**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

# PRODUCT RECOGNITION AND COUNTING IN RETAIL STORES USING COMPUTER VISION

**NOVEMBER, 2017**

**REPUBLIC OF TURKEY**
**YILDIZ TECHNICAL UNIVERSITY**
**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

# PRODUCT RECOGNITION AND COUNTING IN RETAIL STORES USING COMPUTER VISION

**MUHANAD HAMEED ARIF ALGBURI**

**MSc. THESIS**
**DEPARTMENT OF COMPUTER ENGINEERING**
**PROGRAM OF COMPUTER ENGINEERING**

**ADVISER**
**ASSOC. PROF. DR. SONGÜL VARLI ALBAYRAK**

**ISTANBUL, 2017**

**REPUBLIC OF TURKEY**

**YILDIZ TECHNICAL UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

# PRODUCT RECOGNITION AND COUNTING IN RETAIL STORES USING COMPUTER VISION

A thesis submitted by **MUHANAD HAMEED ARIF ALGBURI** in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE** is approved by the committee on 15.11.2017 in Department of Computer Engineering, Computer Engineering Program.

**Thesis Adviser**

Assoc. Prof. Dr. Songül VARLI ALBAYRAK

Yıldız Technical University

**Approved By the Examining Committee**

Assoc. Prof. Dr. Songül VARLI ALBAYRAK

Yıldız Technical University

Prof. Dr. Hasan DAĞ, Member

Kadir Has University

Assist. Prof. Dr. Erkan USLU, Member

Yıldız Technical University

## ACKNOWLEDGMENTS

I would like to thank my supervisor “Assoc. Prof. Dr. Songül VARLI ALBAYRAK ” for all the guidance and assistance she provided during all the stages of the thesis. I also would like to show my gratitude to the one who taught me my very first word, my mother. I would also appreciate all the wisdom provided by the person who put my foot on the first step on the way, my father. Many thanks also go to the one who supported me along the way, my wife. Finally, I would like to thank my daughter who has always lightened my life with her beautiful smile.

# TABLE OF CONTENTS

Page

## LIST OF SYMBOLS

| | |
|---|---|
| Mb | Blue color average in a model image |
| Ob | Blue color average in an observed image |
| Mc | Color matching factor |
| L ( , ) | Gaussian second derivative convolution of a point at a certain scale |
| Mg | Green color average in a model image |
| Og | Green color average in an observed image |
| ( , ) | Hessian matrix of a point at a certain scale |
| Mi | Matched interest points ratio |
| Mt | Overall matching factor |
| Mr | Red color average in a model image |
| Or | Red color average in an observed image |
| σ | Scale of an image |

# LIST OF ABBREVIATIONS

| | |
|---|---|
| GUI | Graphical User Interface |
| HOG | Histogram of Oriented Gradients |
| OCR | Optical Character Recognition |
| OOS | Out-of-stock |
| PCA-SIFT | Principle Component Analysis-SIFT |
| POS | Point of Sale |
| RFID | Radio Frequency Identification |
| SIFT | Scale Invariant Feature Transform |
| SQL | Structured Query Language |
| SURF | Speed Up Robust Features |
| SVM | Support Vector Machines |

## LIST OF FIGURES

Page

# LIST OF TABLES

Page

# ABSTRACT

## PRODUCT RECOGNITION AND COUNTING IN RETAIL STORES USING COMPUTER VISION


Muhanad HAMEED ARIF ALGBURI

Department of Computer Engineering

MSc. Thesis

Adviser: Assoc. Prof. Dr. Songül VARLI ALBAYRAK



Retail business is a very competitive industry and retailers are using every possible means to maintain and grow their businesses. On the other hands, computers are available in almost every retail store, and are usually used to perform traditional tasks, such as sales and financial managements. Moreover, some more sophisticated methods are proposed to assist retailers offering better services to the customers in order to maximize the profits. Some of these methods are used to support the decisions made by the retailers, while other methods are used to monitor the execution of these decisions.

The use of computer vision is found to be more cost efficient than other techniques, but it has its own limitation in classification, especially that most of the computer vision techniques convert the image to grayscale before any processing to detect shapes rather than colors. Thus, these techniques may be considered as color-blinded, while many products are different from each other mainly in color. In this study, computer vision is used to monitor the placement and number of products on the shelves of a retail store by combining the Speed Up Robust Features (SURF), which is used to detect the existence of one image into another, and the color average, to improve the classification performance, especially for products of same brands.

The results of the conducted experiments show the huge improvement in products recognition when the color average is used alongside with the matching ratio resulted from the SURF method. In experiment (A), the accuracy is measured for products recognition using the SURF method only. The detected products are 95% of the total number of products in the images, 61% of these detected products are classified correctly. Then, the color average is used alongside the matching ratio, in experiment (B), and the number of products detected in this scenario is still 95%, because the detection task is achieved by the

SURF algorithm only, and the color average has no role in this task. On the other hand, 88% of the detected product are classified correctly in this scenario, which means that the use of color average improved the results by 27%. In experiment (C), three model images per product are used for training, instead of one model image per product in the previous scenarios, to improve the recognition capabilities of the SURF method. The detected products in this scenario are 99% of the total number of products in the tested images, and 97% of these products are classified correctly into their corresponding classes.

The results of these experiments show the huge improvement in products classification when the color average is used with the ratio of the matched interest points to the total number of interest points in the model image. This improvement is according to the fact that the SURF detects interest points regardless of the colors in that image, while the main difference among products of the same brand is usually the colors of these products. Moreover, the increment of model images per product in the training dataset results a better detection, as well as better classification, especially when model images taken for each product are in different lightings and view angles.

**Keywords:** Computer vision, retail stores, product recognition, product counting.

**YILDIZ TECHNICAL UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

# ÖZET

## BİLGİSAYARLA GÖRME YÖNTEMLERINI KULLANARAK PERAKENDE MAĞAZALARINDA ÜRÜN TANINMA VE SAYMA

Muhanad HAMEED ARIF ALGBURI

Bilgisayar Mühendisliği Anabilim Dalı
Yüksek Lisans Tezi

Tez Danışmanı: Doç. Dr. Songül VARLI ALBAYRAK

Perakende ticaret çok rekabetçi bir endüstridir ve perakendeciler, işletmelerini sürdürmek ve büyütmek için her türlü yolu kullanmaktadır. Öte yandan, bilgisayarlar hemen hemen her perakende mağazasında bulunur ve genellikle satış ve finansal yönetimler gibi geleneksel görevleri yerine getirmek için kullanılırlar. Ayrıca, perakendecilere, karları en üst düzeye çıkarmak için müşterilere daha iyi hizmet sunmalarında yardımcı olacak daha karmaşık yöntemler önerilmektedir. Bu yöntemlerden bazıları perakendecilerin kararlarını desteklemek için kullanılırken, diğer yöntemler bu kararların uygulanmasını izlemek için kullanılır.

Bilgisayarla görme kullanımı diğer tekniklere göre maliyet açısından daha verimlidir, ancak bilgisayarla görme tekniklerinden çoğu görüntüyü gri ölçeğe çevirerek renklerden ziyade şekilleri algılamak için kullanıldığı için sınıflanda başarısı sınırlıdır. Bu nedenle, bu teknikler renk körü olarak kabul edilebilirken, pek çok ürün aslında birbirinden renkleri ile ayrılabilir. Bu çalışmada, "Speed Up Robust Features" (SURF) ile renk ortalamasının birlikte kullanılıasıyla bir perakende mağazanın raflarındaki özellikle aynı markaların ürünlerinin yerlerini ve sayısını bulmak için bilgisayarla görme kullanılmıştır.

Yapılan deneylerin sonuçunda renk ortalamasının SURF yönteminden kaynaklanan eşleme oranı ile birlikte kullanıldığında ürün tanıma alanında büyük başarı elde edilmiştir. Birinci deneyde, doğruluk sadece ürün tanıma için SURF yöntemi kullanılarak ölçülmüştür. Tespit edilen ürünler görüntülerdeki toplam ürün sayısının% 95'idir, burada tespit edilen ürünlerin% 61'i doğru sınıflandırılmıştır. İkinci deneyde, SURF eşleme sonuçuna renk ortalaması eklenmiştir. Tespit görevi sadece SURF algoritması ile elde edildiğinden renk ortalamasının bir etkisi yoktur, ve bu senaryoda tespit edilen ürün sayısı hala% 95'tir. Öte

yandan, tespit edilen ürünün% 88'i bu senaryoda doğru bir şekilde sınıflandırılmıştır; bu, renk ortalamasının kullanılmasının sonuçları% 27 oranında iyileştirdiği anlamına gelmektedir. Üçüncü deneyde, SURF yönteminin tanıma yeteneklerini geliştirmek için, önceki senaryolarda ürün başına bir model görüntüsü yerine, eğitim için ürün başına üç model görüntüsü kullanılır. Bu senaryodaki tespit edilen ürünler, test edilen görüntülerdeki toplam ürün sayısının% 99'udur ve bu ürünlerin% 97'si doğru bir şekilde kendi sınıflarında sınıflandırılmıştır.

Bu deneylerin sonuçları, renk ortalaması eşleşen ilgi noktalarının model resimdeki toplam ilgi noktası sayısına oranı ile birlikte kullanıldığında, ürün sınıflamasında büyük gelişme göstermemiştir. Bu gelişme, aynı markanın ürünleri arasındaki temel farklılık genellikle bu ürünlerin renkleri iken, SURF'in o görüntünün renklerinden bağımsız olarak ilgi noktaları tespit etmesine bağlıdır. Ayrıca, eğitim veri setinde ürün başına model görüntülerinin sayısının artması, özellikle de her bir ürün için alınan model görüntülerin farklı aydınlatma ve görüş açılarında olması durumunda daha iyi bir tespitin yanı sıra daha iyi sınıflandırmaya neden olur.

**Anahtar Kelimeler :** Bilgisayarla görme, perakende mağazaları, ürün tanıma, ürün sayma.

# CHAPTER 1

# INTRODUCTION

## 1.1. Literature Review

The wide availability of computers, in almost every workplace, and the low cost of operating and maintaining these computers, encourage business owners to make the best use of these computers in many tasks to optimize the workflow and minimize the errors in order to maximize their profits [1]. Thus, many computer-based methods are proposed for that purpose. These methods are used to be for financial and management purposes using simple algorithms. Moreover, the rapid progress in computers' specifications allowed the use of complex algorithms, such as machine learning and computer vision, in everyday tasks. One of the important tasks the use computer vision is monitoring products on retail store's shelves.

Retail business is very competitive, as this business is growing quickly. Thus, stores owners are using every available method that may assist maintaining and improving their businesses. The number and placement of products, on store's shelves, is one of the most important keys that lead to maximize the profits. For example, many methods are used to analyze the customer's behavior and the data acquired from points of sales in order to predict the best layout for the retail store, where the number and position of products are suggested for better sales [2].

Another example is when a production company agrees to sell their products through the retail store under a pre-agreed minimum number of these products being shown on the shelves. It is important to keep tracking the number of the displayed products to ensure not to go below the agreed number. Otherwise, the production company may terminate the agreement causing losses to the retail store. It is important to notice that losses, or less profit, are not caused by only not selling a product, but they are considered as the financial losses because of losing a customer. When customers cannot find the required product, easily, they leave that retail store, or they never come back.

The importance of knowing the products' count on shelves led to the proposal of many methods to keep track of the product on the store's shelves. Some of these methods depend on monitoring the number of products in storage and number of products sold at the point of service. Such method may not require long execution time, but the results are not accurate enough as the number of products in the storage area is monitored per parcel, not per product. Thus, the exact number of products on the shelves is still unknown. Another example is the use of methods that forecast the sales according to the old available sales data to predict the number of each product on the shelves [3-5].

All these methods are based on assumptions and forecasts. The more accurate way is based on the actual products on the shelves. The use of human power to monitor these products may be very accurate, but it is very expensive and time-consuming. Thus, it is very important to automate this process by using computers. The best alternative to human power, in this case, is to use computer vision.

Computer vision is the capability of computers to automatically recognize a pattern or a figure in an image. The rapid progress of computers' specification allows complex algorithms, such as computer vision algorithms, to be executed, in a reasonable time, using normal computers, without the need for expensive servers [6]. There are many algorithms designed for computer vision and are widely used in many fields. For example, Optical Character Recognition (OCR) is used to detect written characters and convert an image into editable text. Histogram of Oriented Gradients (HOG), on the other hand, is used to detect humans, in different positions and patterns, in images.

Many computer vision algorithms are also proposed for extracting features out of an image. To make these features repeatable, points of interest, of specific characteristics, are chosen and used for features extraction. Then, features of different images may be compared to find a match between these images or to find one image inside another [7].

Speed Up Robust Features (SURF) is an example of features detector. This algorithm detects points of interest and extracts a descriptor, which is scale and rotation invariant. Thus, this algorithm is very useful in finding matches between two images or detecting the existence of a model image inside a larger image by comparing the features extracted from each image.

Although these algorithms are designed to be as illumination-invariant as possible, they are based on edge detection for the selection interest points. Thus, an instant change in illumination in the same image, shades and lighting reflection may affect the performance of such algorithms [8]. Such effects cause the generation of inaccurate descriptors, which eventually reduces the accuracy of the results. These effects must be eliminated or minimized to make these methods applicable in a real-life, non-controlled, environments, such as retail stores.

### 1.2. Aim of the Thesis

The importance of detecting the number and placement of each product on retail store's shelves, and the high cost of the human power, alongside with the high availability and low cost of computers, it has become more efficient to use computer vision for these tasks. Although most of the computer vision algorithms are designed to be resistant to illumination variation, they are still affected by the illumination, under certain circumstances, in real-life environments.

Another drawback of the currently existing computer vision algorithms is that all the feature detectors are color-blinded. The reason behind this is that the images are converted to grayscale prior to any interest points selection. This poses another challenge for using these algorithms in detecting products on retail store's shelves, which is that most of the products that belong to the same brand are mainly different in color. These products are going to be detected as matches even if the color difference is huge between them. This confusion in detection makes it difficult, if not impossible, to categorize the product into sub-categories in order to count the available number of products of each sub-category in the shelves images.

It is also important to minimize the use of computer vision algorithms, as these algorithms require long execution time, by not repeating the execution of these algorithms whenever a filter or a task is applied. Instead, these algorithms are executed once, then the results are stored in a way the enables filtering and modifying these results without the need to execute the computer vision algorithms again.

### 1.3. Hypothesis

In this thesis, the most appropriate computer vision method is chosen for products detection in shelves images. The selection is based on comparisons made in other studies to illustrate the benefits of using each method for a specific task. The drawbacks of the selected method are also handled in this thesis by using the color as matching factor alongside with the matching factor resulted from the computer vision algorithm. The matching results are stored in a database for any further processing or filtering. Thus, the execution of the computer vision algorithm is not repeated for every query. This also provides the ease of accessing results in order to develop more features that may assist the owners of the retail stores with minimum time consumption.

The remainder of the study is constructed as follow: chapter two reviews literatures related to the subject of the study. Chapter three demonstrates the proposed solution and the details of the algorithm used in features extraction and detection. Chapter four show the experiments conducted to test the performance of the algorithm in different scenarios. Then, these results are discussed in details and compared to results from other studies in chapter five. Finally, the conclusions of this study are summarized in chapter six.

# CHAPTER 2

# RELATED WORK

The rapid growth of the retail business field leads to enormous challenges faced by the store owners in order to maintain and expand their businesses. Thus, many computer-based solutions are proposed to overcome these challenges. One of the key challenges faced by retail stores owners is to avoid running *Out-of-stock* (OOS) as it may affect the purchase decision of the customer causing sales reduction [9]. The OOS event occurs when the product is not on the shelf even if the product is available in the storage area of the store, keeping in mind that the more product types exist on the shelf, the less space for each type [10], thus, monitoring the products on the shelves is the real key to avoid such problems.

Guren and Corsten [11] propose a methodology to predict consumer's behavior in order to redistribute the spaces reserved for each product on the shelf in a way that the more the product is consumed, the more space reserved to avoid OOS events. This methodology is presented as a substitute for using *Radio Frequency Identification* (RFID) applications according to technological and financial reasons. Papakiriakopoulos and Doukidis [12] propose a machine learning based method to predict OOS events, the study depicts that the opportunity for building such mechanism to detect missing products from the shelf is better for retail chains that face high OOS rates and keep extended data sources. Papakiriakopoulos [13] proposes another machine learning based method that collects data from all *Points of Sale* (POS) in a retail chain in order to provide an OOS list to the store manager and the product supplier. This method uses historical data and rule-based system to conclude the OOS list.

Another challenge for retailers is the correct placement of products. For example, a customer walks into a retail store to buy a specific item may be encouraged to buy another related item when placed correctly. Many studies propose methods to decide the correct placement of products depending on the analysis of customers' behavior and psychology while other studies depend on data analysis to find products that are sold together to suggest placing them together to maximize profit [14]. Thus, a misplaced product will affect the sales of that specific product and eventually reduces the overall profit. Surjandari and Seruni [15] propose a method to find the association rules between items sets using Apriori algorithm for data mining to analyze the market basket. The results of this analysis are used to suggest a store layout expected to achieve maximum profit by placing associated products close to each other.

A product may be misplaced according to an employee mistake or when products are moved by customers during their shopping. Thus, to ensure that there are no misplaced products, the store owner needs to maintain the layout proposed by such methods for optimal sales. In summary, a store owner needs to monitor the number of products placed on the shelves and where they are placed to overcome many challenges such as OOS, misplaced products. The methods proposed depending on data analysis are mainly targeted toward retail chains with large sales data.

Another approach is proposed by using robots distributed around the retail chain stores and remotely controlled by an employee who is responsible for monitoring the products on stores shelves. This method is also recommended for retail chains, as it is easier and less expensive for a store owner to walk around his store to monitor the products on the shelves [16].

A more practical methodology to monitor products on shelves, as it is more cost efficient and is suitable for any store whether it is a part of a retail chain or not, is by using *Computer Vision*. Computer vision aims to replicate human visual system using computational models to enable computers to perform some of the human visual tasks. Although computer vision suffers by comparison to the human visual system because, simply, the human visual system is too good to be replicated, many algorithms are proposed to perform a specific task with limited capabilities [17].

Segmenting an image into meaningful segments that do not overlap with each other is called *Image Segmentation*. In order to achieve a proper segmentation, a region of pixels that are similar to each other and significantly different from pixels in adjacent regions is considered a segment [18]. Many methods are proposed to detect the boundaries of a segment, Martin, Fowlkes, and Malik [19] propose a method that depends on detecting a change in pattern to define the boundaries of a segment, while another approach based on reward and punishment to reward similar adjacent pixels and punish non-similar pixels is used to set the boundaries of a segment in [18]. Examples of image segmentation at two different levels are shown in Figure 2.1. Image segmentation is an excellent example to illustrate the principle of computer vision, but it is very sensitive to noise, view angle and illumination of the image, thus, it is rarely used to detect objects in images in real life cases.

Figure 2.1 Examples of image segmentation at two different levels [18]

Text recognition is another example of computer vision, various fonts and styles are provided for each letter to train the computer in order to be able to recognize text in images. This is called *Optical Character Recognition* (OCR). Although using OCR for product recognition is not applicable according to the different designs of logos that make it almost impossible for OCR to detect the characters in the logo, OCR is sometimes used to provide additional support to the product recognition by attempting to recognize text on the product other than the logo [20].

More invariant methods are proposed using Features Detection. Extracting information that may be used to represent image contents, for the purpose of a certain task execution, is fundamental to computer vision. This information is selected based on the properties it exhibits or the patterns it represents. The presented patterns may be blobs, corners, contours, and much more. Image features are the combination of all pattern extracted from an image. Interest-point-based features are widely used in image matching and are preferred over many other feature detection approaches. The quality of a feature detector is measured by their localization, robustness, sensitivity, stability, and complexity. Localization defines the detector's accuracy of the detected location in comparison to the actual location. Robustness is a measure of a detector's sensitivity to the noise in the image being processed. Sensitivity is described as the ability of a detector to detect feature in images with low illumination. Stability defines the ability of the detector to extract the same features from an image from the same locations regardless to the photometric or geometric transformation of that image. Complexity represents the time required by the detector to extract features from an image. To detect a feature point, detectors use various derivative operators to check the variations in image intensity [21]. The detector proposed by Moravec [22] is one of the earliest detectors developed for interest points extraction by measuring the changes in intensity around a pixel in different directions inside a local window. Vertical, horizontal, and two diagonal squared difference sums are computed for each pixel in the image with the surrounding pixels in the local window. Then, the minimum sum is used as variance measure to select points of interest from the image. Another feature detection based method namely *Histogram of Oriented Gradients* (HOG) is proposed by Dalal and Triggs [23] mainly for detecting humans in images. This method uses a dense grid to evaluate normalized local histograms of image gradient orientation. The challenges faced by human detection process in images, according to different patterns that may exist in clothes and different poses of humans, enables the HOG algorithm to be used in various objects detection.

A customer assisting solution is proposed by George, Mircic, et al. [20] that aims to improve the shopping experience by recognizing the fine-grained product classes of products on the customer's shopping list, in images taken by a smart phone camera for shelves in a retail store. This solution uses images of various patterns in each class and uses HOG descriptor to extract the features in those images and provides them to *Support Vector Machine* (SVM) detector for training. This solution also uses OCR to detect text on products. A histogram of words is created for each class, so when a customer searches for a specific item by brand name or type of the product, the solution will attempt to find a match in the words histogram to predict the desired product.

*Scale Invariant Feature Transform* (SIFT) is proposed by Lowe [24] to obtain scale-invariant coordinates that are related to local features from image data. The key aspect of this approach is the dense covering of an image by a huge number of features over the entire range of scales and locations. The matching process in this method is done by individually comparing each feature extracted from the model image to those extracted from the compared image. A set of matched features is generated for further filtering in order to find the correct matches depending on the object and its location, scale, and orientation in the compared image.

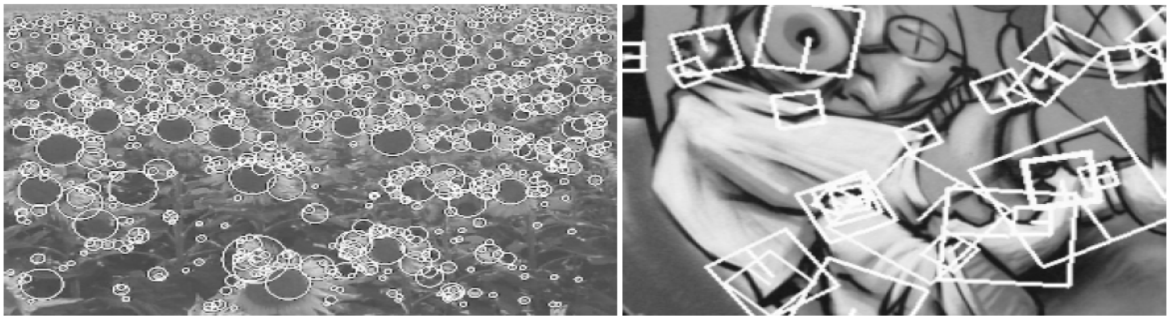
Two characteristics are preferred in the matching methodologies when being used for product recognition; those are, scale- and orientation-invariance. As the images being processed are taken from a real world environment rather than a controlled environment, the images and the products in these images are most likely to be different in scale and orientation than the model image being matched. To achieve orientation-invariant model, the descriptor of the feature points extracted should contain information about the local orientation of this point. When using local orientation in the descriptor, it enables the matching descriptors regardless of the entire image orientation. For scale-invariance, an image pyramid is created where the base of the pyramid is the original image and each layer upward is a reduced size copy of the layer beneath it. A scaling factor is used to set the size of the new layer in the pyramid with respect to the previous layer.

A product recognition method is proposed by Varol and Kuzu [25] based on using HOG and SIFT algorithms. The proposed method decomposes the process into detection and recognition, the detection is done using the HOG algorithm while the recognition is done using SIFT algorithm and color comparison. As this method is targeted to detecting tobacco packs, which are usually arranged in a special container, it is easier to set the boundaries of shelves, then, as tobacco packs have similar shapes, the method uses HOG algorithm to set the boundaries for each pack as shown in Figure 2.2. The segments set by the HOG are further processed in order to conclude the type of the product. The decomposition of the process is used to accelerate the recognition process as the SIFT algorithm densely samples the image, more processing time is required to find matches in the entire image while segmenting the image using the HOG algorithm will produce only relevant sections of the big image to be processed by the SIFT.

Figure 2.2 Shelf boundaries and products segmentation [25]

The warnings on tobacco packs are usually common among all tobacco brands. Thus, only the top 40% of the segment is used for classification as it includes the unique logo for that specific brand. On the other side, the training images for the SIFT algorithm contains only the brand logo, so that the SIFT is not confused by the warning images and texts on the tobacco packs. The combination of shape and color information in recognition produces more accurate results, as brands with similar shape have different colors and vice versa.

A rotation- and scale-invariant detector and descriptor for interest points are proposed by Bay, Tuytelaars, and Gool [26] named Speed Up Robust Features (SURF), which is discussed in details in the next chapter. In this method, Hessian matrix is used for detection according to the high accuracy and low computation time compared to the Hessian-Laplace detector. The determinant of the Hessian matrix is used to select the scale and location of the point of interest rather than using the more complex Hessian-Laplace detector. Moreover, instead of using image pyramids to implement a scale space by iteratively reducing the image size, this method uses a different, and yet faster, method by up-scaling the filter size so that any scale can be derived directly from the original image by using the suitable filter size. The descriptor for points of interest in this method is extracted for the square region centered at the point of interest (Figure 2.3.a) and oriented according to the reproducible orientation found for the circular region around that point (Figure 2.3.b).



(a) (b).

Figure 2.3 (a) Circular regions around the interest points (b) Square regions centered at the interest point [26]

A product recognition and counting method, proposed by Kehriwal, et al. [27], uses SURF to detect the number of each product in shelves images. Six cameras, three on each side, are mounted on a robot that moves through the aisles of the store to take shelves image in order to count visible products. A set of model images for each product is provided to the method to be used for detection. SURF is used to detect these products, then the detected product is removed from the image, and the counter is incremented. A grid search is conducted in the neighborhood of the detected product in an attempt to find any products available. A brand may have many products that share the same logo but different in some other characteristics like color or size which might be very confusing for features detection methods because these methods convert the image to grayscale prior to any feature extraction. Thus, color has no effect on the results of detection or matching. Feature detectors are supposed to be scale-invariant because it is almost impossible to get two identical pictures in a real world non-controlled environment where pictures may be taken from closer, farther, or different point of view. Thus, products with the same logo but mainly different in size may also be confusing to the matching methods as they must be scale-invariant. For these two reasons, this method is not able to detect different categories of the same brand in the shelves images taken by the robot. The experimental results of this method show the efficiency and reliability of using SURF for products detection in a shorter time compared to other detection methods.

Another mechanism is proposed by Skoczylas [28] for products recognition and counting using smart phones. This mechanism uses a sliding window with a size related to the size of the image under investigation. This window is moved from the top left corner to the right down corner with a step equal to half the size of the window. Any product detected inside this window is removed from the descriptor of the image. After finishing the first pass, the window size is doubled then the entire process is repeated until the window size reaches the image size. This approach is proposed to overcome the problem of multiple identical products in one image. This will cause similar feature points to be distributed on the image causing skewed detection results. But, the problem still exists when two or more products are detected in the same window at a certain step. This study also shows the superiority of the SURF detector when used for products recognition.

The comparison made by Juan and Gwun [29] also show the superiority of the SURF method over the SIFT and the Principle Component Analysis (PCA)-SIFT, which computes relatively, compared to the standard SIFT method, smaller feature vector that can be used in the same comparison algorithms by projecting local patches of gradient images using the eigenspace computed for that patch. The comparison illustrates the low processing time required by the SURF and the lower sensitivity to illumination changes in compared images. The low illumination sensitivity is very important for products detection because of the different lightings used in retail stores and also the different effects of the lighting on the products. Figure 2.4 illustrates the performance of the SURF detector in different illuminations compared to the SIFT and PCA-SIFT methods, where 10 correct matches were found by the SURF method out of the 10 matches while 9 correct matches out of the 10 matches found by the SIFT and only 6 out of 10 matches were correct when PCA-SIFT method is used.

Figure 2.4 Images matching in different illuminations [29] Left: using SIFT. Middle: using PCA-SIFT. Right: using SURF

## CHAPTER 3

## FEATURES EXTRACTION AND MATCHING

The proposed method for product recognition and counting is achieved using three main steps that are matching, validation and results management. The matching and validation are done basically using SURF and color average comparison. In addition to few validation rules that will be discussed in details later in this chapter.

### 3.1. SURF (Speed-Up Robust Features)

The matching task is based on detecting interest points in an image and creating descriptors for these points. These points are matched between two images in order to find a match. Thus, for a successful match, the interest points need to robust and immune to noise, and the descriptors are required to be scale and orientation-invariant [30].

#### 3.1.1. SURF Detector

According to the low computation time and high accuracy, the detection in SURF is based on Hessian matrix. The location and scale selection measure in SURF is achieved based on the determinant of the Hessian matrix rather than being based on Hessian-Laplace detector as most remaining detectors. The Hessian matrix $\mathcal{H}(\mathrm{x}, \sigma)$ at the point x = ($x$, $y$) of the image $I$ at scale σ, is defined as

$$\mathcal{H}(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \tag{3.1}$$

where $L_{xx}(\mathrm{x}, \sigma)$ is the convolution of the Gaussian second derivative $\frac{\partial^2}{\partial_x{}^2} g(\sigma)$ with the image $I$ in point x, and so on for $L_{xy}(\mathrm{x}, \sigma)$ and $L_{yy}(\mathrm{x}, \sigma)$ [31].

Instead of using discretized and cropped Guassians (Figure 3.1 left half) to achieve optimal scale-space analysis, SURF uses box filters (Figure 3.1 right half) as an approximation of the second order Guassian derivatives which can be evaluated quickly using integral images regardless of the size [26].

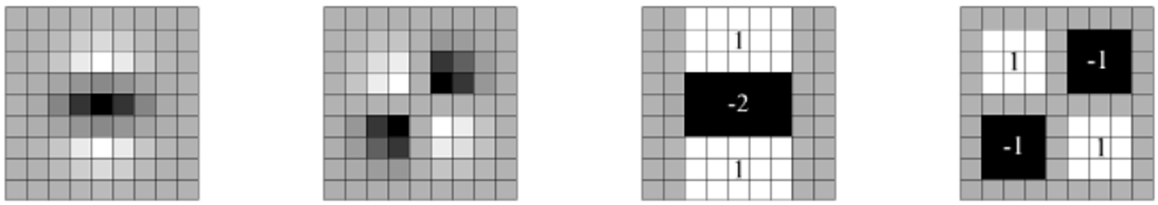


Figure 3.1 Left half: The second order partial derivative of the discretized and cropped Guassian in y-direction and xy-direction. Right half: SURF approximation using box filters[26]

The Gaussian second order derivative is approximated by 9×9 box filters as shown in Figure 3.1 with $\sigma = 1.2$ representing lowest scale. This approximation is denoted by $D_{xx}$, $D_{yy}$ and $D_{xy}$ keeping the computations efficient by maintaining the simplicity of the weights applied to the rectangular regions. Moreover, further balance is needed to the relative weights in the Hessian's determinant expression with $\frac{|L_{xy}(1.2)|_F |D_{xx}(9)|_F}{|L_{xx}(1.2)|_F |D_{xy}(9)|_F} = 0.912\ldots \cong 0.9$, where $|x|_F$ is the Frobenius norm. From the above equation, we conclude

$$det(\mathcal{H}) = D_{xx}D_{yy} - (0.9D_{xy})^2 \tag{3.2}$$

To ensure that the Frobenius norm is constant at any filter size, the filter responses are normalized according to the mask size.

Methods other than SURF, like SIFT, usually implement scale spaces by creating image pyramids using images smoothed with a Guassian and sub-sampled subsequently to achieve the next layer of the pyramid [32]. According to the combined use of box filters and integral images, SURF does not have to apply the same filter repeatedly to the results of the layer filtered before, and it is capable of applying these filters of any size at precisely the same computation time directly from the original image. In other words, the scale space analysis is based on filter size up-scaling rather than reducing the size of the image per iteration. The results from the above 9×9 filter is deemed to be the initial scale layer which is referred as scale $s$=1.2 (corresponding to Guassian derivatives with $\sigma = 1.2$). The remaining layers are obtained by increasing the mask size gradually to filter the image, keeping in mind the specific structure of these filters and the integral images' discrete nature. This will results in filters of size 9×9, 15×15, 21×21, 27×27, etc. For larger scales, the increase in filter size for two consecutive filter sizes should also scale respectively. Moreover, for every new octave, the increment in filter size is doubled. For the interest points localization in the image and over the scales, a non-maximum suppression is applied in a 3×3×3 neighborhood.

### 3.1.2. SURF Descriptor

When an interest point is detected by the SURF, a descriptor is generated for this point. Descriptor generation process can be divided into two steps. The first step aims to find a reproducible orientation for that interest point and the second step generates SURF descriptor using the orientation information computed in the first step.

For the descriptor to be rotation-invariant, it is mandatory to find a reproducible orientation for the interest point by calculating the Haar-wavelet responses in both *x and y* directions for the circular region around the interest point with a radius of 6*s*, where s is the scale that the interest point is detected at. The sampling step is chosen to be *s* which means that it is also scale-dependent as well as the computation of the wavelet response. Thus, Integral images are used for fast filtering as the size of the wavelet gets bigger at higher scales.

After calculating the wavelet responses and weighting them with a Guassian, centered at the detected interest point, these responses are represented as vectors in a space where the strength of the horizontal response is set along the abscissa and the strength of the vertical response is set along the ordinate. Using a sliding orientation window with an angle of $\left(\frac{\pi}{3}\right)$, all responses within this window are summed in order to estimate the dominant orientation by generating a new vector from the horizontal and vertical responses calculated for this window. The orientation of the longest vector resulted is used as orientation of the interest point for the following step.

Next, the descriptor is extracted by constructing a square region centered at the interest point and oriented according to the orientation estimated earlier. This region is then split up into 4×4 smaller square sub-regions to keep the crucial. Few simple features are computed at 5×5 uniformly spaced sample points. Finally, a four-dimensional vector is created using the summation of the Haar wavelet response in horizontal ($d_x$) and vertical ($d_y$) in addition to the summation of the absolute values of these responses, $|dx|$ and $|d_y|$. The vertical and horizontal here are defined with respect to the orientation of the interest point and the summation is for the 4×4 sub-regions creating a vector with the length of 64 for each interest point.

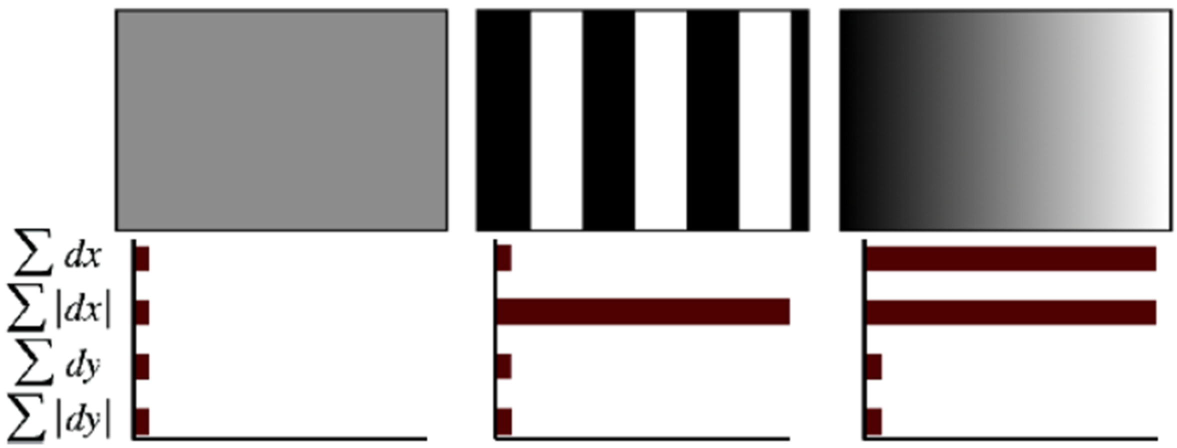



Figure 3.2 Entries of a SURF descriptor for different region natures. Left: homogenous region. Middle: frequencies in the x direction. Right: intensity gradual increment in the x direction [26]

### 3.2. Algorithm

There are two categories of images used by the algorithm, the model images which are the images of the products to be detected and counted and the observed images which are the shelves images where the products are searched in. All images information are stored in the database as well as all the intermediate and final detection results.

For each observed image, the algorithm loads the images and uses SURF to detect interest points and compute descriptor for them. Then, the model images are loaded one by one and SURF is also used to extract descriptors for the interest points in each image. The observed image is then searched for any match with the model image. If so, the matched region is checked for three constraints to make sure that the detected region is acceptable. These constraints are that the ratio of the number of matched points to the total number of interest points detected in the model image is larger than a pre-set threshold, the region is shaped like a rectangle and has a width and height more than a pre-defined number of pixels. Because there is more than one product in every observed image, it is possible to detect some interest points from one product and other points from another product resulting either a non-rectangular or very thin region. Finally, the boundaries of the matched region are checked whether they fall inside the shelves image or not. If the detection is found to match the constraints set, the resulted match stored in a special table set for the intermediate results in order to be validated later and the detected region is removed from the observed image, and a new descriptor is created in order to remove the detected interest points from the descriptor.

The data stored about the found match contains the match ID, model image ID, the observed image ID and the match indicator in one table alongside with the match ID and the four corners of the detected region in another table. The match indicator is used to calculate how close the detected model image in the observed image is. This indicator is calculated using two main factors, the first factor is the ratio of the number of matched interest points to the total number of the interest points found in the model image and the second factor represents the color similarity of the model image and the detected region. To calculate this factor, color averages are computed for both the model image and the detected region. Then, the average color match of the three colors is computed between the model image and the detected region. The equation used for color matching factor is:

$$M_c = \left( \frac{\left(1 - \frac{|M_r - O_r|}{255}\right) + \left(1 - \frac{|M_g - O_g|}{255}\right) + \left(1 - \frac{|M_b - O_b|}{255}\right)}{3} \right) \quad (3.3)$$

where $M_c$ is the color matching factor, $M_r$, $M_g$, $M_b$ are the model image average's red, green and blue components, and the same for the observed image average's values $O_r$, $O_g$, $O_b$. The resulting value from this equation varies from 0 to 1, where 0 means least match and 1 represents the highest possible match.

Some products have very similar color histograms for products in different categories of the same brand as shown in Figure 3.3. Thus, the deviation of the average color among these products is very close to each other resulting very close color matching factors. To overcome this problem, the color matching factor is taken to the power five when the overall matching factor is calculated. The equation of overall matching factor $M_t$ is $M_t = M_i * M_c^5$, where $M_i$ is the ratio of the number of matched interest points to the total number of interest points in the model image.

Figure 3.3 Products of the same brand in different categories

After adding the match information to the database and removing the interest points in the detected region from the observed image, the observed image is searched again for the same product until there is no match found by the SURF or the match found does not comply with the constraints set for the matching task. A fresh copy of the observed image is loaded, and a new descriptor is generated in order to search for the next product in the database. This procedure is repeated until all selected products are searched for in the selected observed images. Then, results validation process is started. Following is the pseudo code of products detection algorithm:

Algorithm: Product Detection and Counting

Input:

O //Shelves images data

M //Products data

Output:

Products detected in shelves images

Begin

| | | |
|---|---|---|
| Recognition | Step1: | Load new shelves image (Oi) |
| | | Generate SURF descriptor (Od) for this image |
| | Step2: | Load new model image (Mi) |
| | | Generate SURF descriptor (Md) for this image |
| | Step3: | Find a match (D) of (Md) in (Od) |
| | | If (D) not found or (D) not comply with constraints the goto step2 |
| | Step4: | Extract corner points (Rp) for the region (D) from (Oi) |
| | | Save (Rp) to data table (Pt) |
| | Step5: | Calculate match factor (Mf) for the detected region and model image |
| | | Save detected match details to data table (Mt) |
| | Step6: | If unprocessed model image exists then goto step2 |
| | Step7: | If unprocessed shelves image exists then goto step1 |
| Validation | Step8: | Arrange (Mt) by (Mf) descending |
| | Step9: | Load new shelves image (Oi) |
| | | Create new blank image (Bi) with identical details of (Oi) |
| | Step10: | For each match in (Mt) related to shelves image (Oi) |
| | | Extract region from (Bi) using corner points position from (Pt) |
| | | If extracted region > 80% blank then Copy match information to results table (Rt) |
| | | End for |
| | Step11: | If unprocessed shelves image exists then goto step9 |

End

A product in the observed image may be detected as a match when compared to more than one model image according to the similarity among these products. On the other hand, a closer match has higher match indicator. Thus, the results from the matching process are arranged according to the match factor in descending order in order to validate the results. A new blank image is created identical to the observed image being validated then the matches found in this image are selected one by one as arranged earlier. The detected region is compared to the blank image created if this region has more than 80% of its area blank, then this match is considered valid, and a new polygon is drawn in this image for further validation. Otherwise, the match is considered invalid and is neglected. If the blank area of the matched region is less than 80%, then the product in this region is already detected as a match to another model image with higher match factor, i.e. better similarity. The validated results are stored in a table for any further processing or display. Each record contains the match ID, observed image ID and model image ID. As the match ID is the same among all database tables, the regions of the validated matches can be retrieved from the same table where they are stored earlier.

For a better illustration, the steps to recognize two products in an image are shown in Appendix A. A real-life shelves image is used, where all the products in that image are removed except two products of the same brand but in different categories. This provides a better understanding of how the color average is used for better classification. To keeps the image as close to the original properties as possible, the removed products are replaced with black polygons instead of deleting them, which results a new image with different dimensions than the original image.

### 3.3. Application

The application to test the performance of the algorithm is built using C# programming language and visual studio 2010 environment. EMGU [33] library, which is a cross platform .Net wrapper to the OpenCV [34] image processing library, is used for the matching process using SURF. The main interface is shown in Figure 3.4 and can be split into three main sections, Products, Images and Process.

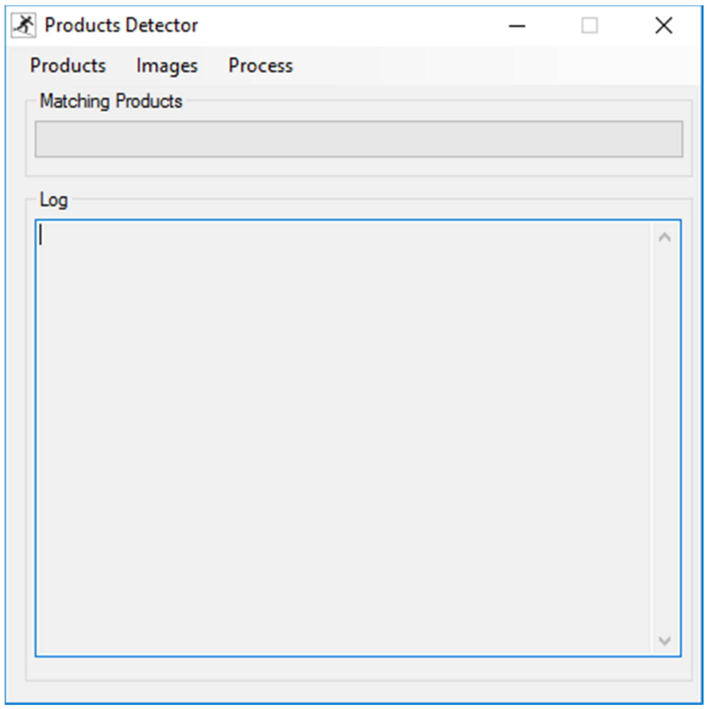



Figure 3.4 Graphical User Interface (GUI) of the application

Products section is used to add new products information as well as the model image for that product. It is possible to add more than one product with the exact same details, but it is impossible to add two identical model images. This is achieved by calculating an MD5 hash for every image file and check if there is an image with the same hash in the database before storing this one. If yes, the product is not added and the user if prompted that an identical image already exists. This is used to avoid losing processing time while searching for matches for the same model image more than once. It is also possible to modify any of the product details other than the model image of the product or deleting the entire product. The information that the user needs to provide, for every new product, are product type, product name, sub-category and model image. All this information is stored in the database alongside with the automatically generated unique ID for each product and the calculated hash for the file. The new product and manage products interfaces are shown in Figure 3.5.

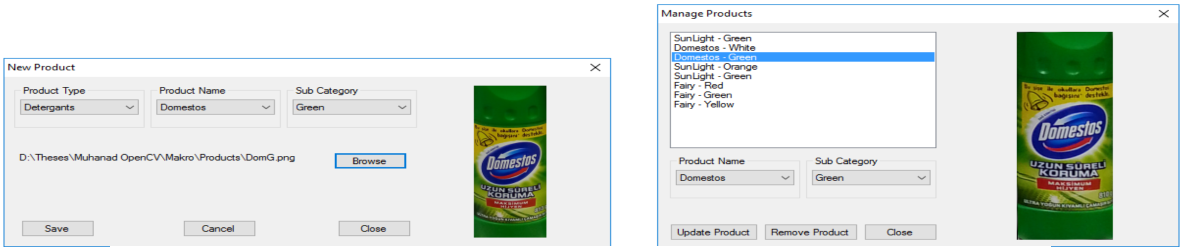



Figure 3.5 Left: New product interface. Right: Manage products interface

The images section is used for adding new shelves image or editing an existing one. Unlike products, shelves images do not require much information. Only image name is required to make it easier for the user to know the image by its name. Filenames of these images are used as image names when more than one file are selected to be added simultaneously. When a new shelves image is added, it is automatically selected for processing. It is possible to deselect one or more images from being processed using the image editing interface to save time in case that the image is not required to be searched in the meantime.

The process section has two options, the first option is to process the data provided to the application and the second option is to display the results of the detection and counting. When selecting to start processing the existing data, the user is required to choose whether to process all products in the database or to search for specific products types. The selected products types are then searched within the observed images that are selected earlier to be processed. The results are stored in the database so that it may be displayed at any time later without the need to run the detection process all over again. This reduces processing time as the results may be processed and displayed in any way required directly from the database and provides accessible data for any other applications that may need these results in the future. For example, in this application, the products in each observed image are grouped according to the product name and the sub-category to provide a list of the detected products as well as the number of these products detected in this observed image. This is used to provide more flexibility as it is possible to provide more than one model image for the same product and each has a different product ID. The interface designed to display the results is shown in Figure 3.6.

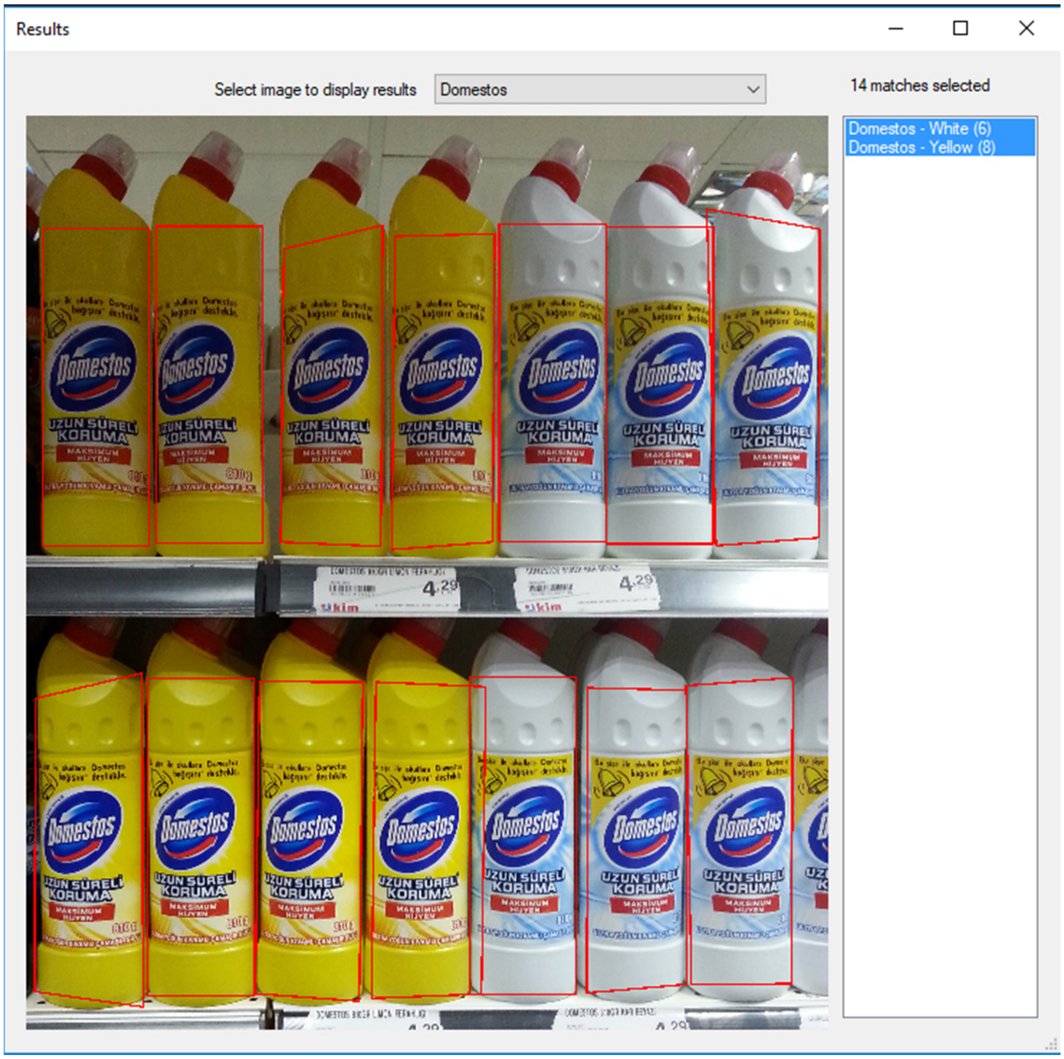



Figure 3.6 Results management interface

# CHAPTER 4

# EXPERIMENTAL RESULTS

Three scenarios are used to conduct three experiments in order to evaluate the performance of the proposed method. Experiment (A) uses only SURF algorithm for detection and classification. This scenario is used to provide control results for the other experiments to be compared to. Experiment (B) uses color average alongside with the ratio of the number of matched interest points to the total number of interest points extracted from the model image for products classification in the sub-categories. These two experiments are conducted using one model image per product to illustrate the effect of using the color average on the detection and classification results. Experiment (C) is conducted using three model images per product to evaluate the performance of the proposed method in real-life and compare it to other methods proposed in other studies.

All images used in these experiments are real-life, un-edited, images taken for shelves in retail stores in Turkey. These images are taken using an eight-megapixel smartphone camera in different illuminations and view angles depending on the environment of the retail store taken from. The model images for products are cropped from shelves images that are not included in the detection process during the experiments. When more than one model image per each product in the training set are used, different illuminations and view angles are chosen for each product to provide more flexibility for the detection and classification.

Two product types are chosen for the experiments, detergents and shampoos. Each product is added to the database with its brand name and sub-category. In total, 50 shelves images taken from four retail stores are used in the experiments containing 675 products from the

32 product types shown in Table 4.1. A sample image for each product category is shown in Appendix E.

Table 4.1 Summary of the products in the database

| **Product Type (#of total subcategories)** | **Product Name** | **No. of sub-categories** |
|---|---|---|
| Detergents (15) | Fairy | 3 |
| | Domestos | 4 |
| | Pril | 5 |
| | Bingo | 3 |
| Shampoos (17) | Head & Shoulder | 5 |
| | Loreal | 3 |
| | Johnson's Baby | 3 |
| | Elidor | 6 |
| Total | | 32 |

The program used to test the method is written using C# programming language and EMGU [33] library, which is a cross platform .Net wrapper to the OpenCV [34] image processing library, for the SURF detection. The program is built using Microsoft Visual Studio development environment and is tested using a computer with an Intel® Core™ i7-4500U @ 2.40GHz processor and 8.00 GB memory with Microsoft Windows 10 64-bit operating system.

### 4.1. Experiment A

Products are detected in shelves images and classified into subcategories using **only SURF method** in this experiment. One model image per product is added to the database for detection and classification resulting in a total of 32 entries. The total number of products detected in the shelves images is 641 (95%) out of the 675 products available in the shelves images while 34 products (5%) are not detected as valid matches. Products are sub-categorized in this experiment depending on **the ratio of the number of matched interest points to the total number of interest points detected in the model image of the product**. When a specific region in the shelves images is matched to more than one product model image, this product is classified according to the match with the highest matching ratio and the remaining matches are considered invalid. This classification scheme results in 391 correctly classified products which represent (61%) of the detected products and (58%) of the total number of products in shelves images. The confusion matrix for this experiment's results is shown Appendix B. The results of one of the processed shelves image, shown in Figure 4.1, illustrate the confusion occurred when products are detected in this experiment. Some products are detected in subcategories that do not exist in the processed image. Both model image and shelves image are converted to grayscale prior to the interest points detection and matching process in SURF causing products classification to be color invariant. The average time consumed by the application to process all the shelves images in this experiment is 7.11 seconds per product.

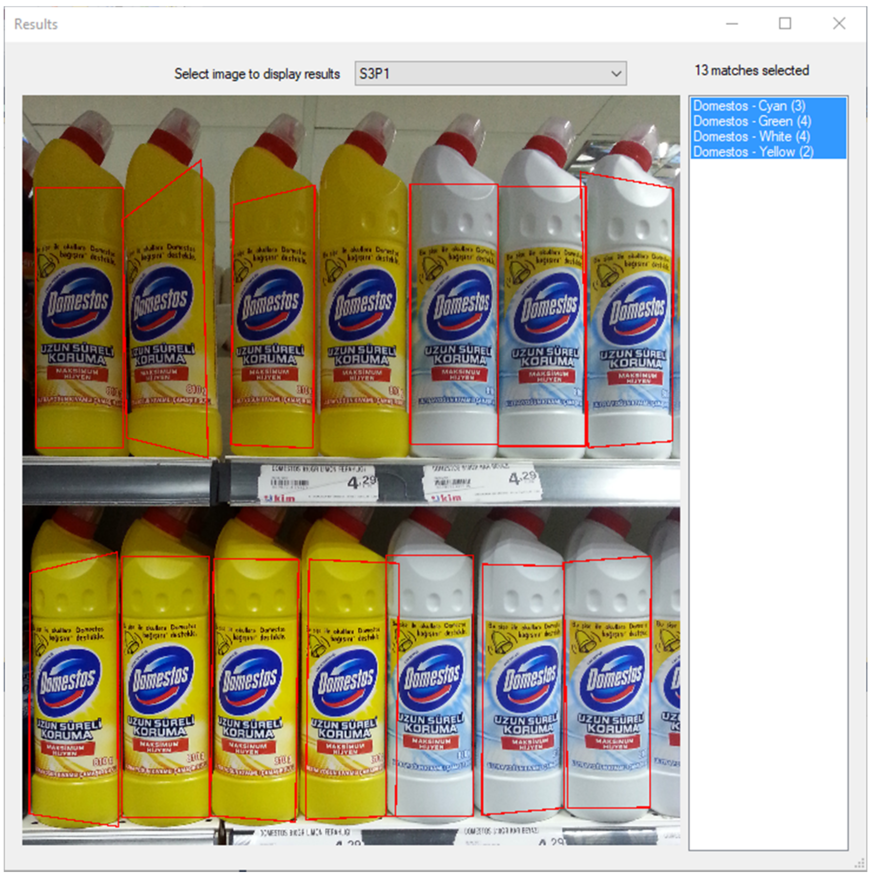



Figure 4.1 Sample results for experiment A

### 4.2. Experiment B

In this experiment, one model image is also used per each product in the database. A total of 32 entries exists in the products' data table of this scenario. The overall matching factor, which is calculated using the standard scheme of the proposed method that relies on both

**the matching ratio and the color average**, is used in this experiment to classify products into sub-categories. The match with the highest overall matching factor is selected as the detected product and is copied to the results data table. The products recognized using this scenario are 641 products (95%) out of the 675 products in the shelves images. The remaining 34 products (5%) are not detected by the method as valid matches. The number of products that are correctly classified is 564 products. This number represents 88% of the 641 detected products and 84% of the total number of products in the shelves images. The remaining 77 products are wrongfully classified, which means that these products are detected as different products than they actually are. The confusion matrix, shown in Appendix C, describes the results of this experiment. The improvement in products classification shows the important role of the addition of color matching to the overall matching factor. The number of detected products remains the same, compared to experiment A, because the detection part of the application relies only on the SURF matching. The total execution time for this experiment is 8.71 seconds per product.

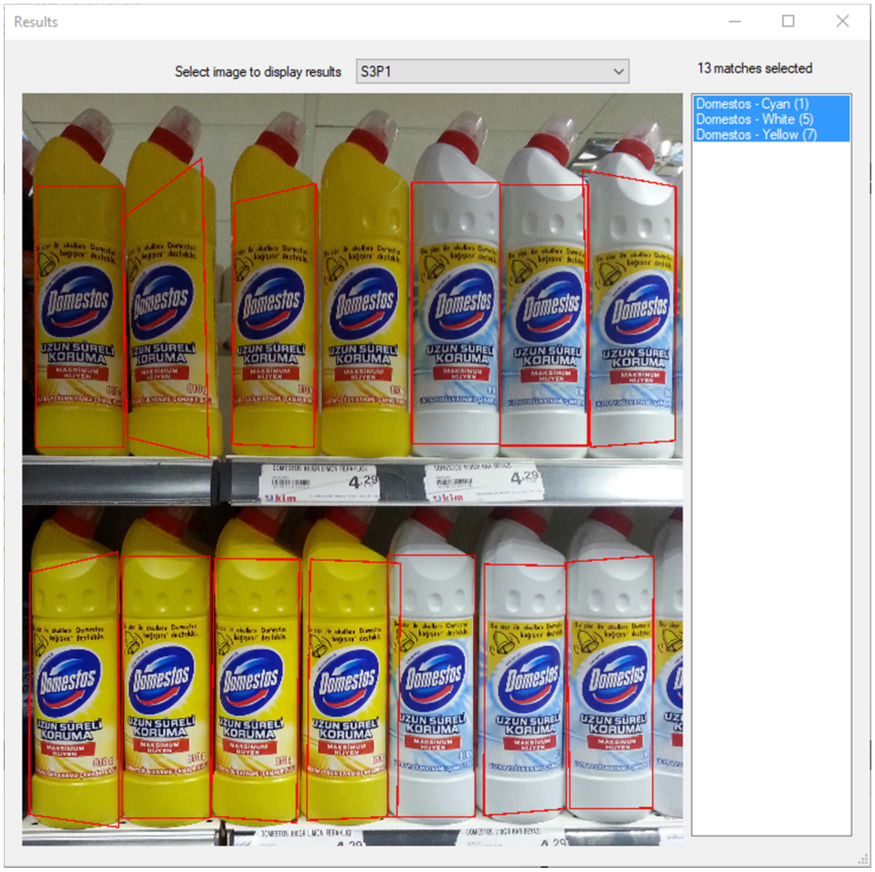



Figure 4.2 Sample results from experiment B

### 4.3. Experiment C

**Three model images** are used for each product in this experiment. These images are cropped from shelves images with different illuminations and view angles. An example of model images added for one product is shown in Figure 4.3. The total number of entries in the products' data table is 96 entries distributed as three entries per each of the 32 products used in the experiments. The proposed method detected 668 products (99%) out of the 675 products in the shelves images while 7 products are not detected at all. The total number of correctly classified products is 648, which is 97% of the detected products and 96% of the total number of products in shelves images, while 20 products are detected but wrongfully classified. A sample of a wrongfully classified product is shown in Figure 4.4. The confusion matrix for this experiment is shown in Appendix D.


Figure 4.3 Model images for one product

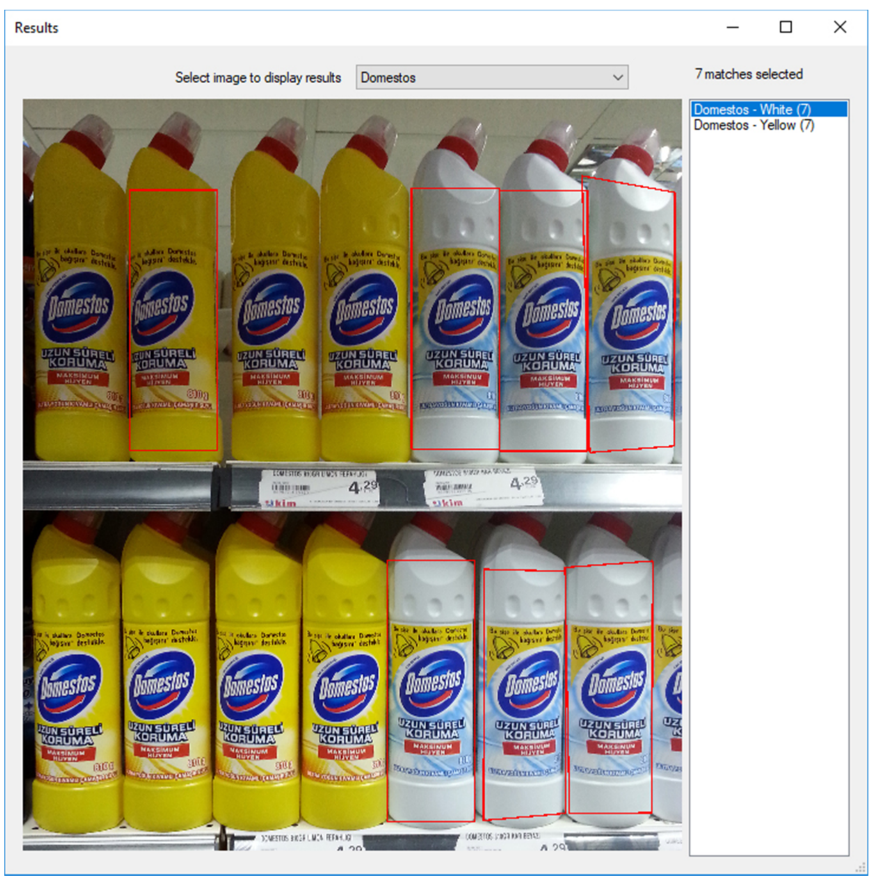



Figure 4.4 A wrongfully classified product

The results of this experiment show improvement in detection as well as classification when compared to both experiment A and B. Sample results of this experiment are shown in Figure 4.5. The number of detected products increased, because of the larger training data generated when three model images are used for each product, and the products are more accurately classified into sub-categories. The time consumed by the detection and classification in this experiment is 19.58 seconds per detected product.

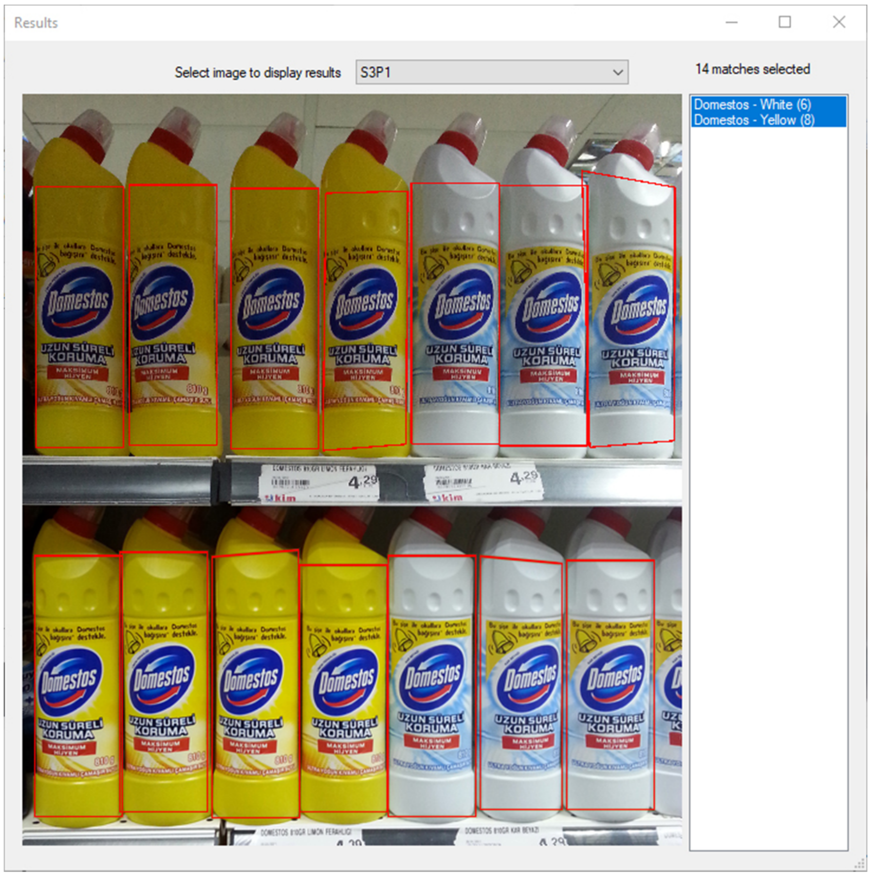



Figure 4.5 Sample results from experiment C

### 4.4. Results Summary and Discussion

In summary, the distribution of the detected products for all experiments among the shelves images taken from retail stores is shown in Table 4.2.

Table 4.2 Summary of detected products distribution

| Store | No. Of images | No. of products | Experiment A | | Experiment B | | Experiment C | |
|---|---|---|---|---|---|---|---|---|
| | | | Detected products | Correct results | Detected products | Correct results | Detected products | Correct results |
| S1 | 12 | 174 | 164 | 102 | 164 | 144 | 171 | 165 |
| S2 | 15 | 183 | 174 | 127 | 174 | 156 | 181 | 178 |
| S3 | 10 | 158 | 150 | 73 | 150 | 128 | 158 | 148 |
| S4 | 13 | 160 | 153 | 89 | 153 | 136 | 158 | 157 |
| Total | 50 | 675 | 641<br>95% | 391<br>61% | 641<br>95% | 564<br>88% | 668<br>99% | 648<br>97% |
| Execution Time (Second) | | | 7.11 | | 8.71 | | 19.58 | |

For a better understanding, the precision and recall are calculated for each experiment. The precision represents the fraction of the correctly classified products among the total number of detected products while the recall is the ratio of the correctly classified products to the total number of products in the shelves images. Using the data in Table 4.2, the overall precision and recall for each experiment are calculated and shown in Table 4.3.

Table 4.3 Precision and recall table

| Store | Experiment A (SURF) | | Experiment B (SURF+Color) | | Experiment C (SURF+Color+Three Model Images) | |
|---|---|---|---|---|---|---|
| | Precision(%) | Recall(%) | Precision(%) | Recall(%) | Precision(%) | Recall(%) |
| S1 | 62.20 | 58.62 | 87.80 | 82.76 | 96.49 | 94.83 |
| S2 | 72.99 | 69.40 | 89.66 | 85.25 | 98.34 | 97.27 |
| S3 | 48.67 | 46.20 | 85.33 | 81.01 | 93.67 | 93.67 |
| S4 | 58.17 | 55.63 | 88.89 | 85.00 | 99.37 | 98.13 |
| Overall | 60.51 | 57.46 | 87.92 | 83.50 | 96.97 | 95.97 |

The good performance of the SURF, even when used with relatively small training data, is well illustrated in experiment A, where 95% of the products are detected using only one model image per product for training. The low precision and recall values of this experiment are according to the use of different products, from the same brand, that are mainly different in color and the fact that no color matching is implemented during this experiment. Thus, the SURF method is evaluated depending on the number of detected products only. A closer look at the confusion matrix for this experiment shown in Appendix B, shows that there is no mix up in brand recognition, which is the responsibility of the SURF detector, and the confusion occurs only on the subcategory classification level. Thus, for the SURF method evaluation only, the recall is 95% while the precision is 100% for brands classification. This perfect precision is achieved by examining the matching results returned from the SURF method against a set of constraints to filter out all the false positive matches as explained in section 3.2.

The sub-category classification is well improved when the color matching factor is added to the overall matching factor. The deviation of color average between the model image, matched by the SURF, and the detected region in the shelves image is calculated and combined with the ratio of the matched interest points to the total number of interest points detected in the model image as explained in section 3.2. Experiment B shows the effect of combining these factors to create an overall matching factor used for classification. The total number of detected products remains the same for experiments A and B. This is according to the fact that the product detection process is not related to the color matching factor and is entirely dependent on the SURF method. On the other hand, the sub-category classification is significantly improved with a relatively small increase in execution time. The execution times are 76 minutes for experiment A and 93 minutes for experiment B. No changes other than the addition of color matching factor between both experiments. Thus, the calculations of color matching factors consumed 17 minutes, for the entire matches found, which represents only 18% of the execution time.

The precision is improved by about 9% in experiment C, when compared to experiment B, by using three model images per product in the training data and the number of detected products is only 4% higher. These improvements consumed 234% of the time consumed in experiment B where only one model image per product is used in training data. The improvements in the results are low compared to the relatively high execution time but, on the other hand, the number of model images used per product requires no modification in the hierarchy of the algorithm or the application and is completely done in the runtime by the user. Thus, the need of such improvement, with that increment in execution time, can be evaluated and changed by the user of the application.

Another comparison between the results of experiment C, and the results of the experiments conducted by Varol and Kuzu [25] shows the superiority of the SURF method over the SIFT, keeping in mind that SURF method is much faster than the SIFT as shown by the experimental results provided by Juan and Gwun [29]. The best results reached by Varol and Kuzu [25] are 94% recall and 81% precision while Table 4.2 shows that experiment C has approximately 96% recall and 97% precision.

The percentage of detected products, precision, recall and execution time are illustrated in Figure 4.6. The execution time values are normalized for better illustration.

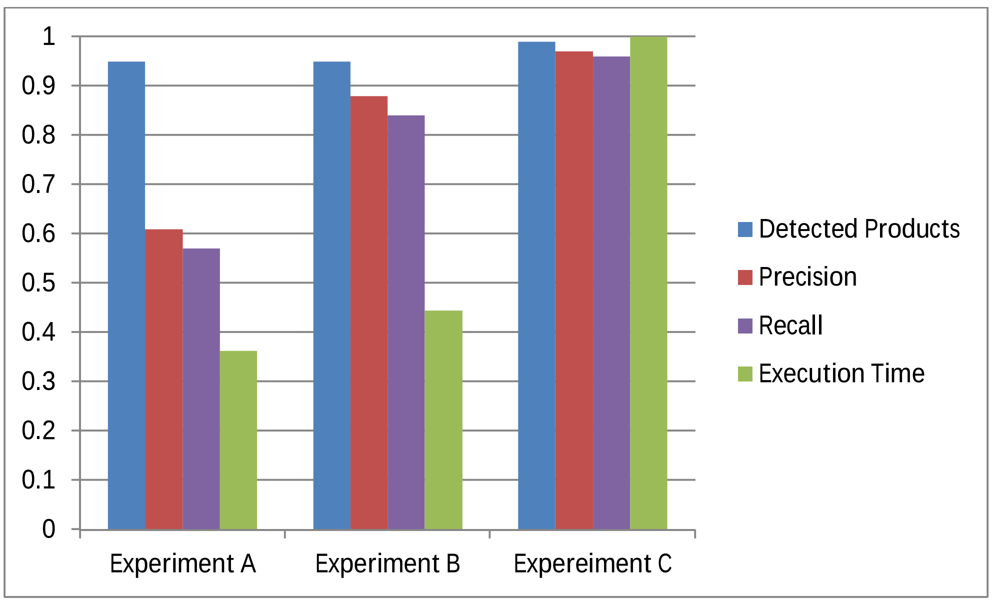



Figure 4.6 Experiments evaluation

The constraints used in the algorithm to check SURF's matching results prior to any further processing has shown significant improvement in the recognition results, when compared to the results of the experiments conducted by Kejriwal et al. [27], by eliminating false positive matches. Although these false positives are mostly going to be dropped in the results validation stage, which is the last stage of the algorithm, dropping such results in the early stages helps to reduce the execution time by eliminating unnecessary further processing of these false positives. Thus, these constraints improved the overall performance of the algorithm by increasing the recognition accuracy and reducing the execution time.

The use of color average, as an addition to the matching factor, helped to improve the accuracy of the classification even when the same training data is used as shown in experiments A and B. This combination outperforms many combinations attempted by George et al [20] where highest accuracy achieved, by combining three methods, is 61.9%. This accuracy may be compared to the recall value of experiment B which is 84%. The methods combined by George et al [20], to score the highest recognition accuracy, are discriminative patches, 2x2 spatial pyramid and the machine learning model, support vector machines (SVM).

In experiment C, the use of three model images per product in the training dataset resulted in more accurate classification as well as better products positioning. Although the execution time is relatively high when compared to experiments A and B, it is still performing faster and more accurate positioning than the sliding window method used by Skoczylas [28] for better product recognition and better positioning. The sliding window moves in overlapping steps to ensure better products recognition regardless of the location and size of the product. This overlapping causes repeated processing of some regions, unless a product is detected in the region, which results in higher execution time. On the other hand, the validation stage at the end of the proposed algorithm makes it possible to process the entire image simultaneously and choose the best matches as recognition results.

# CHAPTER 5

# CONCLUSION

In this study, a new algorithm is proposed for products recognition in shelves images of retail stores. The proposed algorithm is based on computer vision and uses SURF method for features extraction and matching. The matching results are processed further in order to deliver accurate and reliable results to the retail stores owners. The first step is to check each match against a set of constraints. Then the information of that match is stored in an intermediate table if it passes this check alongside with the matching factor, which indicates the similarity of the detected shapes as long as the color average deviation between the model image, that is being searched for, and the region detected in the shelves images where that product exists.

Later, the best match among the matches detected in a specific region is moved to a separate table in the database that is designed to store the final results. This is done by creating a new blank image, that is identical to the shelves image being processed, and check if the detected region is blank on that image or not. If the region is blank, it is considered as a valid match, moved to the results table and the region is marked in black so any further less matched results are not validated. This method is found to be much faster than using the SURF method again for matches validation.

After the results are stored in the database, any kind of further processing can be done directly using that table without the need to repeat the recognition process. This ease of access for the results enables providing many services for the retail stores owners. In the application, built to test the proposed method, the results are summarized per each shelves image. The summary shows the number of products from each brand and sub-category in the selected picture and the position of these products.

Three experiments are conducted to test the results of the proposed method. The first experiment demonstrates the effect of using predefined constraints to judge whether the SURF matching results are acceptable or not. The second experiment shows the improvement of products classification into corresponding sub-categories when the color matching factor is used alongside with the matching factor concluded from the SURF method. In the third experiment, three model images are used per product in training dataset to improve the performance of the algorithm in such scenario. All images used in these experiments are real life images taken in different non-controlled environments. The model images are cropped from shelves images that are not included in the experiments.

Experiment (A) shows an improvement in recognition accuracy when compared to other experiments conducted in different studies. The constraints eliminate any false positive recognition produced by the SURF method, which is generated because of the repeated products in the shelves images causing the SURF method to detect a matching point from one product and another point from another product in the image. All brand names of the recognized products are correctly set, but the sub-category classification is very confusing in this experiment as the color matching factor is not used and the SURF method converts all images to grayscale prior to any processing. Thus, products of the same brand that are mainly different in colors are not accurately classified into their sub-categories.

In experiment (B), the color matching factor is added to the overall matching factor. This resulted in a significantly more accurate classification of detected products into their sub-categories. On the other hand, the execution time is slightly higher than the time consumed in experiment A. Thus, this experiment shows the important role of using the color matching factor alongside with the results of the SURF method.

In experiment (C), three model images are used per product, instead of one in experiment (B), for training instead of only one in the previous experiment. The accuracy of products classification into their corresponding sub-categories is slightly improved, keeping in mind that the accuracy in the second experiment is already high enough, while the execution time increased rapidly when compared to the other experiments.

In summary, the proposed algorithm outperforms most of the known methods that use computer vision for the purpose of products recognition. Changing the number of model images per product requires no change in the way the algorithm works. Thus, it is up to the user to decide the appropriate number to be used, depending on the required accuracy and the flexibility of the execution time, which is also affected by the specifications of the computer used to run the application. Although the results of the second experiment show a huge improvement in results, when compared to the first experiment, with relatively very slight increment in execution time.

For future work, it is recommended to develop more services for the store owners using the results of this algorithm. The ease of access to the recognition and classification results keeps the door wide open for any further services development. Simple examples of these services are to create an alarm when the total number of a product is less than a threshold value to avoid out of stock situations or when a product exists in a shelves image, which is not supposed to be in, to notify the owner of a misplaced product.

# REFERENCES


[1] Strassmann, P.A.,(1990). "The business value of computers: an executive's guide". Information Economics Press.

[2] Ahmed, S.R., (2004). "Applications of data mining in retail business". in Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on.of Conference.: IEEE.

[3] Gruen, T.W., D.S. Corsten, and S. Bharadwaj,(2002). "Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses". Grocery Manufacturers of America Washington, DC.

[4] Emmelhainz, M.A., J.R. Stock, and L.W. Emmelhainz, (1991). "Consumer responses to retail stock-outs", Journal of retailing. 67(2). 138-148.

[5] Corsten, D. and T. Gruen, (2003). "Desperately seeking shelf availability: an examination of the extent, the causes, and the efforts to address retail out-of-stocks", International Journal of Retail & Distribution Management. 31(12). 605-617.

[6] Forsyth, D. and J. Ponce,(2011). "Computer vision: a modern approach". Upper Saddle River, NJ; London: Prentice Hall.

[7] Szeliski, R.,(2010). "Computer vision: algorithms and applications". Springer Science & Business Media.

[8] Steger, C., M. Ulrich, and C. Wiedemann, (2008). "Machine vision algorithms and applications", Illumination. 5(2.1). 1.

[9] Campo, K., E. Gijsbrechts, and P. Nisol, (2003). "The impact of retailer stockouts on whether, how much, and what to buy", International Journal of Research in Marketing. 20(3). 273-286.

[10] Laurent, G., A.-S. Bayle-Tourtoulou, and S. Macé, (2006). "Assessing the Frequency and Causes of Out-of-Stock Events Through Store Scanner Data".

[11] Gruen, T.W. and D.S. Corsten, (2007). "A comprehensive guide to retail out-of-stock reduction in the fast-moving consumer goods industry".

[12] Papakiriakopoulos, D. and G. Doukidis, (2011). "Classification performance for making decisions about products missing from the shelf", Advances in Decision Sciences. 2011.

[13] Papakiriakopoulos, D.A., (2006). "Automatic detection of out-of-shelf products in the retail sector supply chain", International Journal of Retail & Distribution Management.

[14] Rekik, Y., E. Sahin, and Y. Dallery, (2008). "Analysis of the impact of the RFID technology on reducing product misplacement errors at retail stores", International Journal of Production Economics. 112(1). 264-278.

[15] Surjandari, I. and A. Seruni, (2010). "Design of product placement layout in retail shop using market basket analysis", Makara Journal of Technology. 9(2). 43-47.

[16] Kumar, S., et al., (2014). "Remote retail monitoring and stock assessment using mobile robots". in 2014 IEEE International Conference on Technologies for Practical Robot Applications (TePRA).of Conference.

[17] Huang, T., (1996). "Computer Vision: Evolution and Promise".

[18] Li, H., et al., (2013). "A benchmark for semantic image segmentation". in Multimedia and Expo (ICME), 2013 IEEE International Conference on.of Conference.: IEEE.

[19] Martin, D.R., C.C. Fowlkes, and J. Malik, (2004). "Learning to detect natural image boundaries using local brightness, color, and texture cues", IEEE transactions on pattern analysis and machine intelligence. 26(5). 530-549.

[20] George, M., et al., (2015). "Fine-grained product class recognition for assisted shopping". in Proceedings of the IEEE International Conference on Computer Vision Workshops.of Conference.

[21] Bhatia, A., (2007). "Hessian-Laplace feature detector and Haar descriptor for image matching", University of Ottawa (Canada).

[22] Morevec, H.P., (1977). "Towards automatic visual obstacle avoidance", in Proceedings of the 5th international joint conference on Artificial intelligence - Volume 2, Morgan Kaufmann Publishers Inc.: Cambridge, USA. 584-584.

[23] Dalal, N. and B. Triggs, (2005). "Histograms of oriented gradients for human detection". in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.of Conference.: IEEE.

[24] Lowe, D.G., (2004). "Distinctive image features from scale-invariant keypoints", International journal of computer vision. 60(2). 91-110.

[25] Varol, G. and R.S. Kuzu, (2015). "Toward retail product recognition on grocery shelves". in Proc. SPIE.of Conference.

[26] Bay, H., et al., (2008). "Speeded-up robust features (SURF)", Computer vision and image understanding. 110(3). 346-359.

[27] Kejriwal, N., S. Garg, and S. Kumar, (2015). "Product counting using images with application to robot-based retail stock assessment". in Technologies for Practical Robot Applications (TePRA), 2015 IEEE International Conference on.of Conference.: IEEE.

[28] Skoczylas, M., (2014). "Detection of positions and recognition of brand logos visible on images captured using mobile devices". in Electrical and Power Engineering (EPE), 2014 International Conference and Exposition on.of Conference.: IEEE.

[29] Juan, L. and O. Gwun, (2009). "A comparison of sift, pca-sift and surf", International Journal of Image Processing (IJIP). 3(4). 143-152.

[30] Khan, N.Y., B. McCane, and G. Wyvill, (2011). "SIFT and SURF performance evaluation against various image deformations on benchmark dataset". in Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on.of Conference.: IEEE.

[31] Bo, X., et al., (2010). "Image copy-move forgery detection based on SURF". in Multimedia information networking and security (MINES), 2010 international conference on.of Conference.: IEEE.

[32] Yang, J., et al., (2009). "Linear spatial pyramid matching using sparse coding for image classification". in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on.of Conference.: IEEE.

[33] Emgu, C., (2013). "Emgu CV: OpenCV in .NET (C#, VB, C++ and more)", Online: http://www.emgu.com.

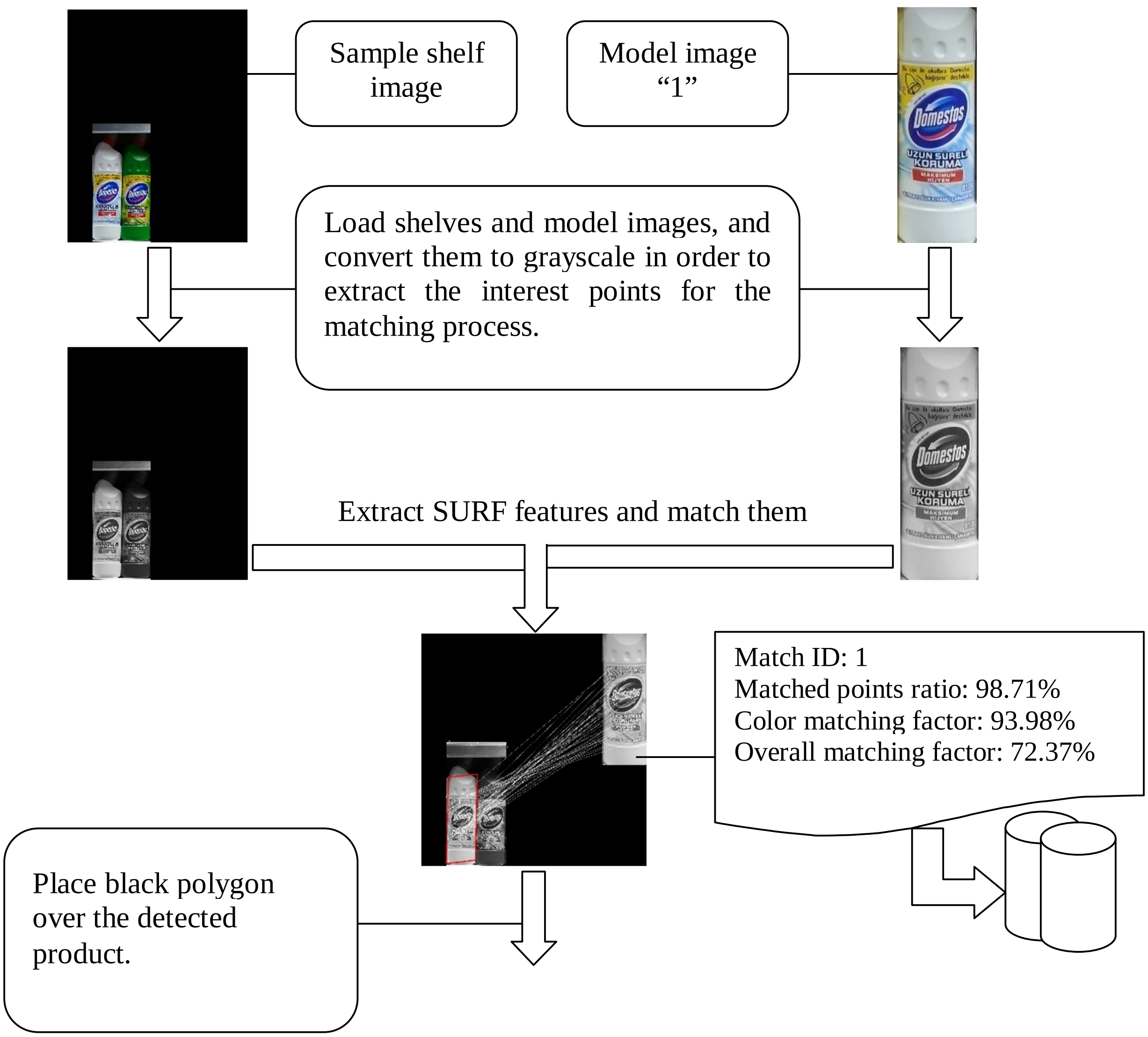
[34] Bradski, G., (2000). "The OpenCV Library", Dr. Dobb's Journal: Software Tools for the Professional Programmer. 25(11). 120-123.

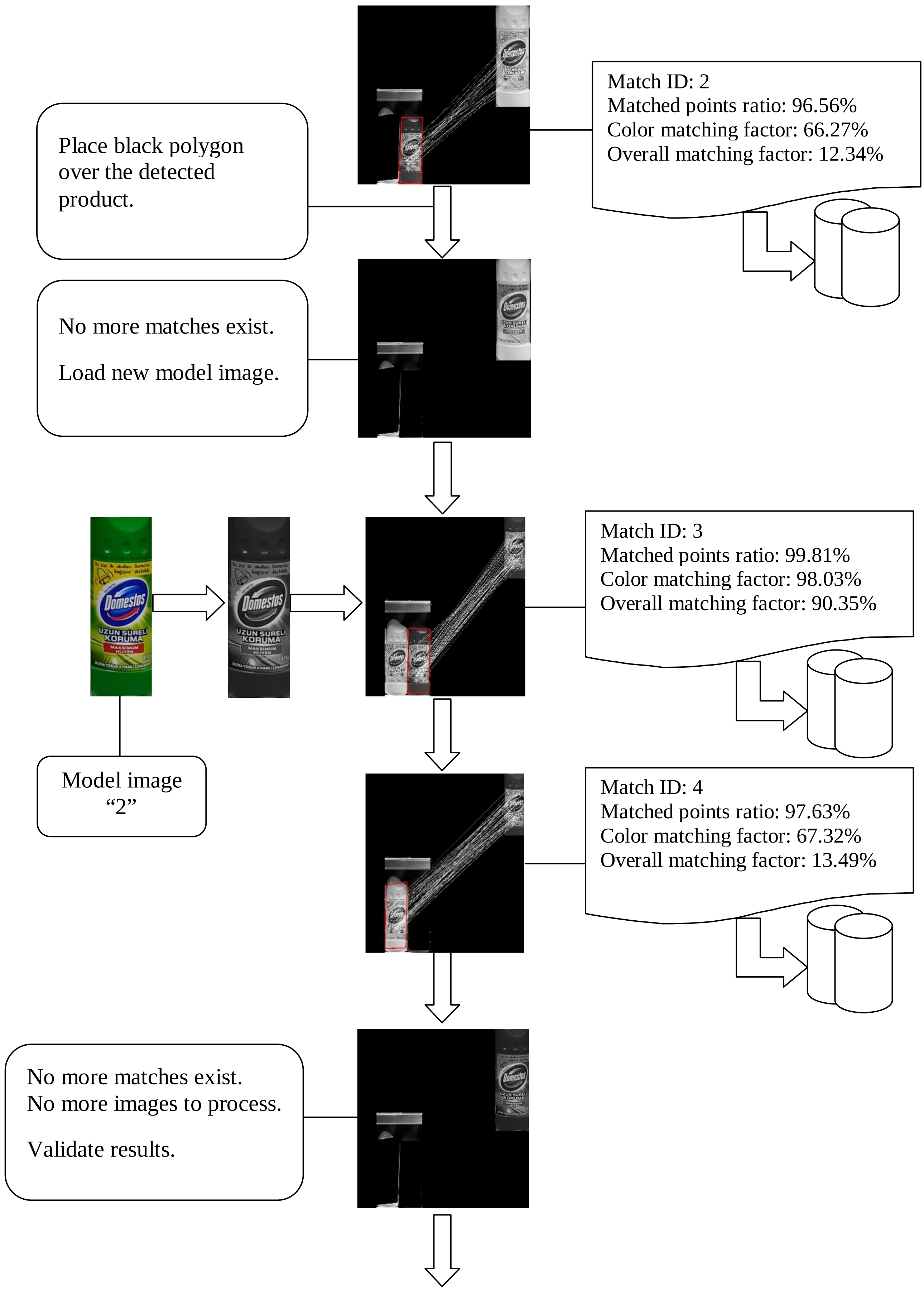
# APPENDIX-A

---

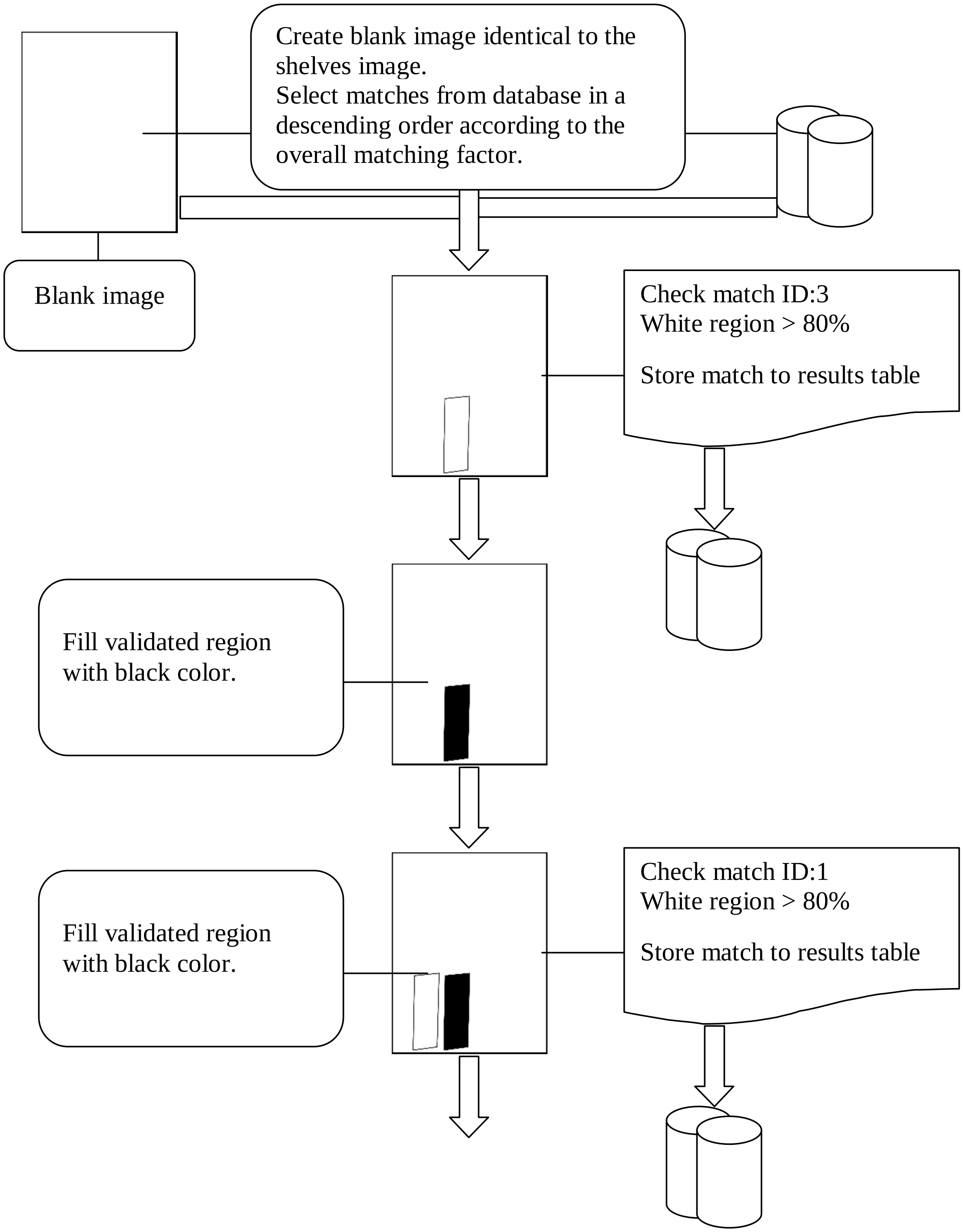
## SAMPLE FLOW CHART

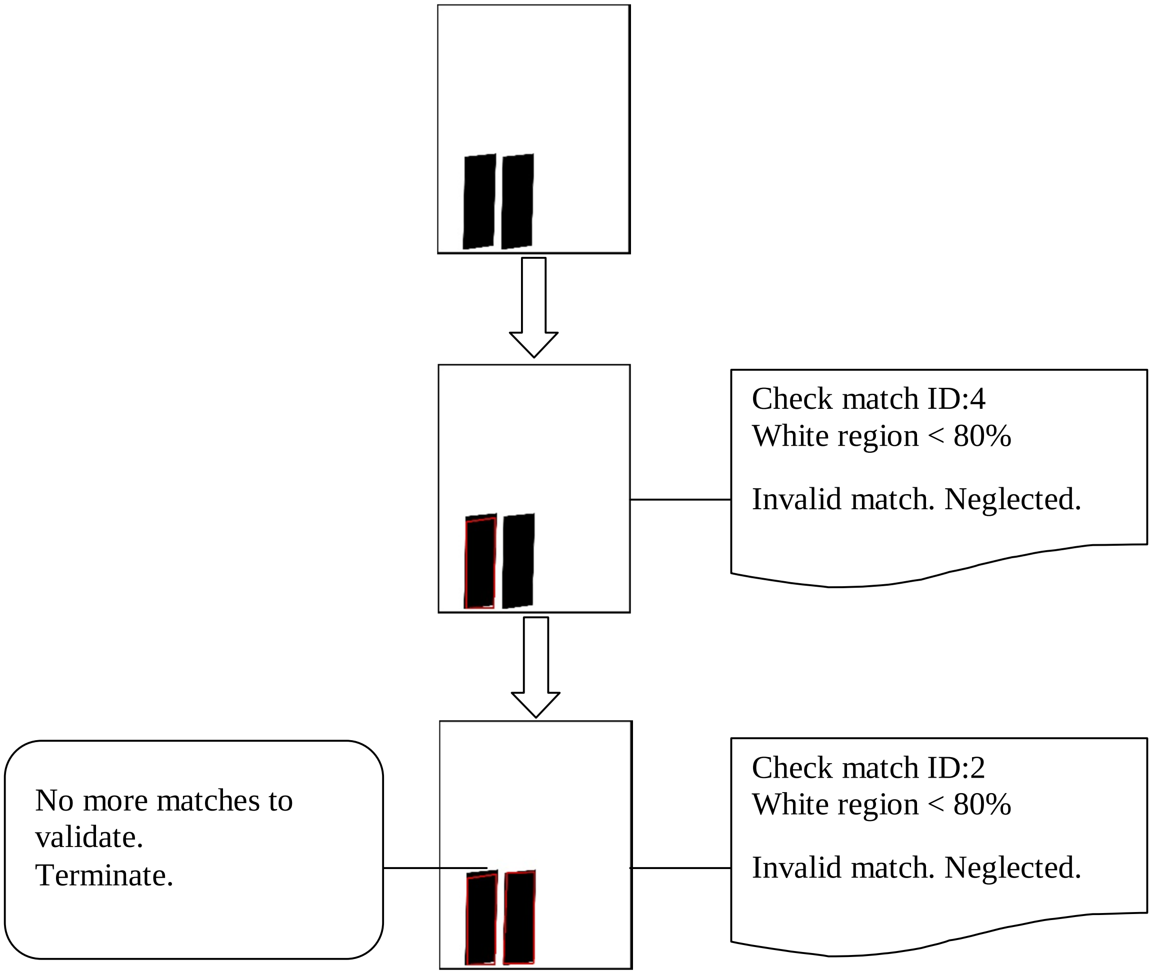
DETECTION PART




Match ID: 2
Matched points ratio: 96.56%
Color matching factor: 66.27%
Overall matching factor: 12.34%
Place black polygon over the detected product.
No more matches exist.
Load new model image.
Match ID: 3
Matched points ratio: 99.81%
Color matching factor: 98.03%
Overall matching factor: 90.35%
Model image "2"
Match ID: 4
Matched points ratio: 97.63%
Color matching factor: 67.32%
Overall matching factor: 13.49%
No more matches exist.
No more images to process.
Validate results.

VALIDATION PART




Check match ID:4
White region < 80%
Invalid match. Neglected.
No more matches to
validate.
Terminate.
Check match ID:2
White region < 80%
Invalid match. Neglected.

# APPENDIX-B

## CONFUSION MATRIX FOR THE RESULTS OF EXPERIMENT (A)

| Actual \ Predicted | | | Detergents | | | | | | | | | | | | | | | Shampoos | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Fairy | | | Domestos | | | | Pril | | | | | Bingo | | | Head & Shoulder | | | | | Loreal | | | Johnson's | | | Elidor | | | | | |
| | | | A | B | C | A | B | C | D | A | B | C | D | E | A | B | C | A | B | C | D | E | A | B | C | A | B | C | A | B | C | D | E | F |
| Detergents | Fairy | A | 25 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 4 | 22 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 3 | 5 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Domestos | A | 0 | 0 | 0 | 8 | 11 | 3 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 8 | 9 | 4 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 5 | 3 | 8 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | D | 0 | 0 | 0 | 1 | 2 | 2 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Pril | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19 | 1 | 3 | 2 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 24 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 4 | 11 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 2 | 2 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Bingo | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 3 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 17 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Shampoos | Head & Shoulde | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 3 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 15 | 4 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 7 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 1 | 9 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 3 | 2 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Loreal | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 3 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 11 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Johnson's Baby | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 17 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 14 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Elidor | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 3 | 2 | 3 | 2 | 2 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 1 | 1 | 3 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 6 | 0 | 0 | 0 |
| | | D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 7 | 1 | 1 |
| | | E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 4 | 9 | 2 |
| | | F | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 1 | 1 | 2 | 1 | 7 |

# APPENDIX-C

## CONFUSION MATRIX FOR THE RESULTS OF EXPERIMENT (B)

| Actual / Predicted | | | Detergents | | | | | | | | | | | | | | | Shampoos | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Fairy | | | Domestos | | | | Pril | | | | | Bingo | | | Head & Shoulder | | | | | Loreal | | | Johnson's | | | Elidor | | | | | |
| | | | A | B | C | A | B | C | D | A | B | C | D | E | A | B | C | A | B | C | D | E | A | B | C | A | B | C | A | B | C | D | E | F |
| Detergents | Fairy | A | 29 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 1 | 27 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 2 | 2 | 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Domestos | A | 0 | 0 | 0 | 20 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 1 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 15 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | D | 0 | 0 | 0 | 1 | 0 | 2 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Pril | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19 | 0 | 1 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 30 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 19 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Bingo | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 23 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Shampoos | Head & Shoulder | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 23 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 15 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Loreal | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 13 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Johnson's Baby | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 17 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 15 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Elidor | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 0 | 1 | 0 | 1 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 10 | 2 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 0 |
| | | D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 1 |
| | | E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 12 | 2 |
| | | F | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 12 |

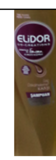
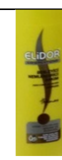
# APPENDIX-D

## CONFUSION MATRIX FOR THE RESULTS OF EXPERIMENT (C)

| Actual \ Predicted | | | Detergents | | | | | | | | | | | | | | | Shampoos | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Fairy | | | Domestos | | | | Pril | | | | | Bingo | | | Head & Shoulder | | | | | Loreal | | | Johnson's | | | Elidor | | | | | |
| | | | A | B | C | A | B | C | D | A | B | C | D | E | A | B | C | A | B | C | D | E | A | B | C | A | B | C | A | B | C | D | E | F |
| Detergents | Fairy | A | 32 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 30 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 2 | 0 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Domestos | A | 0 | 0 | 0 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | D | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Pril | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 27 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 33 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Bingo | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Shampoos | Head & Shoulder | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 19 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Loreal | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Johnson's Baby | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 22 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 16 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Elidor | A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 23 | 0 | 0 | 0 | 0 | 1 |
| | | B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 |
| | | C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 |
| | | D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 0 | 0 |
| | | E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 |
| | | F | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 14 |

## APPENDIX-E

## SAMPLE IMAGE FOR EACH PRODUCT CATEGORY

| **Product Type (#of total categories)** | **Product Name** | **Category** | | | | | |
|---|---|---|---|---|---|---|---|
| Shampoo(17) | Head&Sholder |  |  |  |  |  | |
| | Loreal |   |   |   | | | |
| | Johnson's Baby |  |  |  | | | |
| | Elidor |  |  |  |  |  |  |
| Detergent(15) | Fairy |   |   |   | | | |
| | Domestos |  |  |  |  | | |
| | Pril |   |   |   |   |   | |
| | Bingo |   |   |  | | | |

## CURRICULUM VITAE

### PERSONAL INFORMATION

| | |
|---|---|
| **Name Surname** | : Muhanad Hameed Arif Algburi |
| **Date of birth and place** | : 04/03/1981, Iraq-Diyala |
| **Foreign Languages** | : English, Turkish |
| **E-mail** | : muhanad.hameed.arif.1981@gmail.com |

### EDUCATION

| **Degree** | **Department** | **University** | **Date of Graduation** |
|---|---|---|---|
| B.Sc. | Computer Engineering | Al-Mustansiriyah University | 15/08/2003 |
| High School | Scientific | Al-Markaziyah High School | 09/10/1999 |

### WORK EXPERIENCE

| **Year** | **Corporation/Institute** | **Enrollment** |
|---|---|---|
| 2008 | Al-Markaziyah High School | Teacher |

## PUBLISHMENTS

### Conference Papers

1. Muhanad H. Algburi and Songül Albayrak, “Store Products Recognition and Counting System Using Computer Vision”, CICN2017 Conference, Cyprus, Turkey.