Figure 9: Images of Liquid Foods

- Wrapped foods: The content is always different, based on the type food selected by the patient and the source where the meal comes from (i.e., a restaurant, fast food chain, or homemade, to mention a few sources). Unless the patient is willing to open the wrapped food and take the pictures of its contents so the application can do the entire analysis, the system cannot guess with any level of certainty the ingredients present inside this type of meal. In addition, if the application introduces links to difficult conditions of processes for the patient, the patient is highly likely to respond by showing resistance to the use of the system. Figure 10 shows a set of examples of wrapped foods considered inside this characterization.

a) Italian Turnover Image b) Chicken Wrap Image c) Dumpling Image

Figure 10: Images of Wrapped Foods

- Mixed foods: the variety of food present in this classification is vast, because a mixed food can be a salad (Caesar Salad, Greek Salad or Garden Salad), or rice prepared with different ingredients in it (Chinese food, Mexican food). Italian foods like lasagna, pasta or pizza, are classed the same way due to the variety and the complexity of the ingredients that these types of food can contain. The problems introduced by this type of food are related with occlusion, amount of ingredients and special contents inside each dish. Figure 11 shows a set of examples of mixed foods considered inside this characterization.

a) Mixed Dish with Salad Image  b) Mixed Rice Image  c) Mixed Pasta Image

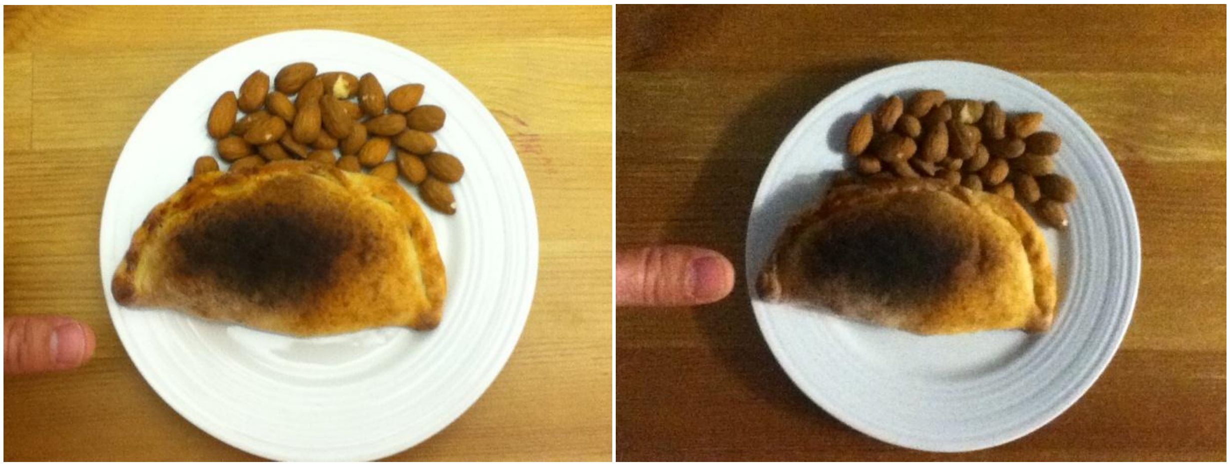
Figure 11: Images of Mixed Foods

### 4.2.2 Limitations introduced by the patient/user

We need to create an application that is simple to use and that can enable the user with a system smart enough to aid the patient in calorie control and dietary planning. Between these characteristics of the application, we also must understand all of the limitations and problems that a regular user can introduce into the application without noticing. The patient cannot be considered an expert in photography; therefore, the system must be able to recognize if the picture taken by the user is not sufficiently clear, or if it has any special conditions that will complicate the analysis process and, in the end, produce erroneous calculations and data for the patient and the dietician. The angle used to take the picture can generate unclear portions for the application, and finally, fewer calories calculated if we compare the number generated by the system and the real number of calories present in the dish.

Location of the patient can become a problem, due to the type of food the user can access based on his/her location. This condition takes on a special meaning when the patient is consuming mixed or wrapped foods, because the location can make the difference between a correct prediction of the food, or a wrong assumption made by the application.

In a combination of the two initial conditions introduced by the user, another problem arises related to the capacity of the patient to input good images into the application and the user's location at the moment of capturing the image for the application. This is related to the available illumination present in the scene for taking the picture. A well-illuminated scene will produce sharp edges, good colour contrasts and well defined contours for the food portions present in the dish. Figure 12 shows a simple example where the food is captured using different types of illumination. In image a, with direct illumination, the definition of the edges and the colour separation for each portion is visually evident, whereas in image b, even though both images have been captured with the same portable device, the definition becomes blurry to some extent, the shadows can produce an overexposed section of the food, and the overall definition of the scene is poor compared with image a. Also note that the distances used to capture both images are slightly different; this also can produce a higher or lower detail of the food, based on the location of the camera with respect to the food at the moment of the snapshot.



a) Image with good Illumination b) Image with poor Illumination

Figure 12: Images of illumination effects

### 4.2.3 Limitations in Image Analysis

In the process of image analysis, we need to consider the amount of blurring in the image to be analyzed as this condition can introduce poor discrimination of edges in the food portions, and will ultimately produce an incorrect contour reconstruction. This can lead to definition of non-present objects, in the case of shadows and ghosting objects, or the consideration of smaller or bigger areas of the portions, with relation to the real life size of them. In this case, another consideration is encountered, when the overlap of two or more food portions produces fewer contours recognized inside the image. This, of course, is a significant error that must be tackled if we want to reduce errors and increase accuracy in our application.

## 4.3 Solution Overview

One of the main characteristics of our application is the definition of Contours to perform the image analysis. The method of contour analysis, definition and extraction is one of the most accurate processes for defining the objects present in the image analyzed. The contour definition will enable the system to define the different subsections of the image and to concentrate the processing over the different sections of interest presented inside the picture. Image analysis is difficult, and if we can divide the image following a special pattern—in this case defined by the contour of each portion—this will enable us to increase the speed of response for presenting the final results to the user.

Another important and simple approach used in our system is how we translate the size of the food present in the images, in order to create a real life size volume measurement. For this, we use the top and side images obtained by the user to extract the area of a food portion and its corresponding height. With those two images, we can produce an approximation of the volume associated with the food portion. The volume of each food portion becomes the input of the following sections where the application takes the nutrition tables and, based on our calculations, generates the number of calories associated with the food, based on these measurements.

Our proposed solution contains details to overcome most of the limitations explained in this chapter. Note that, in the following chapters, a detailed explanation of the work performed will be given. For this section, an outline of the application is presented, to produce a global concept of the system without digging too deeply into the detail of each section.

This work is entirely focused on the development of an application capable of performing the complete set of steps needed for Image Analysis and Image Processing. The sections corresponding to “Food Recognition” and “Calorie Conversion” have been developed by other students from the same investigation team, and these sections are beyond the scope of this work.

## System Preconditions

To initiate the process of calorie control, the system must contain:

- The patient information.
- His/her thumb image and dimensions.
- The number of calories allowed per week, provided by the dietician.

## Image Capture Process

In order to capture the images by the user, the patient must:

- Locate the dish containing the food in a well-illuminated area.
- Include his/her thumb in the image.
- Position the camera as steadily as possible.
- Capture the scene with the entire dish inside of the image.

## Image Analysis

In the image analysis section, the main features applied to the image inside of the application are:

- Contrast increasing method.
- Colour rasterization and texture correction.
- Accentuation of the main characteristics of the objects.
- Contour extraction.
- Area definition
- Translation of size, from pixels to cm.
- Food recognition

- Presentation of the final results of the segmentation

## User Interaction

Once the image has been analyzed, the processed image is shown to the patient and he/she must perform the:

- User verification.
- User Correction.
- User Approval.

## Final methods Applied

In this section, when the user has made all of the corrections in case of misclassifications, the system will:

- Extract the meaningful information from the portions.
- Combine the information with the nutritional tables.
- Generate the final results of calorie consumption.

## Information Storage and Presentation

As a final stage, the system must complete the process as follows:

- Store the images, data and calculations for further analysis.
- Present the final result to the patient on screen.

## 4.4 Database Proposal

In the schematization of the database, it is important to understand that this section will help us enhance the recognition capacity of the database. The set of images in the database must be sharp, with good illumination conditions, in order to facilitate defining the characteristics and using them to perform further analysis of the images obtained by the user/patient.

Once we obtain the image from the user/patient, the image analysis is started. The contours are defined to divide the whole image in sub-sections containing each portion; with this, we can discover special characteristics for each portion, based on physical aspects such as shape, size and colour, to mention a few. These characteristics can be also part of the characterizations of our images database, and we can subcategorize our database as well. With this, we will be able to perform pattern recognitions using the complete image from the database and can try to locate it in the complete image of the food.

The number of images needed to construct the database is in the order of thousands. This number is not trivial at all, and we must understand that each image extracted from the database and compared with the food image will produce a delay in the processing. This delay per image is normally short, due to the actual capacities for the processing even in smartphones, but when we consider several images for performing the same comparison, the delay increases every time we add a new image to the set of pictures to be compared. Now we face a problem of response and delay produced by our application, but this problem can be minimized by using the set of characteristics extracted from our system per portion, and using them to reduce the sample of images extracted and compared.

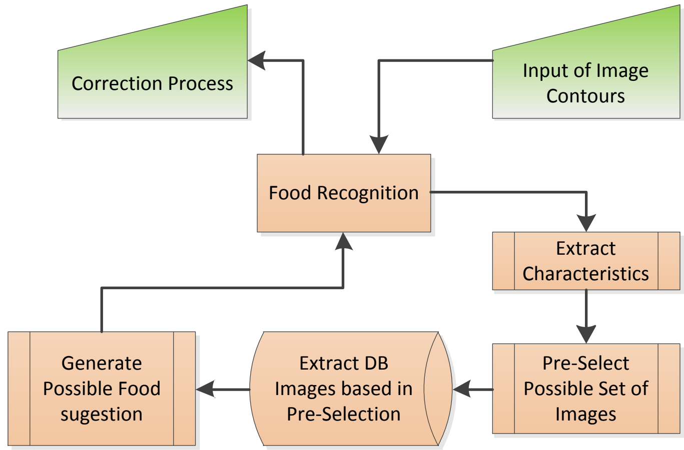



Figure 13: Diagram of food recognition in detail

Figure 13 shows, in detail, the process involved in the database interaction and the use of the images stored in it.

The use of the database of images will increase the accuracy of the food recognition process, for the cases where the complexity of food creates difficulty in the process. In those cases where the portion itself is difficult to define, and its characterization can become diffuse for the application, the comparison of the portion as a whole, with a training image from the database, can allow us to obtain the desired classification, where the normal process is not working properly or is not producing results at all.

# Chapter 5 Proposed System

In this chapter, we are going to discuss in greater depth our proposed system, and how we intend to achieve the goals and objectives defined in the previous sections. Here, we are also going to explain in detail, one by one, the steps followed in the modification process for the images taken by the user and analyzed to extract the corresponding characteristics from each portion of food present inside the images. Once the initial transformations are done, we then focus our explanations on the contour definition analysis, and how this set of implications can be used to perform a positive separation of the food. With the contours defined and the areas of interest separated over the entire picture, then the areas of the objects inside the scene are calculated, and with this, we will be able to specify an image size for each food. Given this point, we must then concentrate on the measurement pattern extraction from the image, and because we are using the patient's thumb, the next section will explain how we manage to extract the skin coloured pixels, and how we use this special characteristic to locate, extract and measure our thumb pattern. With all of the pieces ready for assembly, we then perform the translation from image size into real life size. Then, finally, we only need to present the results to the user, for proper verification and review.

## 5.1 Contrast Increasing Method

Even the clearest images with high resolution can be enhanced to increase the chances of success in an analysis process, to ultimately obtain the values and calculations needed to complete our tasks. One of the commonly used tools is image convolution. With this process, we can produce diverse types of effects over the images, based on the local characteristics of the kernel used, its size, structure and numbers contained in the kernel matrix. The task is not simple, given the variety of images used to perform our evaluation and the different types of smartphones that can generate them. Due to these complications, the task can be done using different procedures; the following are the most common procedures used to complete this type of task. Similarities and differences exist among each when comparing each process and the final results generated. The following are the methods tested and evaluated to produce our final results.

## 5.1.1 Edge Detection with Convolution

Initially proposed by Roberts [43], edge detection can be performed using a double convolution of the image, with two kernels. This double process will produce only one final image. Both kernels are intended to work on the two dimensions of the matrix. To support his theory, Roberts proposed to process the image with a local differential operator, in order to produce a new image that would have the edges highlighted as much as possible. The final image will look like a simple line drawing, following all the possible contours of the objects present in the scene. Roberts's goals were to produce sharp edges, reduce the noise inside the image produced by the background, and finally achieve an image representation of the original picture, close to the human eye perception. He used the following functions to process the images and generate the differential results:

$$y_{i,j} = \sqrt{x_{i,j}}$$

$$z_{i,j} = \sqrt{\left(y_{i,j} - y_{i+1,j+1}\right)^2 \left(y_{i+1,j} - y_{i,j+1}\right)^2}$$

where $x_{i,j}$ is the initial intensity value from the original image and the derivative $z_{i,j}$ is used to store the result of the formula applied. The sub-indexes $i$ and $j$ are the corresponding coordinates for the two dimensional matrix that represents both the input and the output images. In the convolution, the matrix is applied directly to the image, pixel by pixel. The matrix is impressed in a section of the image; to capture the influence of the nearest neighbours of the pixel $\mathrm{i, j}$ analyzed, with this, the new value located in the position of the pixel $P_{i,j}$ will be the result of its own value, plus the influence value of the neighbouring pixels.

This operation showed how the diagonal changes over the image intensity are highlighted; due to this diagonal condition, the same results can be obtained with the convolution of the image using the following two kernels:

$$\begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & +1 \\ -1 & 0 \end{bmatrix}$$

## 5.1.2 Convolution with directional kernels

The previous method was enhanced using a two kernel convolution [44], but applying the combination of two vectors, in horizontal and vertical directions, over the matrix. The convolution is applied to the sections of the original image that includes the pixel analyzed $P_{i,j}$, using a similar method than before the process computes the information taking into account a bigger neighbourhood and the horizontal and vertical results are computed separately. With these changes, the gradient of the intensity can be highlighted with better results. The changes in this case are evaluated with two kernels—one will work for the gradient of $x$ and the other for the gradient of $y$; in this case $G_x$ and $G_y$ respectively. The original image $I$ is convolved separately for each gradient:

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} conv\ I$$

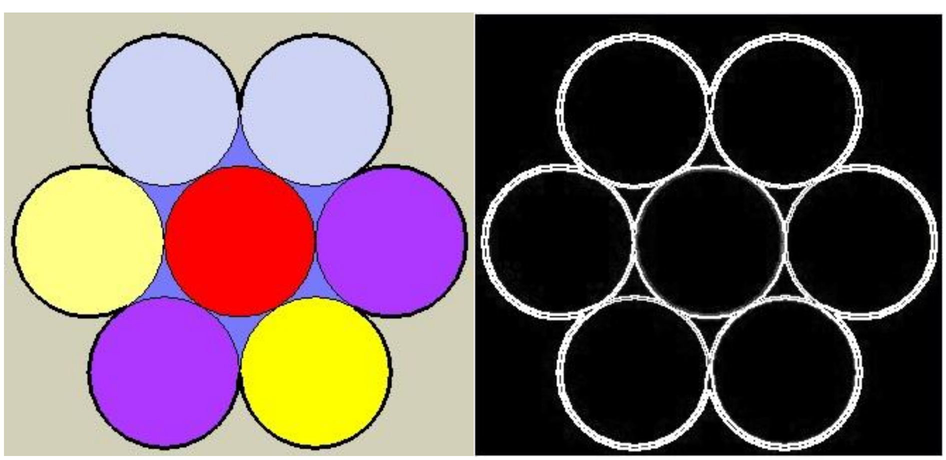
$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} conv\ I$$

This will generate two intermediate images, with the gradients in both directions highlighted; then we need to generate an average of both to produce the output of the procedure:

$$G = \sqrt{G_x^2 + G_y^2}$$

The resulting image $G$ will contain a good result, with the edges highlighted by the directional changes in the gradient of the intensity and with a good resistance to the noise produced by the background.

Figure 14 shows an example of the original image used to perform the 2 directional kernel convolution. This was generated to have different paths to produce similar results, and finally to opt for the most suitable procedure, or set of procedures combined, to generate the most satisfactory results.



a) Original Image b) Image of the edges

Figure 14: Images sample of convolution applied with two directional kernels

### 5.1.3 Convolution with a Gaussian Operator

Developed by Canny [45], image convolution using a Gaussian operator $G_n$ extracted from the first derivative of a two dimensional Gaussian distribution performs good edge detection, using the same $\sigma$ as the first Gaussian derivative. It is similar to the previous procedure exposed in this work, but it appears to be less expensive in computational time, and the noise reduction from the background it is also better with the approach used for this method. The Gaussian operator in any direction $n$ is given by:

$$G = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

Considering the directional changes $\partial$ in $n$ we can use the previous formula to compute the gradient generated by these directional changes:

$$G_n = \frac{\partial G}{\partial n} = n \, . \nabla G$$

Because the direction is not known, it is important to apply a method to generate the estimation of the direction; a smooth gradient operation can be obtained using the next formula, this will generate a good aproximation:

$$n = \frac{\nabla(G \; conv \; Image)}{|\nabla(G \; conv \; Image)|}$$

The edges, by definition, are local maximums in a direction n. To obtain this local maximum, the relationship is described by the following formula:

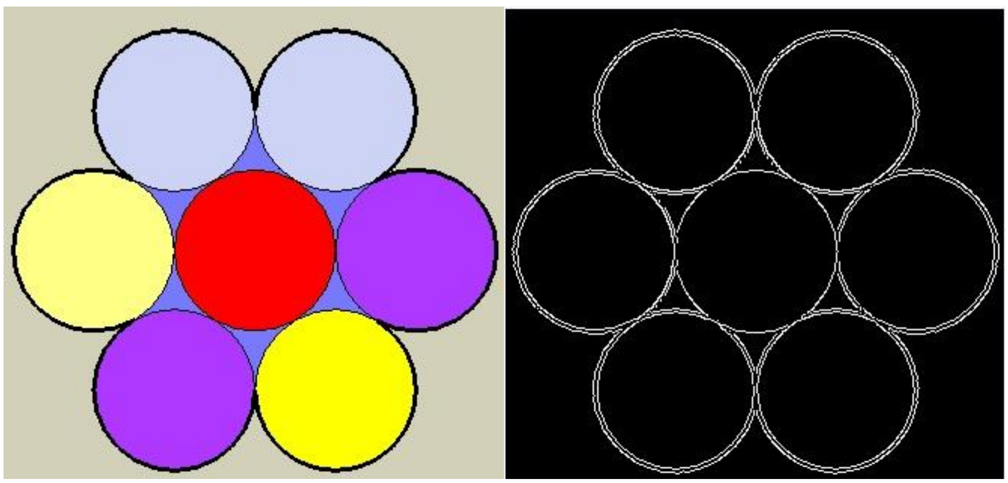
$$\frac{\partial}{\partial n} G_n \; conv \; Image = 0$$

By the substitution of $G_n$ in the previous formula, we can define a new combination as:

$$\frac{\partial^2}{\partial n^2} G \; conv \; Image = 0$$

Finally, the magnitude of the gradient will be given by:

$$|G_n \; conv \; Image| = |\nabla(G \; conv \; Image)|$$



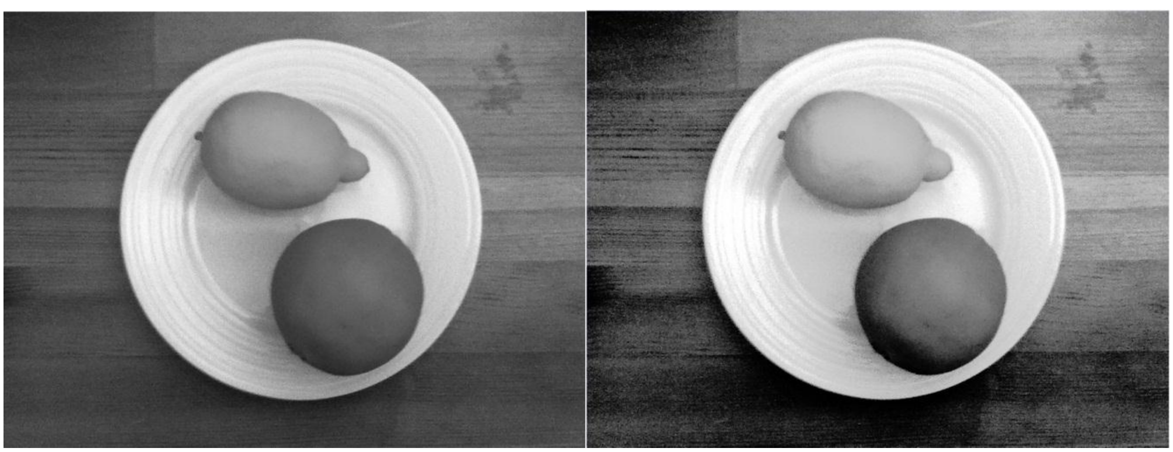
a) Original Image b) Image of the edges

Figure 15: Images sample of a Gaussian convolution applied

Figure 15 shows an example of the original image used to perform the Gaussian convolution, and the result of the procedure, in image a and b, respectively. Note that if we compare this approach with the previous one, the edges here are sharper and more defined, for the inner and the outer contours of the circles.

## 5.1.4 Image Equalization with Histograms

One of the options we have at hand to perform a previous correction, in order to increase the contrast applied to the image, is image equalization. This changes the image brightness to help us reproduce a new image to analyze with a higher dynamic range of intensity between the objects inside the scene. In the end, this can help us gain differences over the edges of the objects, and to include the soft edges present in the picture. In the proposed method [46], the idea is to generate a histogram of the image that will characterize the brightness of the complete set of pixels in the image. The plot of the histogram for the original image normally shows how the brightness distribution is depicted inside the image. The differences appear over the entire image, with sections as peaks in the histogram as the values of regular bright values present, and the valleys for the sections of the image with the not so common bright values attached to those pixels. The distribution of the brightness is the equalized so that every possible value has a similar presence over the image, compared with the other values. The final result normally shows an image where the edges appear sharper and more defined.



a) Original Image b) Histogram Equalized Image

Figure 16: Images with histogram equalization

Figure 16 shows the original image in a, and the result of the image equalization in b. Note how the final result can produce a sharper image, and finally the edges over the objects can be signalized with ease inside the computational analysis.

## 5.2 Colour Rasterization and Texture Correction

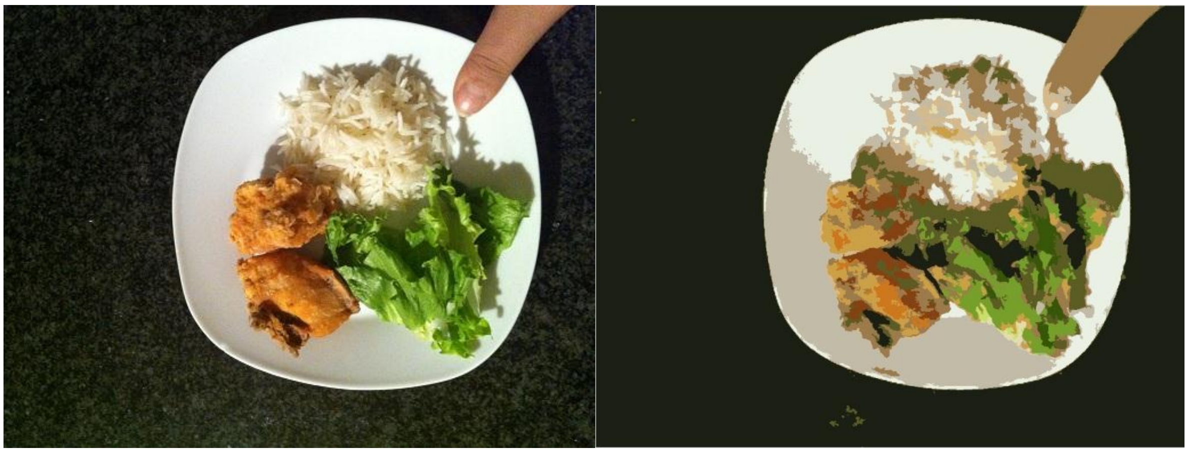
A set of steps must be followed in order to achieve a colour rasterization task. This process implies that the images must be divided in half several times. For this, an initial transformation must be performed, and then the application can proceed with the reduction and rasterization in the different levels of abstraction, to finally reproduce and enhance only the most important characteristics found in every level of the analysis.

### 5.2.1 Pre-Process and Image Modification

The modification of Burt’s algorithm needs initial preconditions for the image in order to apply the general analysis and reproduce the rasterization of colour. To complete the analysis, a simple transformation must be done over the image, where a standard size for the input image must be defined according to the level of analysis one wishes to implement. Once the initial image size has been computed, a change in the size of this initial image must be applied, and the procedure and analysis must be performed with the new sized image. Note that more levels of analysis will produce more exact resultant images, but of course this implies that for every level, a division of the original image into a new image half of its size must be computed, and when the height or the width, or both, are not divisible by half, the procedure will generate errors in the process.

### 5.2.2 Application of the Colour Rasterization

Once the number of levels to be analyzed has been defined and the input image has been adapted to apply the levels, the rasterization can take place, applying the Burt’s algorithm. This generates a set of images to reproduce a pyramidal analysis, where the connected components in each level of the pyramid are calculated and passed as parents of the next level of analysis. This will demarcate the most common characteristics of colour, texture and illumination over the objects present inside the picture. In the end, an image will be output with a smoothed surface, average colour over the objects, and minimal changes and noise eliminated.



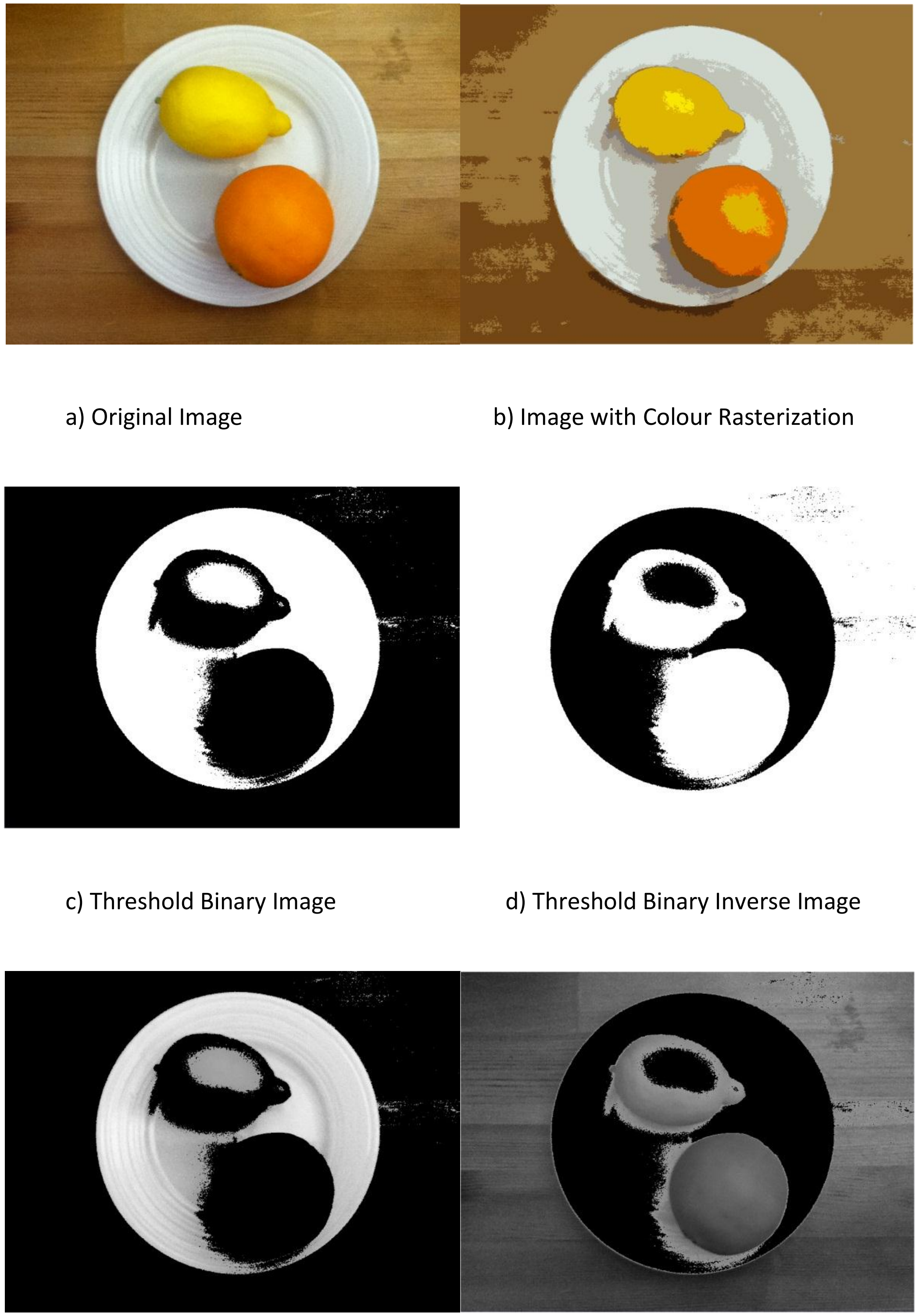
a) Original Image b) Image with Colour Rasterization

Figure 17: Colour Rasterization Images

Figure 17 shows an example of the Burt's algorithm analysis, applying a pyramid with $4^{th}$ levels of depth. Note in image b how the colour and the texture have been smoothed, and small differences over the images have been eliminated, in order to focus our effort on the main characteristics.

## 5.2.3 Accentuation of the main characteristics of the objects

After the application of the pyramid analysis, the resultant image can be used to perform the localization of the main changes over the image with a simple modification of the image, in order to accentuate the important changes over the image, and its particular characteristics. For this section to be developed, we need to move the output image obtained from the previous analysis into a two level image, in order to analyze the changes in intensity of the colours and the objects.



a) Original Image
b) Image with Colour Rasterization
c) Threshold Binary Image
d) Threshold Binary Inverse Image
e) Threshold Binary Avg. Image
f) Threshold Binary Avg. Inverse Image

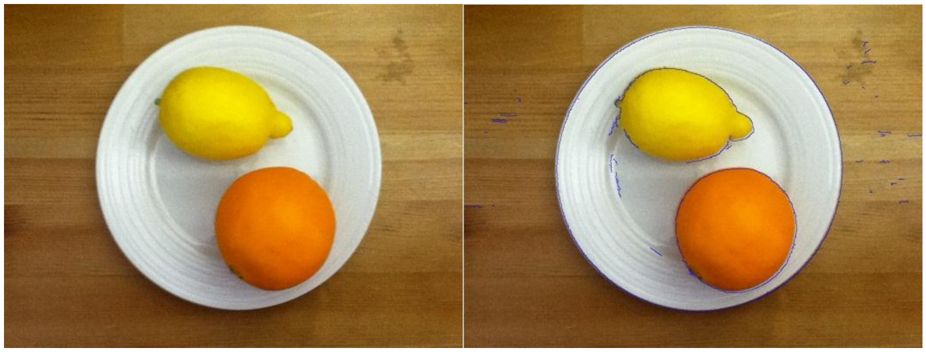
Figure 18: Characteristics of accentuation stages

Once the initial image has been transformed into a binary image and is modified, a special selection of pixels over the image is produced for the analysis of the brightness values from the original image. This concept will help us demarcate the foreground as the sections of the image that may contain regions of interest for the next step in the analysis, and disregard the rest of the image, considered as background. This section of the analysis is fairly simple, but relies in the previous analysis to perform adequately and to produce meaningful data.

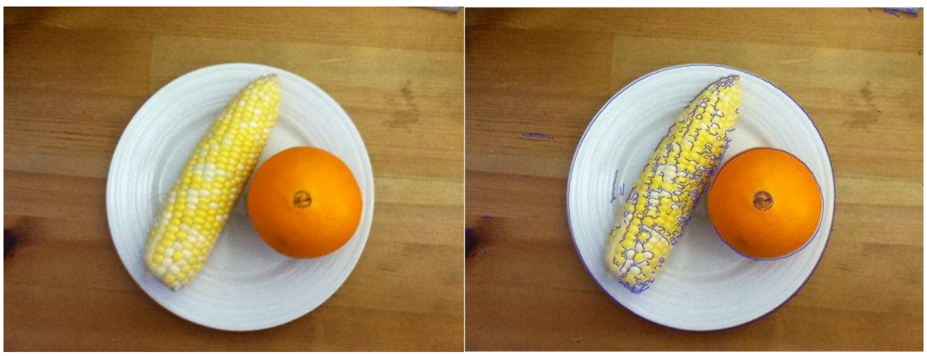
Figure 18 shows four different types of analysis used to enhance the images and reproduce the most important characteristics to be followed in the edge extraction. The different stages were tested to select the procedure most closely related with the quality needed for our application.

## 5.3 Contour Definition

At this point of the image segmentation and analysis, and based on the characteristics demarcated by the pre-processing sections, we simply need to follow the active contours defined by the sharpness incremental procedure and the enhancement of the meaningful characteristics of the image. The extraction of those contours from the image will give us a complete set of pixels with its corresponding coordinates in the plane, which depicts the simple shapes of the objects present inside the scene. Once the set of pixels has been defined, we can recreate the contours artificially, with strong colours such as red, green or blue, just by following these pixels and drawing lines between them.



a) Original Image 1　　　　　b) Contour Image 1



c) Original Image 2　　　　　d) Contour Image 2

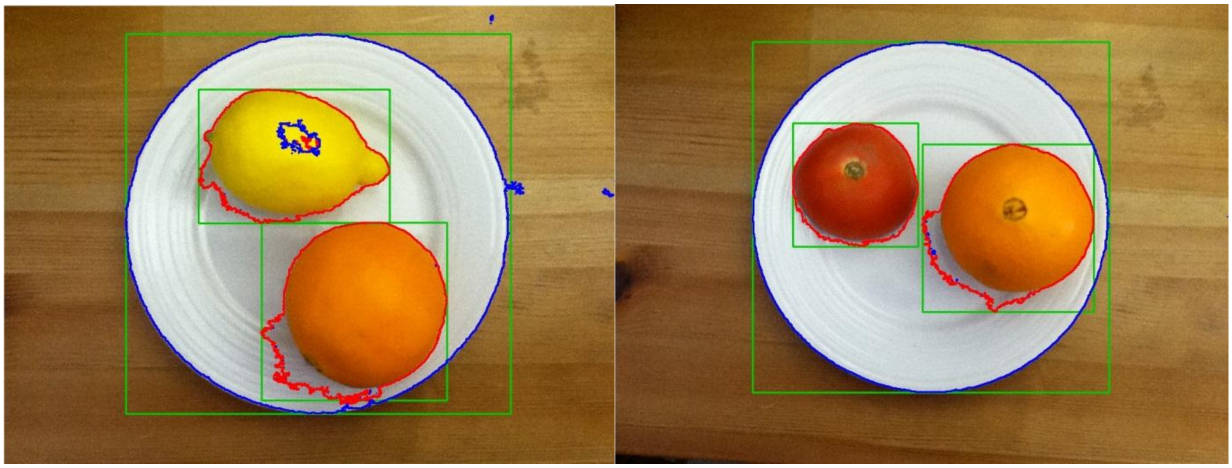
Figure 19: Contour Definition Images

Figure 19 shows two examples of the contour definition, with blue colours over regular images using the entire set of analysis.

## 5.4 Area Definition

### 5.4.1 Region of Interest (ROI) Definition

The first step at this point is to extract, one by one, the objects present inside the image. The objects have been defined in the previous steps, and respective contours have been depicted. Using the images with the contours defined as starting point, we can go directly to each contour defined over the image, analyze its structure and create a specific definition of the different regions of interest, based on the object's distribution over the image.

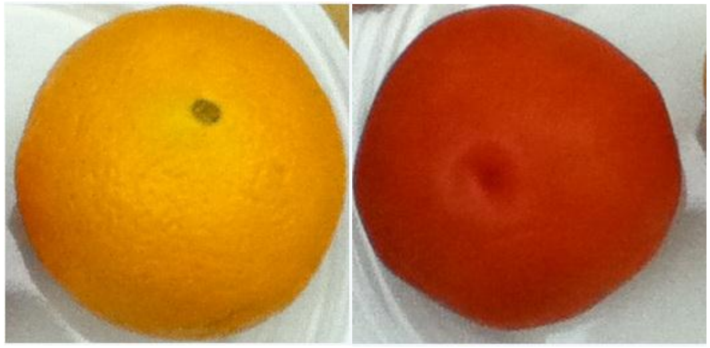


a) Image 1 with Contours and ROI Defined b) Image 2 with Contours and ROI Defined

Figure 20: ROI definition Images

Figure 20 shows how the different regions of interest have been defined using green rectangles to mark the areas where the different objects are located inside the scene. These areas will then be used to extract each section one by one and perform specific size analysis over each object.

The division of the image into sub-images will allow us to focus our efforts, section by section, and work separately on each object of the scene. This will reduce the chances of errors generated by the shared configurations and characteristics between objects. Figure 21 shows two image examples of two objects extracted from the scene.

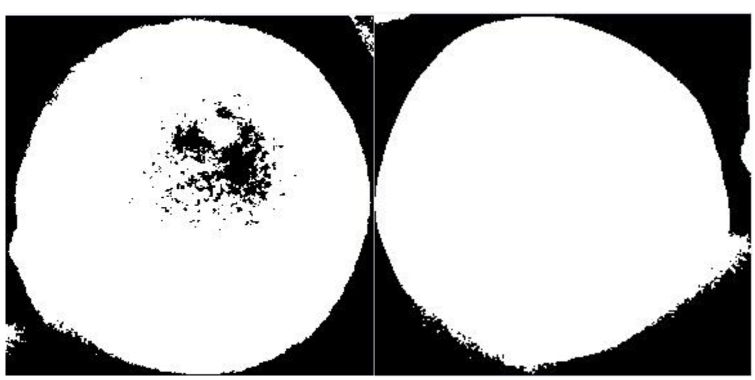


a) ROI 1 extracted b) ROI 2 extracted

Figure 21: ROI images extracted from the original image

### 5.4.2 Calculation of the Binary Large Object (BLOB)

The definition of Binary Large Objects (BLOBs) in these sub-images extracted with the ROI definition is a simple step used to transform only the big object present in each portion of the image, separated by the ROI, and transform it into a binary representation of the object. The sub-image is then processed to obtain a binary image form, and used to focus on the area with the colour of interest from the object, moving white pixels to those positions, and black pixels to the rest of the image.



a) ROI 1 extracted b) ROI 2 extracted

Figure 22: BLOB Definition Images

a) ROI 1 extracted b) ROI 2 extracted

Figure **22** shows the two objects extracted and transformed; it is easy to understand at this point the reason why these types of images are called Binary Large Objects, due to the ratio in size between the object and the image size.

### 5.4.3 Extraction of the Size in Pixels

Using the images extracted from a) ROI 1 extracted b) ROI 2 extracted

Figure **22**, the number of white pixels is counted, and with this number finally generated with a high level of certainty, we can express the size of each object inside the scene, using its size in pixels. In the following section, the size in pixels will be used, and the measurement pattern of our experiment takes a leap of relevance into the whole process.

## 5.5 Size Management and Calculations

### 5.5.1 Skin Colour Section Definition

In order to obtain the sections of the image corresponding to the patient's thumb; a special transformation is performed over the image, to move from RGB colour space to YCbCr colour space. With these changes over the image, the analysis for the specific range of colours where the skin of a person can be located is applied to the images. The sections corresponding to pixels that fall into the category of skin colour are kept in the image, while the other pixels with other set of colours different than the special range are removed, and black pixels are located there instead.

a) Original Image 1 b) Skin colour 1 extracted

c) Original Image 2 d) Skin colour 2 extracted

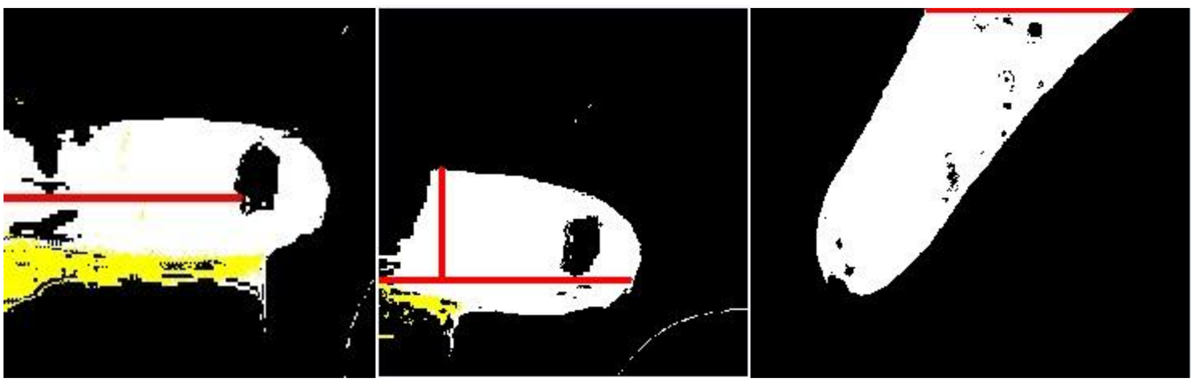
e) Original Image 3 f) Skin colours 3 extracted

Figure 23: Images of skin colour extraction

Figure 23 shows a set of original images (a, c, e) and the extraction of the pixels corresponding to the skin colour (b, d, f). Sections from the objects corresponding to skin colour, which are not real human skin, also appear. To correct for this, a shape and size recognition is applied to disregard the non-important sections.

## 5.5.2 Thumb Dimension Extraction

Using the images with the human skin highlighted, the definition of the section of interest containing the thumb is extracted. The extracted area is then used to produce the BLOB corresponding to the thumb itself. Finally, when the BLOB of the thumb is available, the dimensions of the thumb in pixels are extracted. Even when the BLOB definition transforms the image into black and white pixels, we have been working in a 3-channel image, to be able to use special colours to signalize the dimensions of the thumb.



a) Thumb BLOB 1 b) Thumb BLOB 2 b) Thumb BLOB 3

Figure 24: ROI and BLOBs of user's thumb images

Figure 24 shows the BLOBs of the thumb from 3 different images. Note that before the BLOB extraction, the ROI of the image has been defined per image in order to focus the processing procedure only in a small area of the whole picture, to reduce processing overhead.

### 5.5.3 Transformation from Pixel Size to Real Life Size

With these measurements in pixels, combined with the real life size of the thumb stored inside the application, the system can make a simple number transformation to obtain the real life size of the thumb and its corresponding ratio in this particular picture.

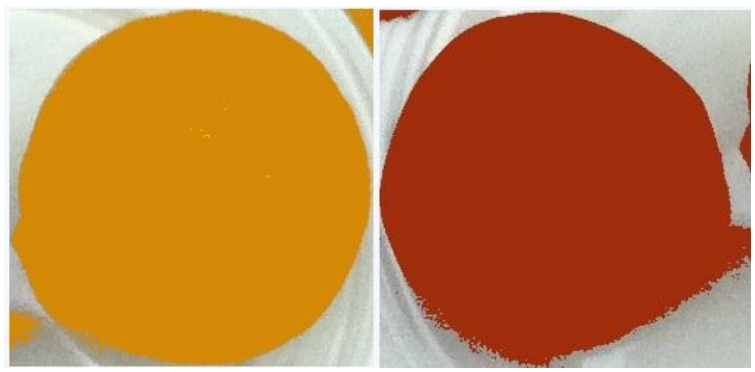
The preliminary results obtained are needed in this case to use the thumb of the patient as a measurement pattern. With both measurements defined, in pixels and centimetres, the area defined in section 4.4.3 for each portion in pixels can now be transformed into the real life size. This is a simple transformation, but is a critical section of the application due to its importance for the calorie translation and food daily intake measurement.

## 5.6 Food Recognition

For the food recognition part of the application, a support vector machine has been applied using characteristics extracted from the portions to generate the initial information needed to run the SVM. The information extracted from the portions is natural characteristics defined as a set of conditions and classifications for each portion. In this case, this is the average of the colour of the portion, the shape that the food represents and its size, where 3 characteristics are taken into account to feed the SVM. Initially, the results were evaluated with only one feature, which was colour, and the SVM gave us a result with 72 percent accuracy. Once this first stage was completed, a second test was performed with more than one feature. For the second test, the SVM used the complete set of properties, which are colour, shape and size. The three recognition process performed well and the final results were adequate. **Error! Reference source not found.** shows that the accuracy of our images in the recognition part was approximately 89 percent.

Table 2: SVM experimental results

| Applications | Training Data | Testing Data | Features | Classes | Accuracy |
|---|---|---|---|---|---|
| Fruit & Food | 120 | 60 | 1 | 3 | 72% |
| Fruit & Food | 140 | 60 | 3 | 3 | 89% |



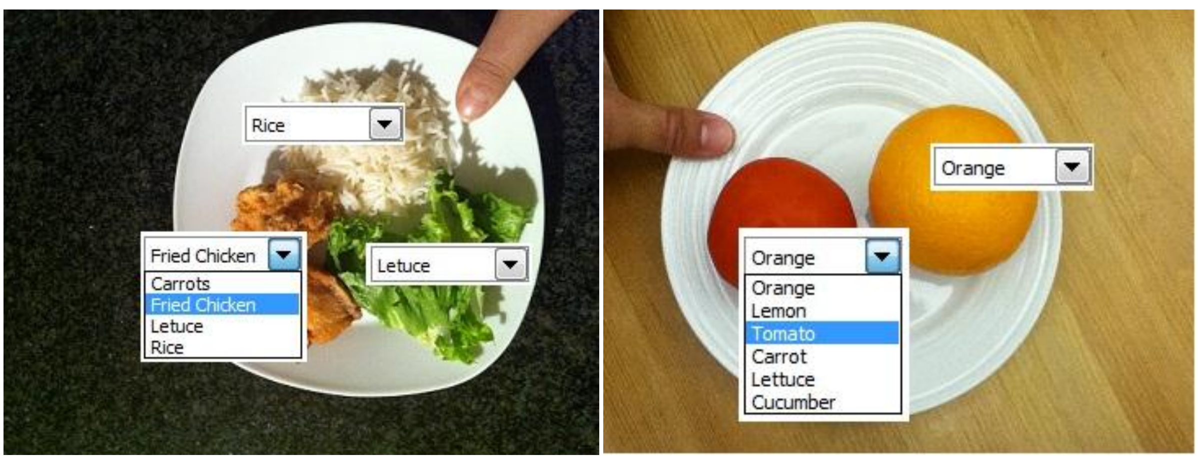
c) Orange ROI with Average Colour  d) Tomato ROI with Average Colour

Figure 25: Average colour Images

Figure 25 shows an example of the average colour computed from the portion definition. The images are initially transformed to a HSI colour space, in order to work with the hue value that represents, in this case, the colour of the objects. The value is then extracted and evaluated with the database configuration to define the range where the number of the colour is located, in order to identify it and feed the SVM procedure.

## 5.7 Presentation of the Results to the User

At this point, we have shown how the system develops and processes the image, extracts and analyzes each portion, extracts special characteristics from them and uses them to perform a decision making process based on special those characteristics. The size of the portions has been computed to translate then into real life size. The only section left is to enable users to verify the complete process, so they can verify the type of food suggested by the system. Once the verification step is done, the application performs a final step, which is to look for the calories related to the portions present in the images. Based on the size calculated per portion, the final number of calories is then computed and stored with the set of images, to enable the dietician to perform *a posteriori* analysis of the daily caloric intake. Figure 26 shows two final images presented to the user, with the results of the food recognition process in it. With this, the user can accept the pre-selection made by the recognition system, or correct the name of the food if a misclassification has occurred.




a) Portion Name Image 1 b) Portion Name Image 2

Figure 26: Final Result images presented to the user

# Chapter 6 Prototype Implementation

## 6.1 System Operation

To start the operation of the application, the user of the system is asked to take a picture of the food with his/her smartphone, before and after eating, to compare the sizes of the portions before and after the food intake. The system will then process the images of the food to detect different types of food and their respective portion sizes.

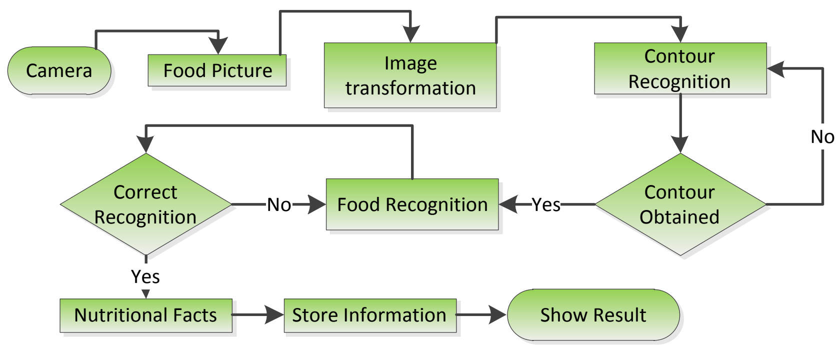
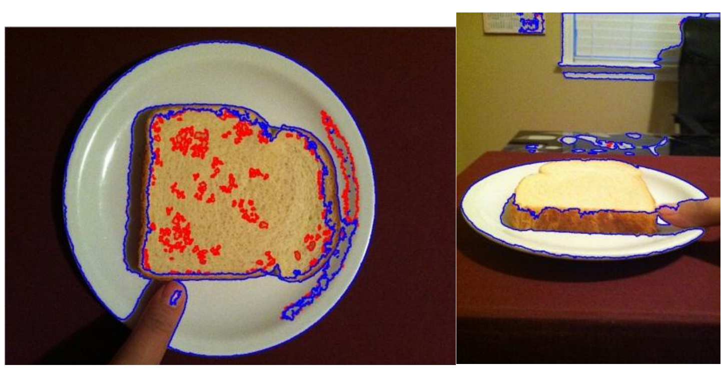



Figure 27: Diagram of the Nutrient Intake Monitoring System

Figure 27 shows the diagram of the proposed system and how the interactions with the user will react and perform over the complete process. The diagram specifies the simple steps followed by the application: once the picture of the food is captured by the user, the image is transformed and prepared to the next step. The contour recognition process then starts, and once everything is set and separated in portions, based on the different contents and objects inside the scene, the system performs the food recognition. At this point, the user will receive the response from the application, with the image taken and the first analyses results, for the user to do the corresponding corrections in case of misclassifications, or the acceptance of the suggestions made by the food recognition process.

The system is now almost ready to perform the calorie calculations. At this point, the user has reviewed the system response and corrected any given misclassification. This gives us a level of certainty to use the portions defined by the application and to extract the areas for each portion. We then make use of the second image of the food: the side image will help us define the height of each portion, and transform the area of the portions into a volume measurement. The thumb of the patient must be present in all images, due to the need for the application to contain a measurement pattern to translate from pixels to centimetres. Once the height is obtained, the area can be simply transformed from $cm^2$ to $cm^3$; this is later used to combine the results and obtain the approximation of the calories and nutritional facts for each food recorded. Figure 28 depicts two sample images, showing the top and the side view of the food to be consumed by the user. These two images of the same food will help us perform volume calculations and generate the results for the following steps of the calorie calculation process.



a) Top View Image b) Side View Image

Figure 28: Sample Images of Top and Side Views to Calculate Volume

Error! Reference source not found. shows the calculations performed by the application to extract the dimensions from two different food portions present inside the images. The volume calculated, combined with the approximated density related to each type of food, helps us to obtain the weight in grams for each portion. In the final column of the table, the real average of weight is then shown; note that our calculations are between the lower and higher thresholds.

Table 3: Volume Calculations vs. Real Volume

| | System Calculations | | | | | Real Life Data |
|---|---|---|---|---|---|---|
| | Diameter | Radius | Volume | Density | grams | Average Weight |
| orange | 6.7 | 3.35 | 157.4791385 | 1.8 | 283.4624494 | Between 255-300 g |
| tomato | 6 | 3 | 113.0973355 | 1.4 | 158.3362697 | Between 120-170 g |

As a final stage, the application retrieves the nutritional information associated with the type of food present inside the image, and after the simple calculations that take into account the calories in the database and the real size of each portion, the information is stored and the final calculations and results are presented to the user.

## 6.2 System Workflow

Three major branches of workflow operate inside our application: the first one is the interaction related with the patient and the application, based on the daily basis usage of the application. The second interaction is the review of the stored data to analyze historical data provided by the patient and its daily use. The final workflow is the dietician interaction with the patient and the system.

### 6.2.1 Daily User Interaction

The daily user interaction workflow has to do with the most important part of the application, because it is the section of the system where the information about the food consumed is obtained.

This is the reason that our application exists, because we want to aid the dietician and the patient in the calorie consumption record.

Figure 29 shows the workflow diagram of the system from the view point of the system. As indicated by the figure, the user should first take a picture of the food with his thumb in the picture. After the picture is taken and analyzed by the system, note that, even when our application performs a food prediction, due to the several conditions that can mislead the application to an erroneous recognition, the application enables the user to perform the appropriate corrections. With the prediction, our intention is to ease the path of usage for the patient, but we also intend to produce a system with high accuracy, which is why the correction section is kept inside this flow.

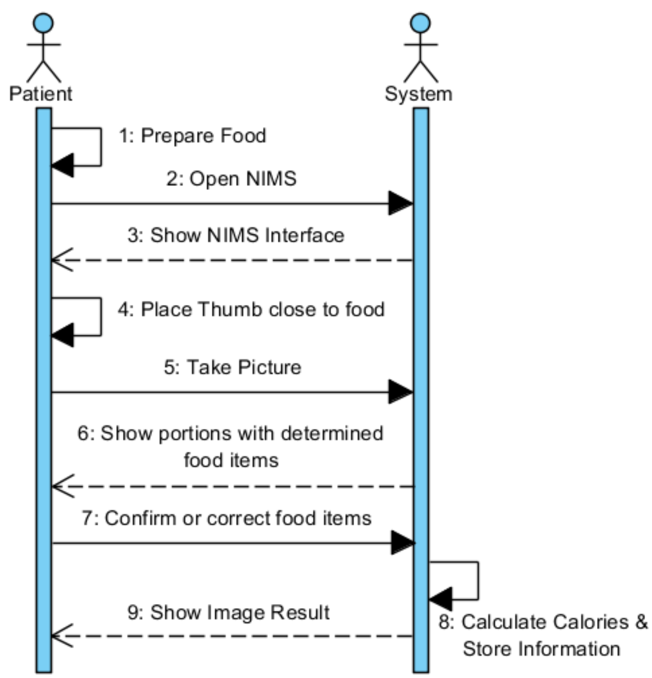



Figure 29: User's interaction with the system

### 6.2.2 Review of the Historical Data

As part of the application, the verification of historical data is an important step for both the user and for the dietician. The historical data can give patients the opportunity to check at any time, their daily consumption, and how their performance is reflected over the entire diet plans proposed by their dieticians. On the other hand, for the dietician, reviewing these historical data is important because this verification will aid the doctor in the decision-making process, to perform changes and redesign the diet plan to promote the patient's health.

Figure 30 shows the form in which the user or the dietician can interact with the system to extract the historical data. The user should also be able to view the results in the form of a bar chart or a pie chart for a specified interval and compare these data to his allowed value of nutrition intake. All of this interaction is considered due to the facility of the application to extract the information using different time frames to proceed with the evaluation.

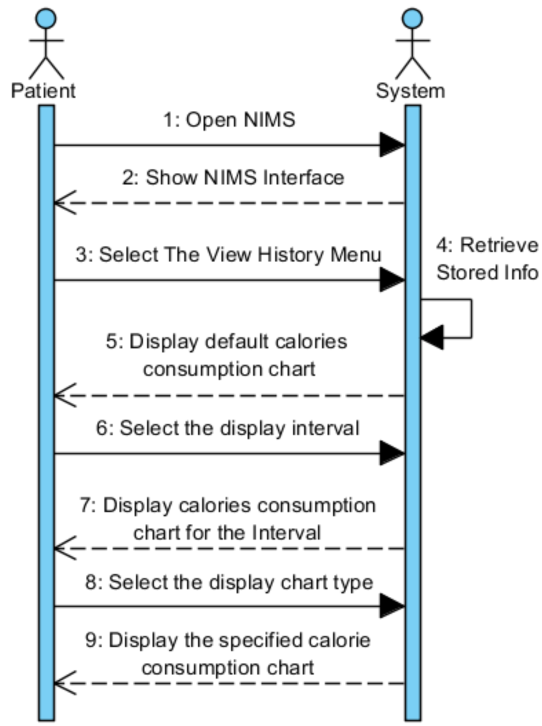



Figure 30: Viewing the nutrition intake record

### 6.2.3 Doctor and Patient Interaction with the System

The initial section of the system is the initial configuration, where the doctor will collect all of the information from the patient to verify the initial conditions for the patient: weight, size, body weight distribution and measurements, to mention just a few of the complete assessments performed by the dietician. Once all the information needed is obtained, the dietician proceeds to design the diet plan for the specific patient, and this is part of the input information needed for this flow over the application. The natural use of the application is then considered, to generate data for *a posteriori* analysis. Once sufficient information is accumulated, the dietician can perform the analysis after a specific time frame, and the application will enable the dietician to generate information from different points of view, in order to make better decisions in the treatment process. Figure 31 shows the diagram of the dietician interaction with the application.

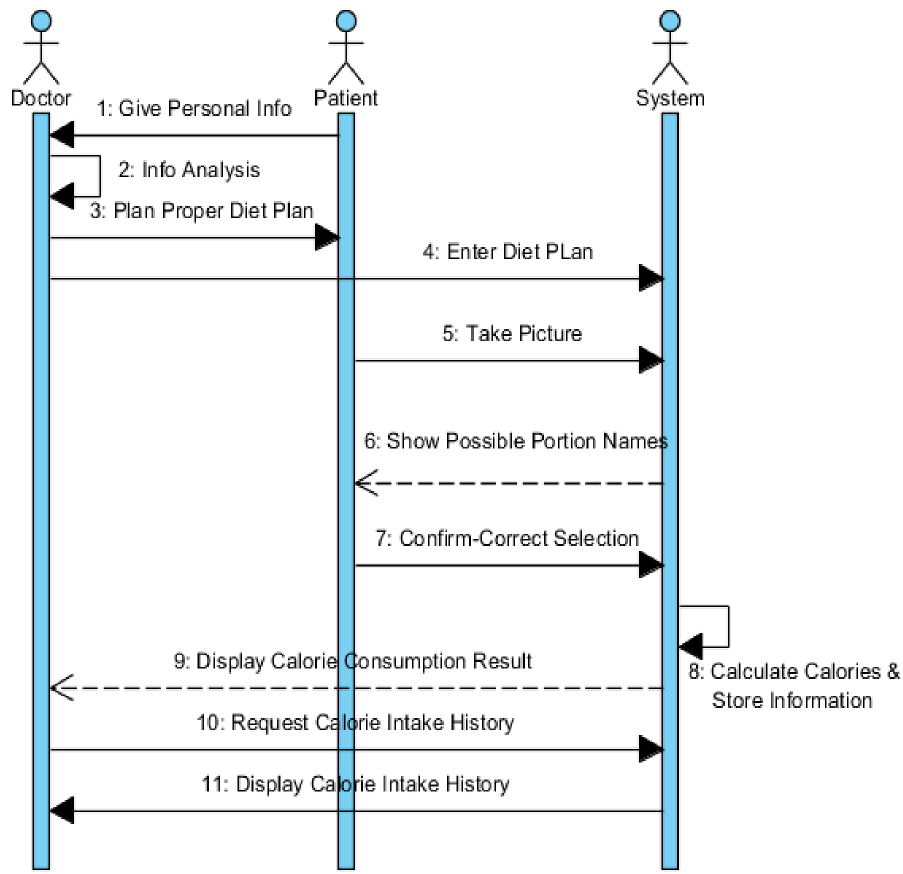



Figure 31: Dietician Interaction

## 6.3 Graphical Interface

We also developed a special interface for the user's application, using the Macintosh Software Development Kit (IOS SDK), to present the actual design for our current application, and how the user will navigate over the different screens of the system to perform all the tasks proposed for this system. The system is then designed to work over the IPhone or the IPod Touch 4$^{th}$ generation, and up.\

## 6.3.1 Main Menu

As usual, our application is managed by a main menu, where the user can reach all of the system's functions. The different buttons will guide the user as well as the dietician through the application, to allow them to perform the regular tasks expected from each participant in the treatment. Figure 32 shows a snapshot of the user interface of the system developed in IOS SDK, which is the user interface of the standard software development kit for IPhone applications released by Apple Inc.

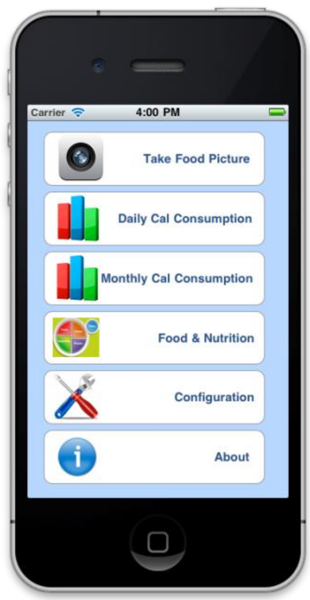



Figure 32: The user interface of the system

## 6.3.2 System Configuration

The dietician must enter the patient information as well as the information regarding the diet plan of the patient and allowed calorie intake. **Error! Reference source not found.** shows the interface where the doctor enters the patient information. This screen also provides access to the screen where the patient captures the thumb picture due to the similar type of data; in this case, the pre-configuration information needed to run the application correctly.


Figure 33: The interface for entering the patient information

### 6.3.3 Parameterization and Measurement Process

Included as part of our proposed system, we have a practical approach for the measurement pattern. It introduces the translation of image size to real life size; in this case, the thumb of the patient is the measurement pattern proposed. The thumb use is expected to be easy because the user does not need to carry a specific artefact that can be forgotten at home or another place, or an external object that will become a burden to carry around just for the sake of the system functionality. In this case, thumb is more flexible, stable and controllable.

Figure 34 shows the screen for the thumb calibration inside the application. The dietician or the patient will also be able to introduce the dimensions of the thumb in centimetres or inches, to perform the corresponding size translation.

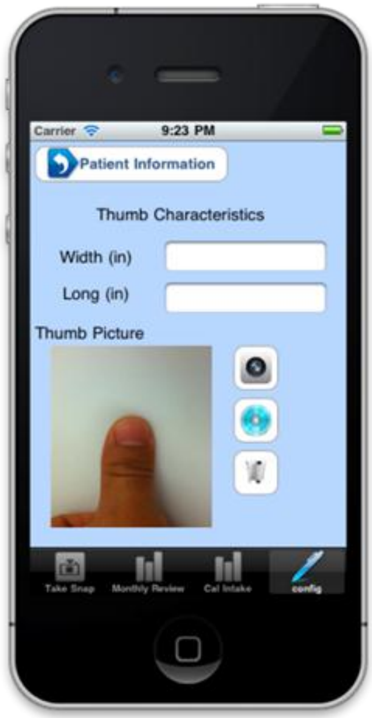



Figure 34: Thumb calibration Image

### 6.3.4 User Results Presentation

Figure 35 shows the technique of the photo capture using the thumb. After the picture is taken, the system starts processing the image to detect the plate and it segments several types of food that might be on the plate.

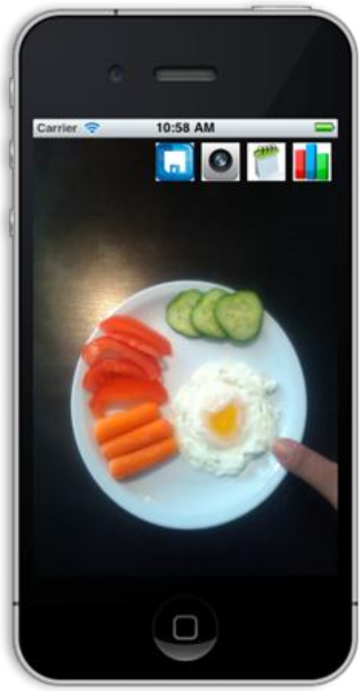



Figure 35: A Snapshot of the system user interface showing the user ready to take a picture of the food, in line with his thumb.

Figure 36 shows a sample of this type of selection process by the user. Over time and using appropriate computational intelligence techniques, the system can improve itself in determining what food is in the picture.


Figure 36: The system allows the user to define the type of food with default values automatically determined by the system

## 6.3.5 Statistical Data Presentation

In the statistical information presentation and analysis, the system has the capacity to present the information using different schemes. Based on the type of information required, the design of each screen has been adapted to fit into the patient's and dietician's needs.

Figure 37 shows how users can see a graph of the consumed calories for a specific interval and compare it to their allowed calorie intakes.

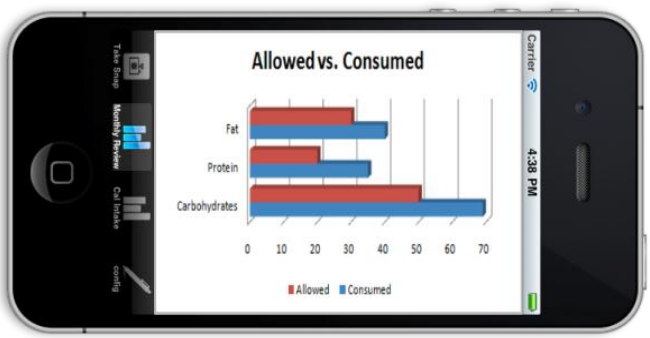



Figure 37: Intake calories and the allowed value

The different types of information must be presented in diverse arrangements, in order to make the information easy to read, understand and analyze. Figure 38 shows how the system can display the results in different formats as requested by the user; in this case, as a pie chart of the nutrition intake during a specified interval.

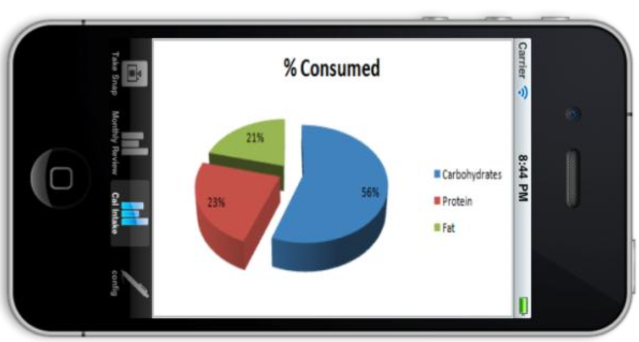


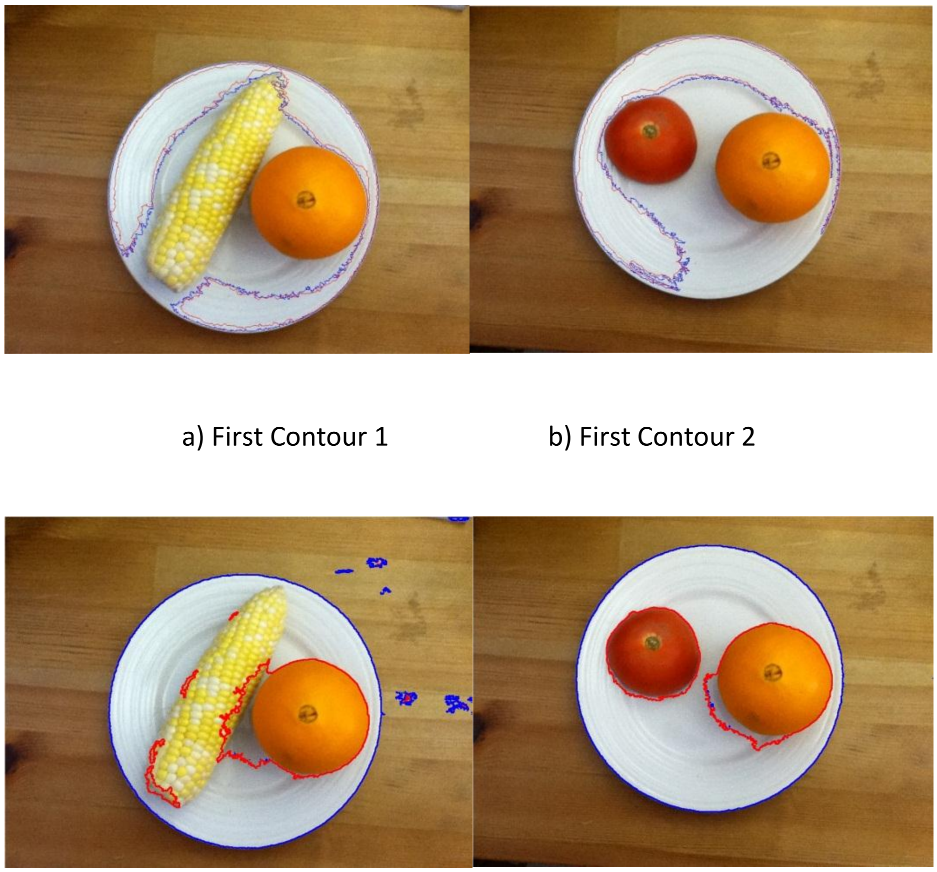

Figure 38: Pie chart of the nutrition intake

# Chapter 7 Experimental Results and Evaluation

For the development of our application, we tested different approaches in order to define the best path to follow. The images used to test the different options available were obtained with similar devices to the ones that the patient will be using, in order to produce similar conditions and generate a better analysis. In the image segmentation area, different techniques are available to perform the object extraction: one of the first approaches is the semi-automatic contour definition, but this option was completely rejected right at the beginning, due to the high interaction needed from the user to perform this process. Another method analyzed in the process was the Watershed Transformation; this method is based solely on gradient analysis, where the pixels with the higher gradient intensity will define the boundaries of the objects, and the pixels where a gradient converges to a local intensity minimum are considered pixels from the same segment. This method was rejected due to the poor results obtained in our initial tests to introduce this image analysis process inside our system.

The method of snake definition, where the contour moves over the images in order to approximate the shape of the images, was also tested, but several tests only produced unsuccessful contour definitions.

Figure 39 shows images a and b as an example of the inappropriate results obtained with the images tested in the initial stages of our application. The better results obtained with our final procedure are shown in images c and d.



a) First Contour 1 b) First Contour 2

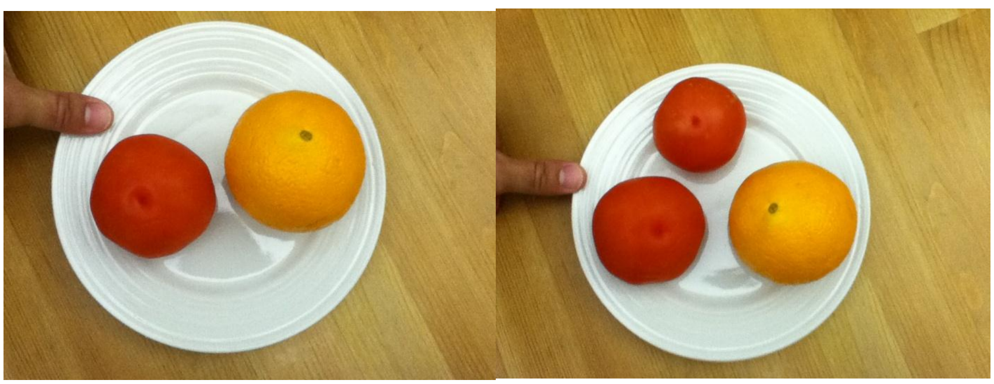
c) Contour selected 1 d) Contour selected 2

Figure 39: Image comparison of approaches

The following sections will explain the method followed by our application, in order to produce the results shown in the images over the sections.

## 7.1 Set of Images Used to Perform the Analysis

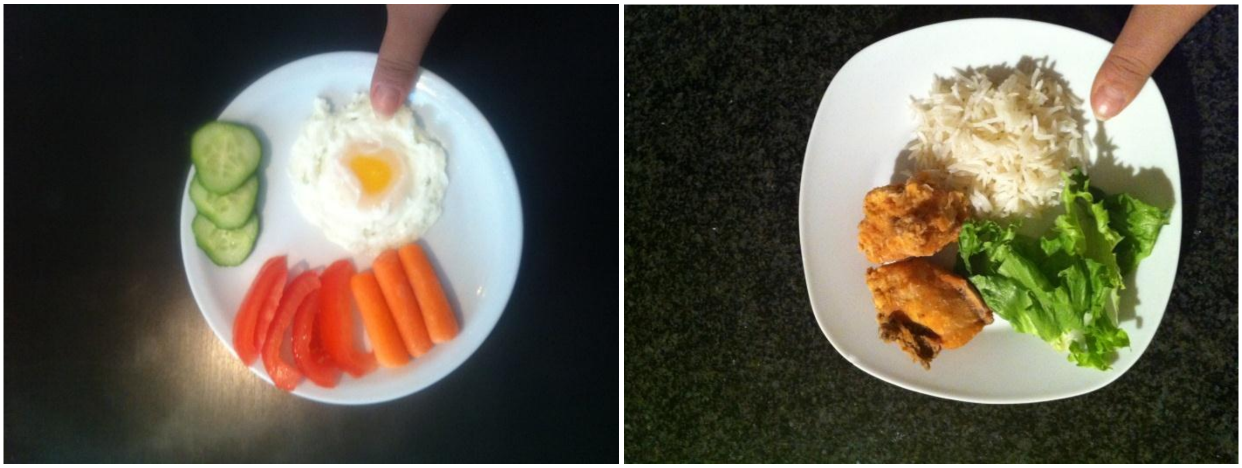
The set of images used to realize the development of the application has been obtained from real food dishes, and in all the cases an IPhone 4 or an IPod Touch 4 has been used to obtain the snapshots of the complete set of test images. This ensured that we worked on images produced with portable devices similar to the devices the real patient will be using for the regular execution of the application. Figure 40 shows two sample images with simple shaped food inside the dish, used to test the performance the application. This kind of image, due to the simplicity of its structures, shapes and colours, produced high quality results, and approximations close to reality.



a) Simple Food Image 1 b) Simple Food Image 2

Figure 40: Images to test the application with simple food

Figure 41 shows two sample images of complex configurations of the meal pictures. In this type of food presentation, due to the higher complexity and variety, the quality of the analysis and final results decrease if we compare the results with the simple set of images.



a) Complex Food Image 1 b) Complex Food Image 2

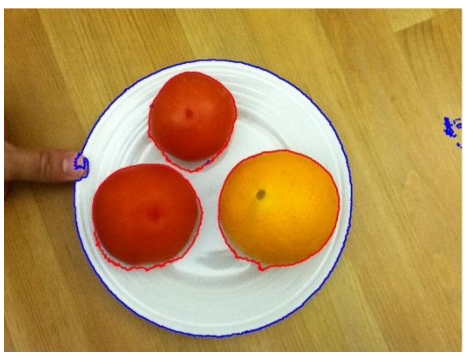
Figure 41: Images to test the application with complex food

## 7.2 Images and Different Formats Used to Extract Information

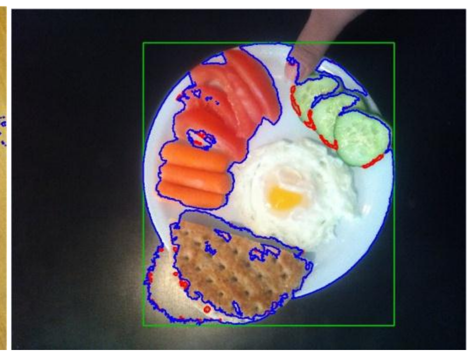
As discussed before in this thesis, we have been working with different colour spaces to perform the different analyses. One of the most important areas to highlight is the use of the colour space YCbCr, due to the capabilities attached to this colour space in the skin colour analysis of the pixels present inside the image. This colour space made possible the thumb recognition and localization in the scene: no matter what section or location the user chose to accommodate his/her finger, we were able to extract it and perform the corresponding measurements and translations. Another was the use of colour space single channel images: these images allowed us to represent some of the characteristics from the contours, and to obtain the pixel content covered by a specific area inside the picture, extracting areas in picture size. We also used the colour space for Hue Saturation and Illumination HSI inside our application. The utilization of this colour space was due to average colour analysis performed over the food portions. This, combined with the colour definition designed for the project, allowed us to categorize the different colours present inside each portion and to feed the recognition process of the Support Vector Machine SVM.

## 7.3 Contour Extraction

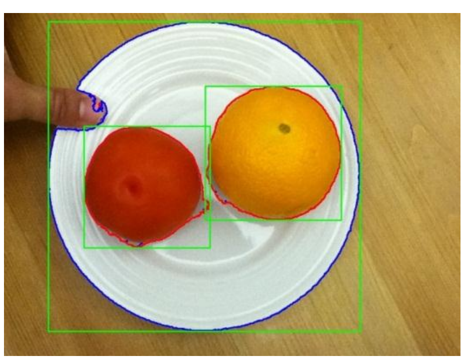
For the contour extraction, we were able to define with a good level of certainty the contours of the objects present inside the image provided by the user. Figure 42 shows a set of images with its contours defined by our application. The application has been tested with different types of pictures and foods present inside the images. The images presented in Figure 42 are a representative example of what has been used, and the final results obtained.



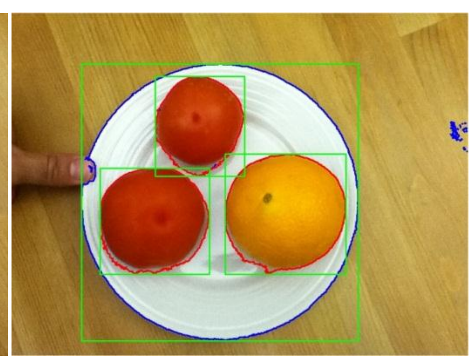
a) Contour Image 1



b) Contour Image 2



c) Contour Image 3



d) Contour Image 4

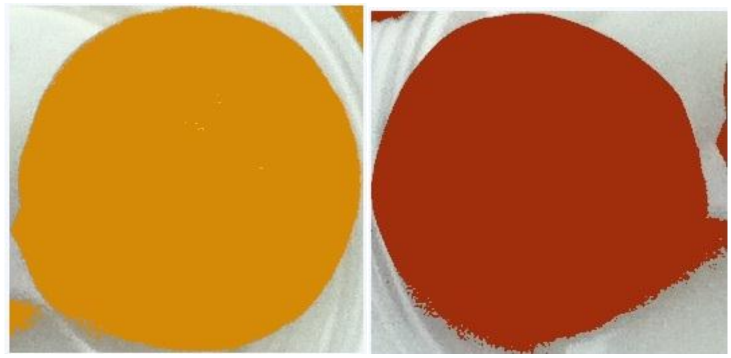
Figure 42: Contour Defined Images

## 7.4 Segmentation Results

For the segmentation area, after the contour has been defined, we have been able to extract, one by one, the portions present inside the dish. The following analyses are performed in the subsections of the image where each portion that is segmented and extracted is located.

In the process, the final images where obtained after the process of color rasterization, where the size of the image was reduced to its half, inside of a $4^{th}$ level pyramid, in order to generate a coarse small image and reproduce the color rasterization, from the upper-small level of the pyramid, to the final lower-big level, where the image returns to its original size, but at this point with smooth textures over the objects of the scene and the color rasterization performed. The color rasterization process eliminates the problems introduced by different colors and illuminations in the scene, with this task performed, our different images analyzed performed fairly similar in the following processes. The extraction of the meaningful edges is then performed, and reproduced with a gradient calculation over the image, to be able to extract the contours of the objects, and use these contours to depict and separate each portion of food. These contours are then used to extract each region of interest, per portion and continue the set of analysis one portion at a time. Given this point, we then move to the skin color extraction from the image, in order to locate and measure our measurement pattern. With the thumb isolated, and its pixels counted, each portion is then translated from pixels size into real life size; this will give us a good approximation of the size of the portions consumed by the patient.

Figure 43 shows a good example of the image segmentation and portion extraction. Here, the portions presented are also modified with the average colour obtained from each portion, and that colour has been applied to the corresponding portion to make a graphical representation of this characteristic. This is used later in our analysis to do the recognition and type of food suggestion.



c) Portion Segmentation Image 1 d) Portion Segmentation Image 2

Figure 43: Image result of portion segmentation

We have presented our steps, one by one, performing a comprehensive explanation of every step. We have also explained which path we selected to generate our solution and the different stages over the entire application, as a general review of the most important sections. We have retrieved the images from the user, generate a special colour rasterization to facilitate our analysis, obtained the measurement pattern from the image, in order to further analyze and extract the different portions of food present inside the image, and to evaluate them on areas such as colour, shape and size, to produce a better approximation of what type of food is present and the approximate value of calories present per portion. Figure 44 is a graphical walkthrough of these steps of analysis.

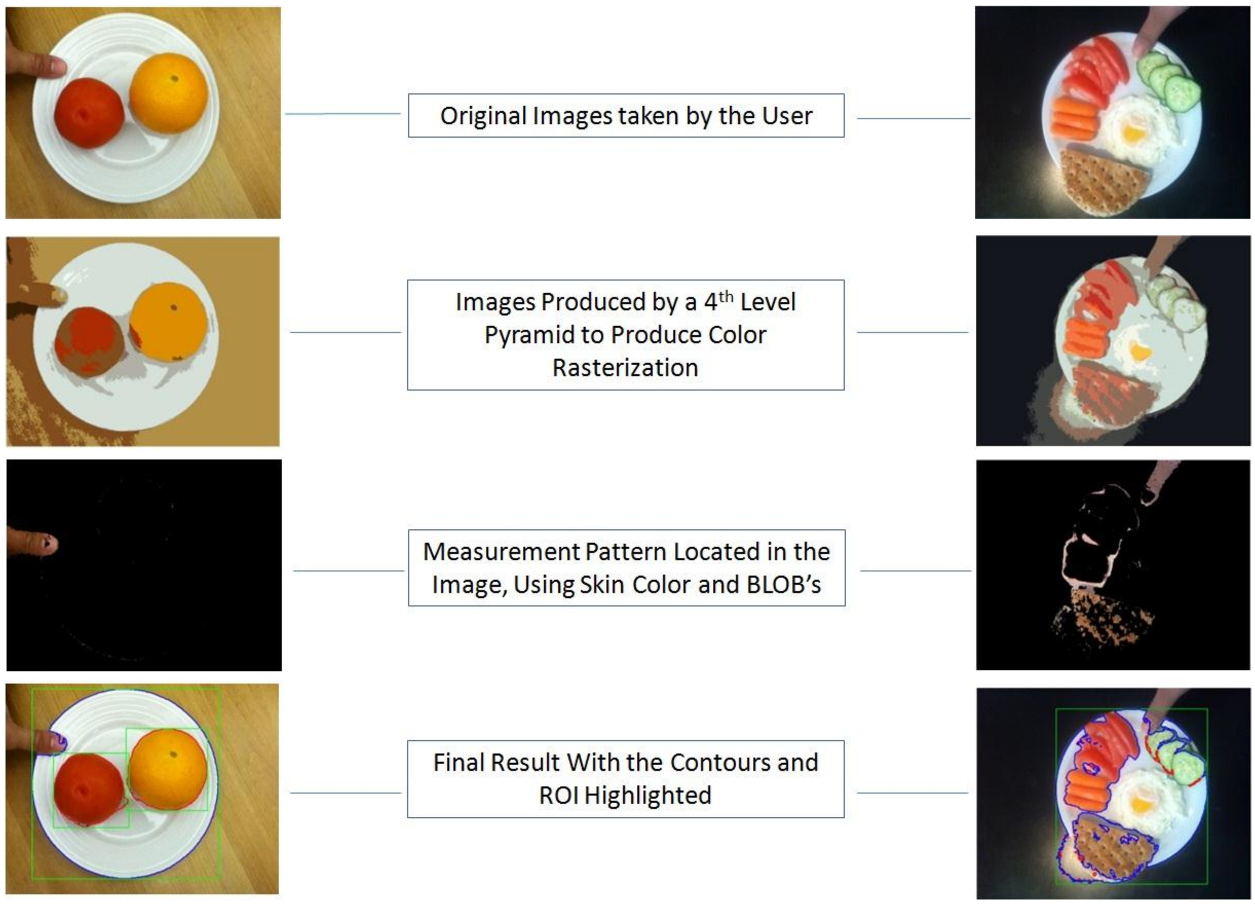



Figure 44: Image overview of the main steps of analysis

# Chapter 8 Conclusions

## 8.1 Discussion

In this work, we proposed a nutrient intake monitoring application that can be used as a personal assistive system on smartphones. The system analyzes food images using image segmentation techniques and takes advantage of a database composed of a set of training images, in order to perform semi-automatic recognition of the food present inside the image. To increase the accuracy of the recorded calories, images and the data stored inside the database are used as templates to recognize the shape, colour and texture of the portions inside of the image, to produce a semi-automatic prediction of the basic characteristics of the food present in the plate. The system will ask the user to confirm if the prediction made by the application is correct; if not correct, the user can correct any mistakes made in automatic prediction. The system then keeps a record of the intake calories for future reference.

We have created a semi-automatic application with the expert-out-of-the-loop; the food analysis is an autonomous procedure, where the important sections of extraction and calculations, normally performed by the dietician, are now performed inside our application. As shown in chapter 6, we have designed a system prototype, following conditions such as same language, in shapes, colours, distribution and design, to produce a simple but functional interface for the user to interact with our application.

We have demonstrated that our proposed measurement pattern is functional, controllable and simple enough to increase the positive qualities of the application, and to encourage the patient to use the system. This, of course, will increase the amount of food reported by the patient, and will generate more historical data for the dietician to analyze so that the control and diet efforts can be focused in the areas that will generate the best results in the treatment of obesity. In this case, the extraction of the thumb located inside the image has been performed successfully, and the use of its dimensions as a way to transform the image size into the real life size has proven to be useful.

The application has generated the data needed, within the regular estimation and results. The combination of the information extracted from the user images, the calculations performed, and the interaction with the standards defined, tied with the database analysis, has generated data close to the real life calculations and conditions. This is mainly to produce an approximation of the volume corresponding to each portion—information that will be combined with the nutrition tables to produce the regular measurements of calories present in the food that are needed to follow through with the treatment given by the dietician.

Because the system uses smartphones, it is easy to use and carry around. This will facilitate its use by patients and help their diet plans to become more effective because the patients can forget about the underreporting problem, and have fun with the application while they are compiling more data into the system for reporting to their doctor.

## 8.2 Future Work

As future work, an in-depth analysis can be performed of other contour analysis systems already in use, to analyze whether the combination of more approaches can generate better results, or if incorporating new steps will produce simple steps for the user, thereby increasing food reporting and analysis.

The extension and increase in numbers of test images and sample images for the training database is also one of the next steps to follow, as this the system will perform with better datasets and the overall processes in the food recognition will work better for the same reason. A larger number of test images will enable the application to evaluate a higher number of foods with good results and good performance.

The introduction of images to be used as templates is another path to be considered; this can give the application the opportunity to recognize portions with a high complexity, and the process can improve the actual response of the application presented. For these conditions, the image template containing only one portion of a specific food type will be used. This image will then be compared with the portions located inside the food image obtained from the user, and if the comparison gives positive results, we can skip steps in the analysis process. We can focus only after this point in the area and volume calculation, to generate the final results for the calories present in the specific portion analyzed. This will be supported by the proposed database, and the union of these features will increase the usability of the application.

A concept that has been considered, but not yet applied, is the use of context. With this, the system can learn from the user behaviour and understand the user schedule, location, time of the day, and regular types of food consumed by the patient. This type of information can be used to reduce the set of images needed to perform the food recognition, which will be helpful if we intend to reduce the response time of our system. The context also can open up more possibilities: in cases where the user is located inside a specific restaurant, we can verify the restaurant website, extract the nutritional facts and food images from the website, and finally perform simple food recognition based on the restaurant’s website information.

# References


[1] W. H. Organization, October 2011. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs311/en/index.html.

[2] W. H. Organization, November 2011. [Online]. Available: http://www.who.int/topics/obesity/en/.

[3] G. Bray and C. Bouchard, Handbook of Obesity, 2nd ed., New York: Marcel Dekker, 2004.

[4] R. Klesges, L. Eck and J. Ray, "Who underreports dietary intake in a dietary recall? Evidence from the Second National Health and Nutrition Examination Survey," *Consult.Clin Psychol,* vol. 63, p. 438–444, 1995.

[5] R. Johnson, R. Soultanakis and D. Matthews, "Literacy and body fatness are associated with underreporting of energy intake in US low-income women using the multiple-pass 24-hour recall: a doubly labeled water study," *J.Am.Diet.Assoc,* vol. 98, p. 1136–1140, 1998.

[6] J. Tooze, A. Subar, F. Thompson, R. Troiano, A. Schatzkin and V. Kipnis, "Psychosocial predictors of energy underreporting in a large doubly labeled water study," *Am.J.Clin.Nutr,* vol. 79, p. 795–804, 2004.

[7] R. Gonzalez and R. Woods, Digital Image Processing, 3rd ed., Upper Saddle River, New Jersey: Pearson Prentice Hall, 2008.

[8] G. Villalobos, R. Almaghrabi, B. Hariri and S. Shirmohammadi, "A Personal Assistive System for Nutrient Intake Monitoring," *International ACM Workshop On Ubiquitous Meta User Interfaces,* (Accepted to be published) Nov 2011.

[9] E. Adelson, C. Anderson, J. Bergen, P. Burt and J. Ogden, "Pyramid methods in image processing," *RCA Corporation,* Nov-Dec 1984.

[10] P. Burt and E. Adelson, "The Laplacian Pyramid as a Compact Image Code," *IEEE TRANSACTIONS ON COMMUNICATIONS,* Vols. COM-31, no. 4, April 1983.

[11] M. Kass, A. Witkin and D. Terzopoulos, "Snakes: Active Contour Models," *International Journal of Computer Vision,* vol. 1, no. 4, pp. 321-331, 1988.

[12] A. Myronenko and X. Song, "Global Active Contour-based Image Segmentation via Probability Alignment," *Computer Vision and Pattern Recognition IEEE,* pp. 2798 - 2804, August 2009.

[13] H. Wang and S. Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video," *IEEE Transactions on Circuits and Systems for Video Thechnologies,* vol. 7, no. 4, 1997.

[14] A. Albiol, L. Torres, C. Bouman and E. Delp, "A Simple and Efficient Face Detection Algorithm for Video Database Applications," 2000.

[15] B. Schölkopf, A. Smola, R. Williamson and P. Bartlett, "New support vector algorithms," *Neural Computation, MIT Press Cambridge,* vol. 12, no. 5, pp. 1207-1245, May 2000.

[16] B. Schölkopf, J. Platt, J. Shawe, A. Smola and R. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation, MIT Press Cambridge,* vol. 13, no. 7, pp. 1443-1471, July 2001.

[17] A. Lenhart, R. Ling, S. Campbell and K. Purcell, "Report: Teens, mobile teens and mobile phones," April 2010. [Online]. Available: http://www.pewinternet.org/Reports/2010/Teens-and-Mobile-Phones.aspx.

[18] M. Livingstone, P. Robson and J. Wallace, "Issues in dietary intake assessment of children and adolescents," *Br.J.Nutr,* vol. 92, p. S213–S222, 2004.

[19] L. Bandini, A. Must, H. Cyr, S. Anderson, J. Spadano and W. Dietz, "Longitudinal changes in the accuracy of reported energy intake in girls 10-15 y of age," *Am.J.Clin.Nutr.,* vol. 78, p. 480–484, 2003.

[20] L. Harnack, L. Steffen, D. Arnett, S. Gao and R. Luepker, "Accuracy of estimation of large food portions," *J.Am.Diet.Assoc.,* vol. 104, p. 804–806, 2004.

[21] F. Zhu, A. Mariappan, D. Kerr, C. Boushey, K. Lutes, D. Ebert and E. Delp, "Technology-assisted dietary assessment," *Proceedings of the IS&T/SPIE Conference on Computational Imaging VI,* vol. 6814, January 2008.

[22] S. Franka, N. Laharnara, S. Kullmanna, R. Veita, C. Canovaa, Y. Hegnera, A. Fritschec and H. Preissla, "Processing of food pictures: Influence of hunger, gender and calorie content," *Neural Mechanisms of Ingestive Behaviour and Obesity,* vol. 1350, pp. 159-166, September 2010.

[23] G. Bathalon, K. Tucker, N. Hays, A. Vinken, A. Greenberg, M. McCrory and S. Roberts, "Psychological measures of eating behavior and the accuracy of 3 common dietary assessment methods in healthy postmenopausal women," *Am.J.Clin.Nutr,* vol. 71, p. 739–745, 2000.

[24] A. Sawaya, K. Tucker, R. Tsay, W. Willett, E. Saltzman, G. Dallal and S. Roberts, "Evaluation of four methods for determining energy intake in young and older women: comparison with doubly labeled water measurements of total energy expenditure," *Am.J.Clin.Nutr.,* vol. 63, p. 491–499, 1996.

[25] W. Luo, H. Morrison, M. De Groh, C. Waters, M. DesMeules, E. Jones, D. S. Ugnat A, M. Lim and Y. Mao, "The burden of adult obesity in Canada," *Chronic Dis Can,* vol. 27, pp. 135-144., September 2007.

[26] M. Sun, Q. Liu, K. Schmidt, J. Yang, N. Yao, J. Fernstrom, M. Fernstrom, J. DeLany and R. Sclabassi, "Determination of Food Portion Size by Image Processing," *30th Annual International IEEE EMBS Conference,* pp. 871 - 874, August 2008.

[27] Y. Saeki and F. Takeda, "Proposal of Food Intake Measuring System in Medical Use and Its Discussion of Practical Capability," *Springer-Verlag Berlin Heidelberg,* vol. 3683, p. 1266–1273, 2005.

[28] C. Martin, S. Kaya and B. Gunturk, "Quantification of Food Intake Using Food Image Analysis," *Conf Proc IEEE Eng Med Biol Soc.,* p. 6869–6872, 2009.

[29] P. Chi, J. Chen, H. Chu and J. Lo, "Enabling Calorie-Aware Cooking in a Smart Kitchen," *3rd international conference on Persuasive Technology,* 2008.

[30] M. Westerterp, "Eating behavior in humans, characterized by cumulative food intake curves-a review," *Neuroscience and Biobehavioral Reviews,* vol. 24, p. 239–248, March 2000.

[31] K. Chang, S. Liu, H. Chu, J. Hsu, C. Chen, T. Lin, C. Chen and P. Huang, "The diet-aware dining table: Observing dietary behaviors over a tabletop surface," *Proceedings of Pervasive Computing - 4th International Conference,* p. 366–382, May 2006.

[32] T. Chan and L. Vese, "Active Contours Without Edges," *IEEE Transactions on Image Processing,* vol. 10, no. 2, February 2001.

[33] L. Cohen, "On active contour models and balloons," *CVGIP: Image understanding,* vol. 53, no. 2, p. 211–218, March 1991.

[34] L. Cohen and I. Cohen, "Finite-element methods for active contour models and balloons for 2-D and 3-D images," *Pattern Analysis and Machine Intelligence, IEEE Transactions,* vol. 15, no. 11, p. 1131–1147, 1993.

[35] T. McInerney and D. Terzopoulos, "T-snakes: Topology adaptive snakes," *Medical Image Analysis 4,* pp. 73-91, 2000.

[36] C. Xu and J. Prince, "Snakes, Shapes, and Gradient Vector Flow," *IEEE Transactions on Image Processing,* vol. 7, no. 3, March 1998.

[37] R. Liu and H. Zhang, "Mesh Segmentation via Spectral Embedding and Contour Analysis," *Computer Graphics Forum (Special Issue of Eurographics),* vol. 26, no. 3, 2007.

[38] A. Jalba, M. Wilkinson, J. Roerdink, M. Bayer and S. Juggins, "Automatic Diatom Identification using Contour Analysis by Morphological Curvature Scale Spaces," *Machine Vision and Applications,* vol. 16, no. 4, pp. 217-228, 2005.

[39] K. Siddiqi, Y. Lauzière and A. Tannenbaum, "Area and length minimizing flows for shape segmentation," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society,* vol. 7, no. 3, pp. 433-43, January 1998.

[40] L. Álvarez, L. Baumela, P. Henríquez and P. Márquez, "Morphological Snakes," *Inteligencia Artificial,* 2010.

[41] J. Angulo and J. Serra, "Color Segmentation by Ordered Mergings," *IEEE Transactions on Image Processing,* vol. 3, pp. 125-128, September 2003.

[42] R. Malladi, J. Sethian and B. Vemuri, "Shape modeling with front propagation: A level set approach," *Pattern Analysis and Machine Intelligence; IEEE Transactions,* vol. 17, no. 2, p. 158–175, 1995.

[43] L. Roberts, "Machine perception of 3-D solids," *Optical and Electro-Optical Information Processing. MIT Press,* pp. 159-197, 1965.

[44] I. Sobel, "Camera Models and Machine Perception," 1970.

[45] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Inteligence,* Vols. PAMI-8, no. 6, 1986.

[46] J. Stark and W. Fitzgerald, "An alternative algorithm for adaptive histogram equalization," *Graphical Models and Image Processing,* vol. 58, no. 2, pp. 180-185, March 1996.

[47] "Eclipse Framework," October 2011. [Online]. Available: http://www.eclipse.org/downloads/packages/eclipse-modeling-tools-includes-incubating-components/heliossr1.

[48] K. Fujiwara, F. Takeda, H. Uchida and S. Lahta, "Dishes Extraction with Neural Network for Food Intake Measuring System," *SICE,* pp. 5-7, August 2002.