Product Reading For Visually Impaired Persons

Gayatri Adhav\textsuperscript{1}, Mandar Gawade\textsuperscript{2}, Shivprasad Gadgile\textsuperscript{3} & Suraj Ghuge\textsuperscript{4}
\textsuperscript{1,2,3,4} B.E Students of Information Technology, International Institute of Information Technology, Pune

Abstract: Product’s Printed label became one of the essential, to get details of the product. It became tedious for blind people to read this text. We present a camera-based product information reader to help blind persons to read information of the products. Camera acts as main vision in detecting the label image of the product, then image is processed internally and matched with preloaded image in database and finally identifies the product name and identified product information is pronounced.

1. Introduction

We proposed a camera based framework to help blind person to read the text on various super markets products. An android application captures image of product label with help of phone camera. This image is then sent to the main server that is database server through the network where the main processing is done. Then the image undergone through the preprocessing. The main server is used for deploying our proposed algorithms, includes preprocessing of image, matching the processed image with previously loaded product images in the database to recognize the product. The audio output device such as a Speaker informs the blind user about the name and specifications of the product held by user.

Today, there are already a few systems that have some promise for portable use, but they can’t solve the problem efficiently. For example, portable barcode readers designed to help blind people identify different products in an extensive product database can enable users who are blind to access information about these products. Some reading assistive systems such as pen scanners might be employed in these and similar situations.\textsuperscript{[1]} These systems have integrated OCR software to offer the function of scanning and recognition of text and some have integrated voice output.

Above systems are generally designed for and perform best with document images with simple backgrounds, standard fonts and well-organized characters rather than commercial product boxes with multiple decorative patterns. Most OCR Software cannot directly handle images with complex backgrounds and variable fonts.

There have been few software application are available in market for blind people to read text from product but none of the software works in case of complex background with irregular fonts. To assist blind persons to read text from complex background and irregular pattern of super market products, we are going to propose an efficient solution.

All the mentioned problems are get avoided in our system. Our system works as follows. User capture product image from phone through android application, it send image to the server. Once we get the image of object, we apply the pre-processing on images like noise removal, normalization etc. Extract each color feature like RGB feature (Color Space) from the capture image. Captured Image feature will match with the template image which stored in the database. If matching image is found in database, user will get informed about the product name and its specifications through audio device. Shop owner can add, update, and delete the Product.

2. Existing System

\textsuperscript{[2]} A barcode is an optical machine-readable representation of data relating to the object to which it is attached. Originally barcodes systematically represented data by varying the widths and spacing of parallel lines, and may be referred to as linear or one-dimensional (1D). Later they evolved into rectangles, dots, hexagons and other geometric patterns in two dimensions (2D). Although 2D systems use a variety of symbols, and they are generally referred to as barcodes as well. The captured two dimensional signals are sampled and quantized to yield digital images. It can be read by optical scanners called barcode readers, or scanned from an image by special software. Limitation to this technique is that it is very hard for blind users to find the position of the bar code and to correctly point the bar code reader at the bar code.

\textsuperscript{[3]} K Reader Mobile software package runs on a multifunction cell phone and allows users to snap a picture of virtually any document, including mail, receipts, handouts, memos and many other documents. The disadvantage of K Reader Mobile is that document must be nearly flat, placed on a clear, dark surface and should mostly text. It can’t identify
text written in various colors or text colored background.

[4] Pen scanners are reading-assistive systems, these systems are generally designed for reading and perform best with document images with simple backgrounds, standard fonts, a small range of font sizes and well organized characters rather than commercial product boxes with multiple decorative patterns. As it’s designed for simple background, standard font and well organized character, it fails to work in case supermarket product where product has text in different font with irregular character alignment.

3. Proposed System

To overcome problems of existing systems, we proposed system where user captures product image from phone and it’s send to the main server. Once image received, it undergoes through pre-processing like RGB to Gray Scale, Noise Removal, and Normalization. Initially image is converted to Gray scale by taking the average of each pixel RGB. In Normalization, It changes the range of pixel intensity values of image to avoid mental distraction or fatigue from the images. In noise removal, we are removing errors which occur during image acquisition process, so processed image reflect true intensity of real event. After preprocessing, feature extraction is applied in which it extract each color feature like RGB feature (Color Space) from the capture image. Finally, Captured Processed Image will match with the preloaded image which is stored in the database. If product image and details is present in database then it will produce audio output to blind user and provides all necessary details about product else it produce audio output stating product not found.

4. Hardware & Software Requirement

4.1. Hardware Requirement

A] Server side
   Processor: 2.6 GHz Intel Pentium IV or equivalent
   Memory: 2 GB
   Disk space: 1 GB of free disk space
   Dedicated Mobile (For sending SMS)

B] Client side
   Processor: 1.2GHz cortex
   RAM: 512MB
   Storage: 100MB
   Android phone with Camera
   Speaker or Headphone

4.2. Software Requirement

A] Server side
   Eclipse
   JDK 1.6 or above
   Netbeans
   MySQL
   XAMP
   MYSQL GUI Browser 1.2 or above
   Microsoft Windows 7 Professional or greater

B] Client side
   Operating system: Android 3.2 and greater

5. Architecture and Implementation

System architecture for product reading for visually impaired person is given below.

**Figure 1. System Architecture**

To help blind people to read the information of product, we implemented a framework consists of five components.

1. Image Capture
2. Pre-processing
3. Image Matching
4. Audio Output
5. Manage Product
5.1. Image Capture

We have developed an android application which is used to capture image of the super market product and to give audio output. We have used JDK and Eclipse as a platform. The application is connected to the main server via network. HTTP server is created for communication purpose. As the user starts the android application, camera starts automatically. User has to click anywhere on the screen to capture the image. As soon as the image is captured, this bitmap image is sent automatically via network to the main server.

5.2. Image Preprocessing

Actual processing of image is done at main server. We have used java environment and netbeans as a platform for main server. Once the image is received at main server, various pre-processing’s like RGB to Gray conversion, Image normalization and noise removal are performed. Initially we converted image to Gray scale by taking the average of each pixel RGB. In Normalization, It changes the range of pixel intensity values of image to avoid mental distraction or fatigue from the images. At last in preprocessing noise removal is performed, in this we are removing errors which occur during image acquisition process, so processed image reflect true intensity of real event.

5.3. Image Matching

We have used OpenCV framework for image matching. It is open source C++ library for image processing. It has C++, C, Python and java interfaces. Pre-processed image is matched with preloaded database images. Speed up Robust Feature (SURF) algorithm is used for matching which works as follows: Algorithm automatically detects interest points of image. These points do not change with change in position or orientation of the image. Each interest point has descriptor that does not depend on feature scale or rotation. Then descriptor of captured image is paired with descriptors of each database image and compared. The eligibility criteria are that the value of matching pairs should be greater than threshold value which is pre-calculated. The image among all eligible images with highest matching value is resultant image.

5.4. Audio Output

The specifications of matched image are sent to the android application and user is informed about product name and other specification through audio output. In case the image is not matched, the user is notified by audio notification.

5.5. Manage Product

We have created UI for admin by using Java Swing which is connected to the database. The shop owner will be the admin of system. He has to maintain the own database. He can add or delete products. Also he can update the information regarding any product.

6. Performance Metrics

We used Precision, Recall, F-Score to evaluate performance of the system. We calculate precisions and recall rates based on number of correctly matched product images to evaluate the efficiency and robustness of the system.

Following figure shows the confusion matrix also known as error matrix that describes the performance of the system.

![Confusion Matrix](image)

When we matched the captured image of product with preloaded database images, system gives specification of matched database image predicted image as an output. This predicted image may or may not be same as actual image of product. To evaluate the system performance, we calculated four parameters i.e. true positive, true negative, false positive, false negative ratios. True positive: When system predicts, captured image which is same as actual image. True negative: When captured image is not matched with any of the database image and predicted result tells us the same. False positive: When any of the preloaded database images is not matched with captured image of product but then also system predicts image found. False negative: When captured image is not matched with any of the database image and predicted result tells us the same. These values occur when actual class contradicts with predicted class. False positive: When any of the preloaded database images is not matched with captured image of product but then also system predicts image found. False negative: When matching image is present in database but not predicted by system.

By using these four parameters, Precision, Recall, F-Score and Accuracy of the system is calculated.
Precision: It is the ratio of correctly matched images to the sum of correctly matched images and false positives.

\[ P = \frac{TP}{TP + FP} \]

Recall (Sensitivity): It is the ratio of correctly matched images to the sum of correctly matched images and false negatives.

\[ R = \frac{TP}{TP + FN} \]

F-Score: It is the weighted average of Precision and Recall.

\[ \text{F-Score} = \frac{2 \times P \times R}{P + R} \]

Accuracy: It is intuitive performance measure which is the ratio of correctly predicted observations to the total observations.

\[ \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \]

Precision, Recall, F-Score for captured images of different products taken from different positions or orientations are used to determine the accuracy of proposed system and to analyze the success and limitations of system. The performance of the system is shown in figure.

Table 1. Performance Percentage

<table>
<thead>
<tr>
<th>Images</th>
<th>Precision Rate (%)</th>
<th>Recall Rate (%)</th>
<th>F-Score (%)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

The average precision of the system is 90 and the average recall rate of system is 90. The system is tested on various products result of testing shows that the proposed implemented system performs better than existing systems.

7. Conclusion

We effectively separated the objects of interest from complex background followed by image preprocessing. Image threshold algorithm is applied to capture image to choose an intensity value as a threshold level and the values below threshold level become 0 (Black) and the values above threshold level become 1 (white). Surf algorithm is used to match capture image with preloaded image of database. At successful match, it provides audio output to user regarding product details.

8. Acknowledgements

We have taken lots of efforts in this project, although it wouldn't possible with grace of God. I would like to extend my sincere thanks to all who assist us to complete our project. I would like to thank our Principal Prof. Vijaya Patil and our HOD Prof. Vilas Mankar for encouraging us. We also like to thank our project guide Prof. Darshana Shimpi for her help & Guidance throughout the project. At last, my thanks and appreciation also goes to my group members in developing the project.

9. References

[3] Solutions for the Blind or Visually Impaired and for People with Learning and Reading Difficulties http://www.sightandsound.co.uk/downloads/docs/Sight_And_Sound_Brochure.pdf