
The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)

Volume 1, Pages 224-228

ICONTES2017: International Conference on Technology, Engineering and Science

ANALYZING THE CONTENT OF DAIRY PRODUCTS USING IMAGE PROCESSING TECHNIQUES

Cagla Yildirim
Pamukkale University

Ceren Atik
Pamukkale University

Meric Cetin
Pamukkale University

Abstract: Considering human health, it is very important for individuals to select the products they will use correctly in order to reduce the effect caused by any allergic or chronic health condition. In addition, health problems caused by nutrients are increasing day by day. Thus, people should carefully check the components and contents of their purchased products in case of any health problem. In this study, it was aimed to determine the nutrients that could trigger some health problems related to the consumption of dairy products and to prevent the consumption of these nutrients. For this purpose, an application which is user friendly and easy to use in daily life has been developed by using image processing based techniques. The mobile application, called "Dairy Products", was developed using Tesseract's Tess-two library on Android Studio with Java language. It was aimed to perform content analysis of dairy products by application developed using Optical character recognition (OCR). The objective of OCR system for general purpose is to convert scanned or photographed pictured format documents into corresponding usable symbolic representation. With this application, the dairy products that should not be consumed as a result of the content analysis can be determined specifically for the person.

Keywords: Optical character recognition, image processing, android mobile applications, content analysis.

Introduction

Today, health problems that people live in as a result of unconsciously consumed foods are increasing day by day. Many people have at least one allergic or chronic condition. Almost all of these diseases are caused by food consumption. Therefore, the individual should carefully evaluate the products they will purchase. People are confronted with many different foods throughout their lives and must have safe food to keep their lives healthy [1]. There are smaller building blocks called nutrients (proteins, fats, carbohydrates, vitamins and minerals) in foods. Food allergies are a result of an immune system that is a consequence of the perception of food components, especially protein, by the body as foreign. Antibodies to these foreign substances in the nutrient can cause unwanted reactions in the body [2].

Food allergies affect 1% of adults and 7% of children. However, they may be seen more often in some children [3]. Food intolerance is seen more frequently. Most people may experience food intolerance-related discomfort in part of their lives. Additive ingredients added to ready-made food can cause allergies in individuals. Some people may tolerate specific substances; the most common of which is "lactose intolerance" and affects 15% of the population. Lactose intolerance cases, the amount of lactase enzyme that is effective in digesting lactose in the digestive tract is absent or reduced [1, 2, 4].

In general, food items are clearly stated on packaged products. However it is a known fact, most people do not check or cannot check on these values. They often do not want to spend time on reading and evaluating content, or those who are susceptible to this issue cannot understand or cannot review the content. Although important

- This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 Unported License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

- Selection and peer-review under responsibility of the Organizing Committee of the conference

*Corresponding author: Cagla Yildirim- E-mail: cagla070194@gmail.com

information is contained in the labels on packaged foods that are readily available, many consumers do not know what the information on the labels mean. In Turkey, where label reading habits are around 50 percent, consumers often look at their expiration date. Consumers do not pay attention to sweeteners or food dyes in different codes in food. The labels on packaged foods that every conscious consumer should read carefully give important information about the health effects of the product we buy. Character recognition process has become one of the main application areas of image processing and pattern recognition [5-8]. Optical Character Recognition (OCR) systems are widely used for the automatic conversion of image-formatted documents into digital form that can be edited in the computer environment. Documents are usually obtained using a scanner or a camera. Mobile devices have become an alternative to scanners with the increased quality of their cameras. In addition, the mobile industry is paying more attention to the barcode applications that users need. Almost all products have its unique barcode numbers. These unique number sequences represent a lot of information about a product.

In this study, a mobile application aimed at analyzing the content of dairy products has been developed, taking into account the daily use of individuals. This application determines using barcode recognition for content and quantities of the substances that could trigger some health problems in the consumption of dairy products. In this way, it may be possible for individuals with certain health problems to gain knowledge of which foods may or may not consume harmful nutrients. The developed application provides an opportunity users to obtain information about nutritional facts of dairy product practically in their daily life.

Image Recognition

Image recognition is the process of finding common features and classifying objects of interest or various patterns in the form of signs. The main purpose in image recognition is to classify and describe a pattern [9]. Complex signs or objects that can be correlated with each other through their common properties can be classified by some identified properties or characters [10].

In order to be able to process a computer image, it has to be made suitable for quantification. For this reason, sampling and quantization steps are applied to digitize an image data. The pre-processing step is the process of making the image to be processed suitable for other image processing steps. A feature extraction is the process of extracting common properties belonging to the class that the data to be processed has. Various feature extraction methods are used for this purpose (Karhunen-Loeve transformation, angular-radial transformation, Zernike moments and the presence of geometric attributes). This feature extraction process may vary depending on the type of data [11].

Optical Character Recognition

OCR (Optical Character Recognition) is the translation and digitization of documents written by a tool or machine or by handwriting into computer language. Character recognition is one of the most frequently used pattern recognition processes today [9, 12-15]. The numerical data must be evaluated with the basic steps of image processing to understand data.

The image processing steps to be used for character recognition can be described as follows: Image restriction, image capture, preprocessing, feature extraction, interpretation [13]. Image restriction is the first operation of the pattern recognition system. There are two main purposes of the image restriction process. The first is to increase the use of the unprocessed image, and the other is to restrict the image to the extent possible to remove image analysis problems. Image capture is the digitization of the image to be given to the system. Pre-processing is a very important step in a pattern recognition system. This step basically consists of two sub-processes; noise filtering and edge enhancement. Segmentation is the separation of the subspecies that make up the image itself [10]. The main task of feature extraction is to specify the character points to be defined from the binary image. In this step, the common characteristics of characters are very important. These features can be features like shape, size.

Designed Mobile Application

This section gives the technical details of designed application. The developed mobile application is based on OCR system. Character recognition is the processing of text based data by the machine to produce meaningful output. The mentioned application has been developed with Java language in Android Studio and Tesseract Tess-Two library. In addition, a free software compatible with the Android system operating, has been used for character recognition in product barcode.

Tesseract Tess-Two Library

Tesseract is a free optical character recognition motor developed for various operating system. Developed with Tesseract, OCR software has a structure that recognizes and reads the black dot community, letters, numbers and symbols.

Working Principle

Each words is analyzed separately and the pages are divided into lines. During analysis the structural characters, heights and widths of the letters are evaluated. The information store of the program is filled with various parameters defined for each character, and if the detected pixels match on of these parameters, it becomes possible to reveal the alphabet.

The Software Mechanism of the Application

Database System

The application was developed with SQLite, which is the local database of Android systems. In the database, the barcode numbers of different dairy products of different bands and the nutritional facts they contain are registered (Figure 1.a). In addition, tables and fields are prepared in the database in order to enable users register their profile information.

User Information Registration

At this stage, it is aimed that the users choose the nutritional facts which may be harmful to them and to save in their own profile (Figure 1.b).

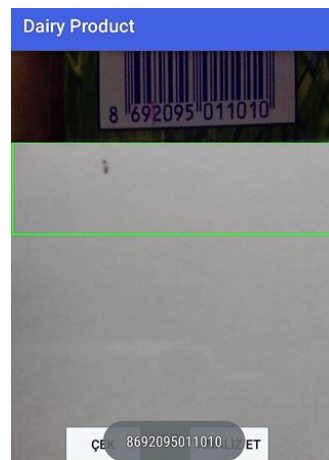


Figure 1a. Database

A screenshot of a mobile application interface titled "Dairy Product" for user registration. It contains several input fields: "Çağla", "Ceren", and "23". Below these is a section titled "Select the item you are allergic to" with a list of items: Cinnamon (checked), Vanilla (checked), Strawberry (unchecked), Banana (checked), Chocolate (unchecked), and Lactose (checked).

Figure 1b. User information

Extraction and Processing of Barcode Image

When the users analyzes the barcode image of the dairy product they wants to consume, the barcode number is recognized with the help of various methods and is matched in the database. Finally, considering the profile of the user, if there is a harmful substance in the product, it is determined and informed to the user (Figure 2).

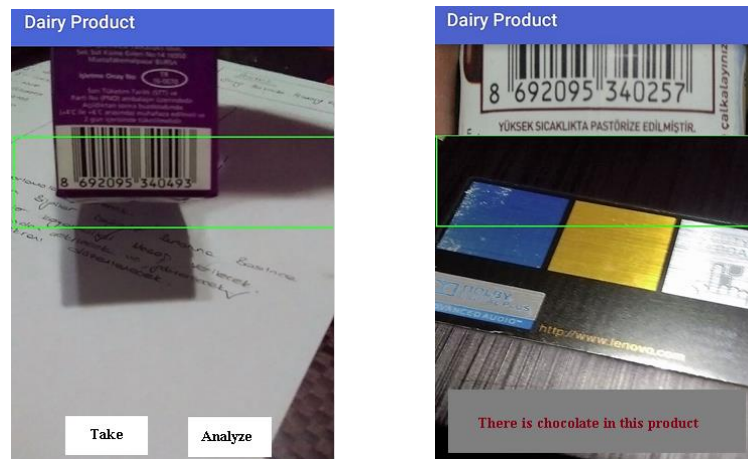


Figure 2. Processing of barcode image

The application has been tried many times under various conditions and achieves on average of 70 percent success. Based on the test results, the reasons for failure of 30 percent were researched and evaluated. These evaluations show that the cause of failures depends on the slope of the view and also on the light in the environment.

Conclusion

OCR technologies are used to solve many problems. The most important reason for the increased use of this technology is its ability to transfer thousands of page of data to computers or databases in a short period of time. It is aimed to provide healthy food consumption for the individuals by utilizing OCR technology in this study. The content analysis of dairy products could be carried out by using the image processing techniques of the barcode images on the products. The application asks the user to enter a list of allergenic foods that are objectionable to be consumed. It then analyzes the data by taking the barcode image as the parameter of the product to be analyzed and informs the user whether or not he can consume the product. Thanks to this, users with different health problems can avoid the symptoms that they may encounter in the consumption of dairy products.

Recommendations

This study has been developed only for dairy products, but has the capacity to be applied to all products. It can be used in all products with complicated and more comprehensive database. Another way to improve is to create a more detailed user profile. The user profile can be generated by taking the parameters of the patient's assay results so nutritional consumption advice can be offered almost all illnesses.

References

- Öztürk, U. D. M., & Besler, H. T. (2006). Besin alerjileri.
- Boğa, A., & Binokay, S. (2010). Food additives and effects to human health. Archives Medical Review Journal, 19(3), 141-154.
- Hocaoğlu, A. B. Çocukluk Çağında Astım ve Gıda Alerjileri.
- Göney, G., Yalçın, A. D., & Ünitesi, A. ALERJİ VE ALERJENLER.
- Burges, C. J. C., Be, J. I., & Nohl, C. R. (1992, November). Recognition of handwritten cursive postal words using neural networks. In Proc. USPS 5th Adv. Technology Conf (Vol. 117).
- de Campos, T., Babu, B. R., & Varma, M. (2009). Character recognition in natural images.
- Kharna, N. N., & Ward, R. K. (1999). Character Recognition Systems for the non-expert. IEEE Canadian Review, 33, 5-8.
- Li, X., & Yeung, D. Y. (1997). On-line handwritten alphanumeric character recognition using dominant points in strokes. Pattern recognition, 30(1), 31-44.
- Avci, E., & Akpolat, Z. H. (2006). Speech recognition using a wavelet packet adaptive network based fuzzy inference system. Expert Systems with Applications, 31(3), 495-503.
- Gonzalez, R. C., & Woods, R. E. (2002). Digital image processing prentice hall. Upper Saddle River, NJ.

- Engin, A. V. C. I. (2007). Doku tipi imgelerin sınıflandırılması için bir uyarlamalı entropi tabanlı dalgacık-yapay sinir ağı sistemi. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 22(1).
- Mao, J., Sinha, P., & Mohiuddin, K. (1998). A system for cursive handwritten address recognition. In *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on* (Vol. 2, pp. 1285-1287). IEEE.
- Ng, C. M., Ng, V., & Lau, Y. (1999). Regular feature extraction for recognition of Braille. In *Computational Intelligence and Multimedia Applications, 1999. ICCIMA'99. Proceedings. Third International Conference on* (pp. 302-306). IEEE.
- Stutz, M., Thomas, V. M., & Saar, S. (2004, May). Linking bar codes to recycling information for mobile phones. In *Electronics and the Environment, 2004. Conference Record. 2004 IEEE International Symposium on* (pp. 313-316). IEEE.
- Liu, X., & Lin, C. (2007, April). Information Management System of Grocery Production Processing Based on a Bar Code Identification Technology. In *Anti-counterfeiting, Security, Identification, 2007 IEEE International Workshop on* (pp. 164-168). IEEE.