Journal of Information Systems Engineering and Management
2025, 10(31s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Web Application Development for Product Label Reading using Machine Learning Techniques

Dr. B. Anuradha[1] , M.Kasipandian[2], Dr. M. Rajesh Babu[3], Dr. S. Uma[4], Dr. S. Palanisamy5, B. Venkatraj[6]
[1]Associate Professor, Department of Computer Science & Design, SNS College of Engineering, Coimbatore, India
Email: pskanu80@gmail.com
[2]Assistant Professor, Department of MCA, Bharath Niketan Engineering College, Aundipatti, Theni(Dt)., TamilNadu, India
Email: kasipandi1985@gmail.com
[3]Professor & Associate Dean, School of Computing, Rathinam Technical Campus, Coimbatore, India
Email: drmrajeshbabu@gmail.com
[4]Professor, PG In charge, Department of CSE, Hindusthan College of Engineering and Technology, Coimbatore, India
Email: druma.cse@hicet.ac.in
[5]Assistant Professor [SG], Department of CSE, Government College of Engineering, Erode, India
Email: spsyaso@gmail.com
[6]PG Student, Department of CSE, Hindusthan College of Engineering and Technology, Coimbatore, India
Email: venkatbaskar32@gmail.com
Corresponding author email id: pskanu80@gmail.com

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

This paper analyse and develop the web application using camera based assistive text reading of product labels from objects to help visually challenged people. Camera serves as main source of input. To detect the objects, the user will move the object in front of camera and this moving object will be detected by Background Subtraction (BGS) Method. Text region will be automatically localized as Region of Interest (ROI). Text is extracted from ROI by combining both rule based and learning based method. A novel rule based text localization algorithm is employed by identifying geometric features like pixel value, color intensity, character size etc. and also features like Gradient magnitude, gradient width and stroke width are learned using SVM classifier and a model is built to differentiate text and non-text region.

Keywords: *Web application, Moving object detection, Background subtraction method, Text localization, SVM classifier*

1. INTRODUCTION

Now-a-days web engineering [2] placed an major role in different web application developments because it provides the collection of design, tools, techniques, modules for analyzing the web application. The developed web application system [3] helps in different people for making their business effective, research process and so on. Even in developed countries like United States (US), 60,393 students are visually challenged. These rates have increased rapidly over years. Approximately 90% of visually impaired people lead their lives in developing countries [4]. Recent research works are done in developing braille character recognition, digital camera and portable cameras which assist blind people in their daily life. Presently visually challenged people are marking their achievement in almost all fields with the help of digital improvement.

Reading is an essential component in today's competitive world. Everything around us are in the form of reports, receipts, bank statements, product packages, restaurant menus etc. For instance on considering the Braille Character Recognition (BCR) [5], it recognizes the braille characters and reads it out to the user and even there are portable bar code readers to identify various information about the product. To overcome the disadvantage of developing the effective web application that locating the region of interest (ROI) automatically, we propose a motion detection algorithm to locate ROI. In order to avoid that, a person will move the object in front of the camera which will be detected by motion detection algorithm. It is a quite challenging task to automatically localize Region of Interest (ROI) [6] from complex background as the object is moving. Hence the moving object will be tracked throughout and in the last frame the moving object will be identified, provided that the object has to be still at least for a couple of seconds towards the end. There are two methods to extract text from the object - rule based and learning based techniques. Rule Based Techniques uses only the geometric properties of the images like pixel value, Color intensity, Character size, aspect ratio, Color uniformity across text etc. to detect [7]. The extended the learning algorithm[8] by training additional feature like Stroke components. Here a novel algorithm based on stroke components and descriptive Gabor filters, is used to detect text regions in natural scene images. A model is built based on many features like Gradient magnitude, gradient width and stroke width which are enough to differentiate both text and non-text region. Thus, this paper introduces the novel rule-based text localization algorithm for creating the web application for Product Label Reading From Visually Challenged People with effective manner.

Then the rest of the section is organized as follows, section 2 deals the related works about the product label reading process, section 3 examines the system structure and working process of novel rule based text localization algorithm, section 4 evaluates the efficiency of system and concludes in section 5.

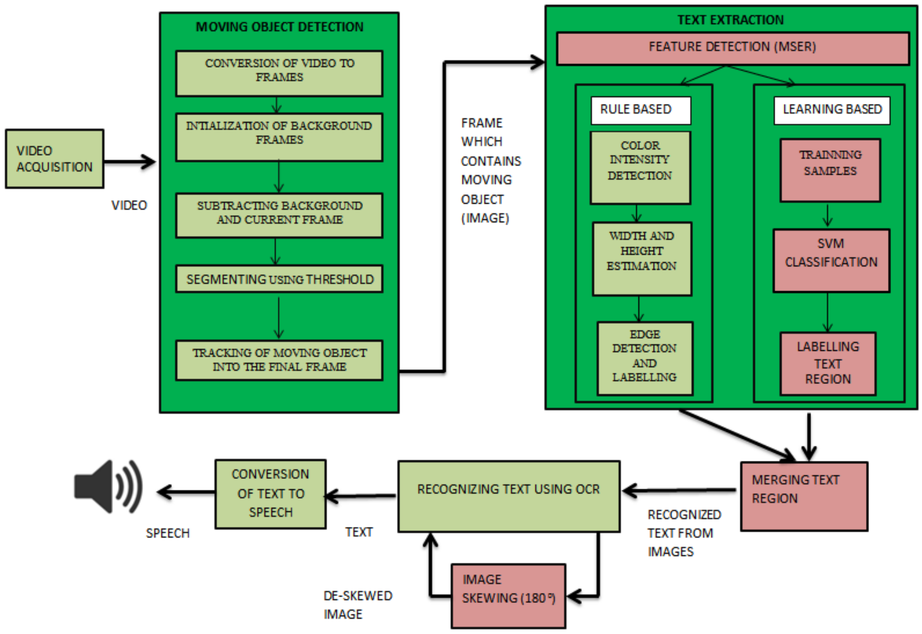
2. RELATED WORK

Arumalla Raja et al. [10] present a camera-based label reader to help blind persons to read names of labels on the products. This uses opencv library to separate the labels from images. The received label is then converted to text by using tesseract library. Aseema et al.[11] described about the different methods for detecting moving object like optical flow method , segmentation method, temporal differencing method and background subtraction method for moving object detection. Background subtraction method in general will fix the background frame and then find the foreground object. Kalaiselvi et al. [12] describes a method for reading text for visually impaired people. This method uses Stationary wavelet transform to identify the text regions from the image.

Priyanka et al. [13] explores various techniques to remove text out of video. This paper performs a extensive survey of various techniques available for text detection and removal from video and images automatically to create visually plausible output. Techniques are proposed based on morphological operators, wavelet transform artificial neural network, skeletonization operation, edge detection algorithm, histogram technique etc. ChucaiYi et al.[14] proposed a mosaic model to unwrap the mentioned text label on the cylinder object surface and reconstruct the whole label for recognizing text information. It uses Adaboost classifier to detect text in each pixel in the cylinder surface will be projected to its tangent plane. This paper employs Scale-invariant feature transform (SIFT) to calculate the matching points between two overlapped images of different portion of the object label. It implies RANSAC algorithm to estimate the transformation matrix in order to stitching all the flattened images together.

Adam et al. [15] proposes a different method to read out text from photographs. Two key components of this system is text detection from images and character recognition. There are many methods available already for doing both the steps. This paper explains about the feature learning method. The machine learning methods are used for learning the features from unlabelled data and will construct classifiers for both detection and recognition for better precision and accuracy. Chucai Yi et al. [16] extended the learning algorithm by training additional feature like Stroke components. Here a novel algorithm based on stroke components and descriptive Gabor filters, is used to detect text regions in natural scene images. Due to the advantage of this method, combination of both rule and learning approach for developing the effective web application is proposed to readthe label for physically challenged people. Then the detail explanation of the system is explained as follows.

3. SYSTEM ARCHITECTURE



Traditional Rule based Approach

Combine Rule and Learning based Approach

Figure 1 Overall Architecture

The above figure 1 depicted that the web application development processing steps for reading product label from the physically challenged people. The architecture includes different processing steps background subtraction, region of interest and rule based reading process. Each and every step is explained as follows.

1. Video Acquisition:

The real time video is captured with the help of a camera. The video should last for at least 5 seconds and in order to extract text from the object it should be still at least in the last frame, hence the user keeps the object still for 2 seconds at the end. The moving object [17] is tracked and identified by drawing the bounding box in the last frame.

2. Moving Object Detection:

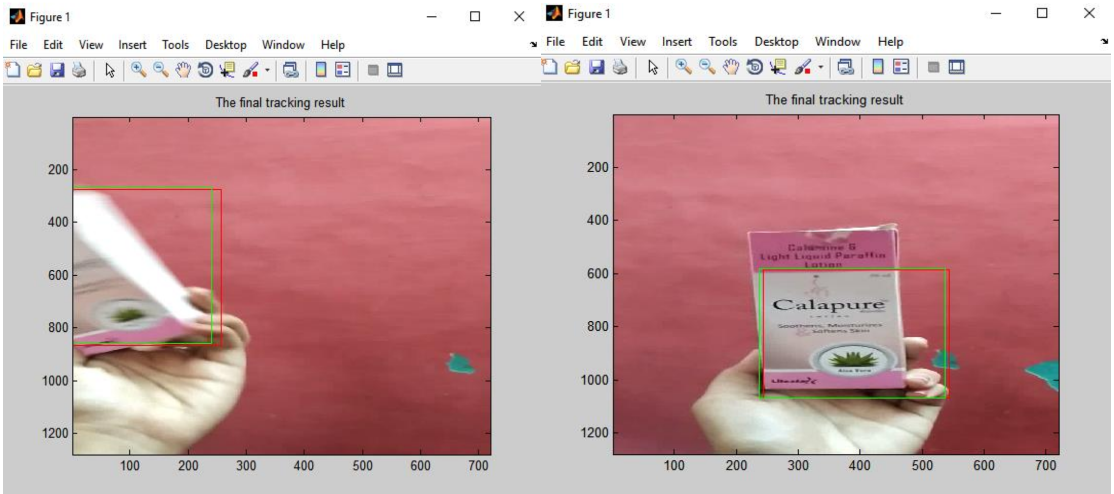
In order to eliminate the background from the object, background subtraction algorithm [18] is applied. Initially video sequence is converted to frames and each frame is processed separately. Background subtraction is a method which involves computational vision process of extracting foreground objects in a particular scene. A foreground object can be described as an object of attention which helps in reducing the amount of data to be processed as well as to provide important information to the task under consideration. Background is estimated using “Mean filtering” i.e., the background frame is estimated using the mean pixel values of n frames.

$$B(x,y,t)= \frac{1}{n} \sum_{i=1}^{n-1} I(x,y,t-i) \quad (1)$$

Once the background frame is fixed, the current frame pixel value is subtracted from the initialized background frame. Difference value above threshold is estimated to be the moving pixel. The key paramater in thresholding process is the choice of threshold value. Threshold value which is too high neogiates moving object and threshold value which is too low adds many object [18] which is not moving, hence it should be a nominal value which is around 50

$$\text{Difference} = I(x,y,t) - B(x,y,t) > \text{Threshold} \quad (2)$$

Moving object is completely tracked and the moving pixels are finally identified in the final frame and a bounding is drawn across the moving object. The frame which has the identified moving object is send to the next phase.



(a) Moving Object in one place (b) Moving Object in other place

Figure 2 Moving Object Detection

3. Text Extraction:

The system builds an automatic text localization and extraction system which is able to accept different types of still images (or video frames) possibly with a complex background. Text extraction combines both rule-based and learning-based techniques to localize text region.

3.1 Feature detection:

Maximally Stable Extremal Region (MSER) [20] is a method for blob detection in images. The MSER algorithm extracts from an image a number of co-variant regions, called MSERs: an MSER is a stable connected component of some gray-level sets of the image. These white spots will eventually merge, until the whole image is white. The set of all connected components in the sequence is the set of all extremal regions. Elliptical frames are attached to the MSERs by fitting ellipses to the regions. Those regions descriptors are kept as features.The word extremal refers to the property that all pixels inside the MSER have either higher (bright extremal regions) or lower (dark extremal regions) intensity than all the pixels on its outer boundary. The feature detector finds both text and non-text region. Further text is extracted from these detected MSER region.

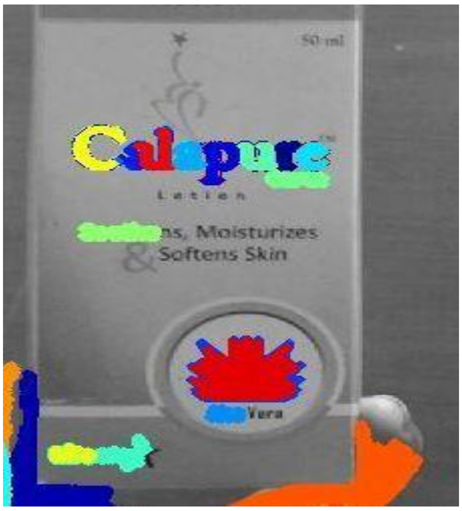


Figure 3 Detected feature using MSER

3.2 Rule-Based Text Extraction:

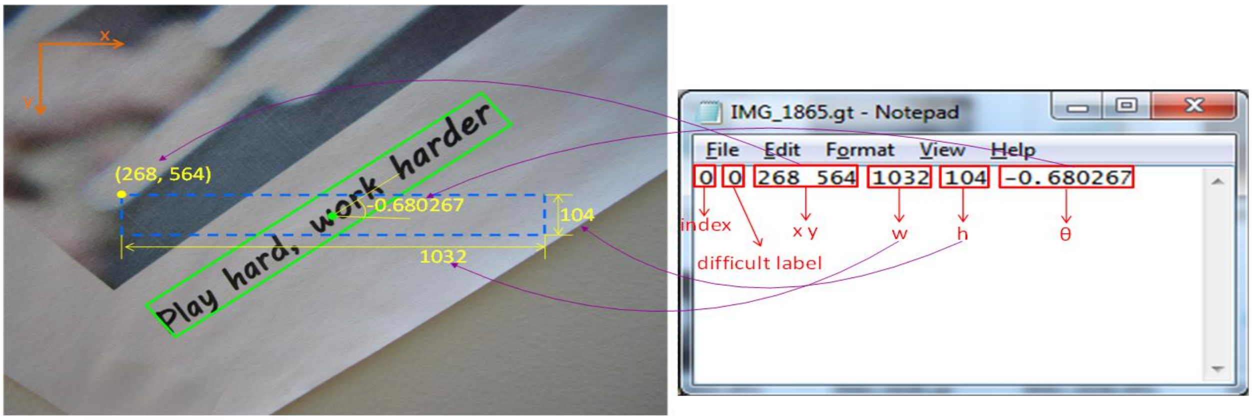
Image partition is first performed to group together pixels that belong to the same text character, obtaining a binary map of candidate character components for extracting text information from a complex background, Rule-based algorithms extract text information from predefined text layouts such as character size, aspect ratio, edge density, character structure, color uniformity of text string [21], etc. for applying pixel-level image processing to. In our system the properties which are taken into consideration are Color intensity, Width and Height and Edge. Text characters will have smaller width and height than its background and the color uniformity will be same to certain extend. Edge is the distinct characteristic which can be used to find possible text areas. Text is mainly composed of the strokes in horizontal, vertical, up-right, up-left direction, so it can be considered that the region with higher edge strength in these directions is the text region. We use the edge detector to get the edge maps in four directions. Canny edge detector is employed to obtain edge image.

Figure 4 Rule Based Result

3.3 Learning-Based Text Extraction:

Even the rule-based text Extraction can detect some non-text region in addition. To improve the performance of the system we integrate the results with learning-based technique which uses machine learning algorithm (Support Vector Machine) to classify non-text region from text region. The proposed method poses text detection as a texture classification problem where problem-specific knowledge is available prior to classification (Supervised classification). Features which are trained to build the model are Gradient magnitude, Gradient orientation, Gradient in both horizontal and vertical. Along with this gradient feature edge orientation, color histogram, Stoke width transform, wavelet transform is also learned.

(a) Sample Dataset Images



IMG_1865.JPG IMG_1865.gt

(b) Captured Images (c) Notepad Image

Figure 5 MSRA dataset which is used to train the SVM classifier

Support Vector Machine:

SVMs have been recently introduced as a method for pattern classification and nonlinear regression. Given a set of labeled training examples $(x_i, y_i) \in R^N \times \{\pm 1\}$, i=1,...l ,an SVM constructs a linear classifier in a feature space F, which is nonlinearly related to the input space via a map $\phi : R^N \rightarrow F$.

4. Image Skewing:

The labels on the product need not be always straight. In order to find product labels, the labels should be made straight. The labels may be slanting or upside down. Upside down image is identified only after making the image to undergo text recognition using OCR. If the OCR conversion ratio is less than 30% then it is rotated to 180^o and again text is identified using OCR (considered only image which is upside down).

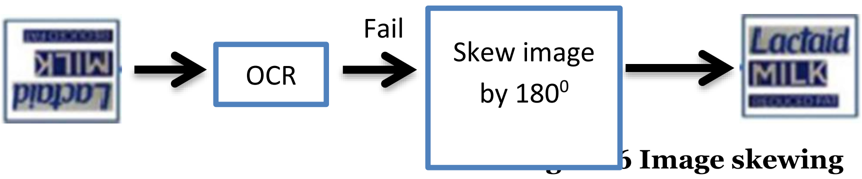


Figure 5 Image skewing

4. EVALUATION METRICS

We evaluate the proposed algorithm ICDAR datasets. ICDAR 2003 Robust Reading Datasets contains 507 images in total. The image sizes range from 640×480 to 1600×1200. There are 2258 ground truth text regions in total. We evaluate the performance of our algorithm by comparing the detected text regions with the ground truth text regions. We define “**precision**” (p) as the ratio of total match score to the total number of localized regions and “**recall**” as the ratio between the total match score and the total number of ground truth regions. Here area means the number of pixels in the image region. “**f-measure**” is defined as the combination of precision and the recall by the harmonic mean.

$$Precision = \frac{total\ match\ score}{total\ no.\ of\ localized\ region} \quad (3)$$

$$Recall = \frac{total\ match\ score}{total\ no.\ of\ ground\ truth\ region} \quad (4)$$

$$F-Measure = 1 / \left(\frac{\alpha}{p} + \frac{(1-\alpha)}{r} \right) \quad (5)$$

where ‘α’ represents relative weight between the above two metrics (α = 0.5)

The ICDAR dataset has been evaluated with the help of software called DetEval. DetEval is software for the evaluation of object detection algorithms. It reads XML files containing the detection results as well as the ground truth information and writes the evaluation output in XML or Latex format. ICDAR evaluation algorithm has been adopted to measure the values.

Table 1 Comparison of accuracy of existing and proposed method

SNO	METHOD	ACCURACY
1	Rule based	68.4
2	Learning based	72.3
3	Combining both rule and learning based results	83.6

The table 1, clearly indicates that the accuracy (83.6%) of the developed product label-based reading web application development process. The combined rule and learning based approach attains the highest accuracy when compared to the normal rule (68.4%) and learning based approach (72.3%). Then the graphical representation of accuracy metric is shown in figure 12.

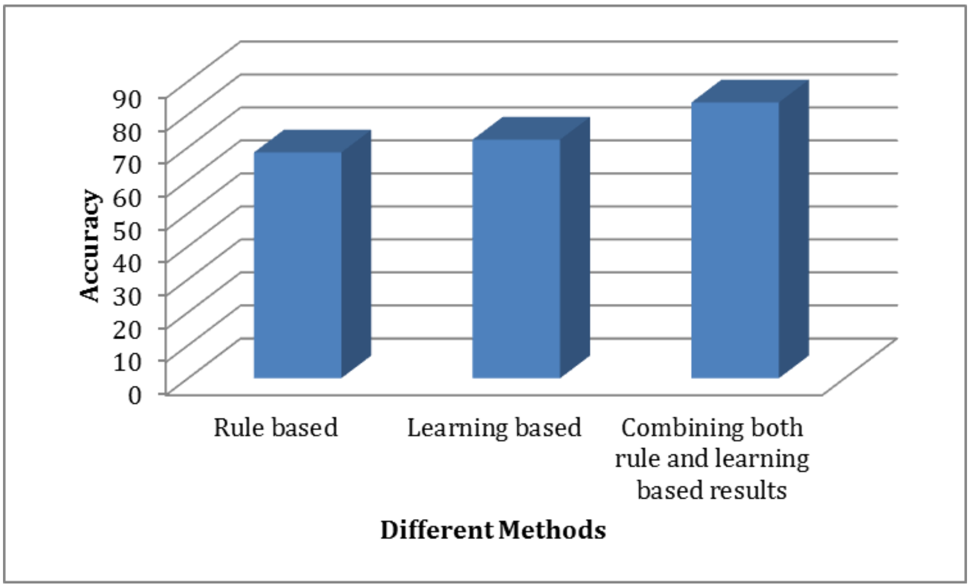


Figure 7 Comparison of Accuracy of existing and proposed method

Figure 7 compares the accuracy of the existing and proposed method. It can be seen that accuracy increases significantly when rule based and learning based approach is combined while detecting the product label.

Table 2 Comparison of precision of existing and proposed method

S.No.	METHOD	PRECISION
1	Rule based	73.3
2	Learning based	82.60
3	Combining both rule and learning based results	89.33

The table 2, clearly indicates that the precision of the developed product label based reading web application development process. The combined rule and learning based approach attains the highest precision (89.33%) when compared to the normal rule (73.3%) and learning based approach (82.60%). Then the graphical representation of accuracy metric is shown in figure 8.

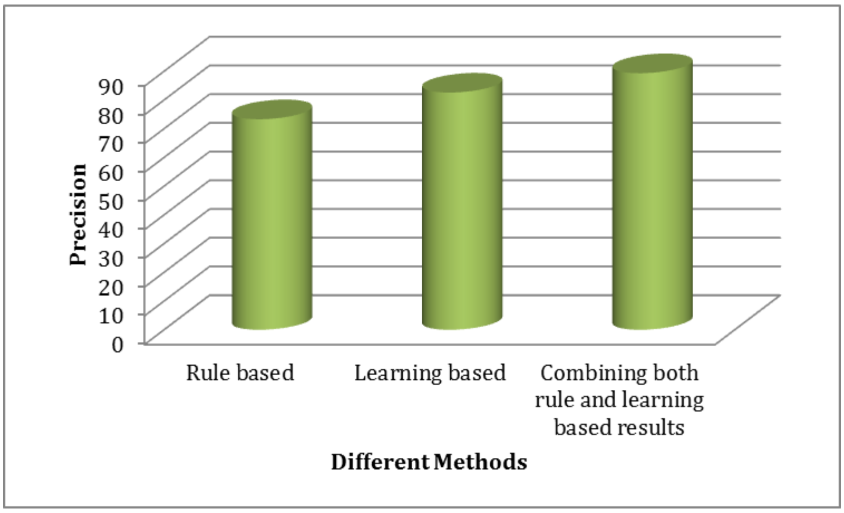


Figure 8 Comparison of Precision of existing and proposed method

Figure 13 compares Precision of the existing and proposed method. It can be seen that precision increases significantly when rule based and learning based approach is combined while detecting the product label.

Table 3 Comparison of recall of existing and proposed method

SNO	METHOD	RECALL
1	Rule based	56.89
2	Learning based	65.46
3	Combining both rule and learning based results	86.89

The table 3, clearly indicates that the recall of the developed product label-based reading web application development process. The combined rule and learning based approach attains the highest recall (86.89%) when compared to the normal rule (56.89%) and learning based approach (65.46%). Then the graphical representation of accuracy metric is shown in figure 13.

5. CONCLUSION

This paper describes a prototype system to read printed text on objects for assisting visually challenged people. In order to solve the common focusing problem for blind users, we have proposed a motion-based method to detect the object of interest, while the blind user simply moves the object for some seconds. This method distinguishes the object of interest from background or other objects which are in the camera view. Text is extracted from the moving object by using two methods one is rule-based and learning-based techniques. Extracted text is merged into words and de-skewed in order to remove the alignment problems. The text is digitalized using OCR (Optical Character Recognition) and it is given to the user in form of speech. Our future work can be extended to localization algorithm which can efficiently process text strings with fewer characters and focuses on designing robust block patterns for text feature extraction. We will also extend our algorithm to handle non-horizontal text strings. Furthermore, we will address the significant human interface issues associated with reading text by blind users.

REFERENCES

[1] Charlyn Nayve Villavicencio, Julio Jerison Macrohon, Xavier Alphonse Inbaraj, Jyh-Horng Jeng and Jer-Guang Hsieh (2024), "Development of a Machine Learning Based Web Application for Early Diagnosis of COVID-19 Based on Symptoms", Diagnostics. Vol.12, Issue 4.

[2] Roger S Pressman (2001), "What a Tangled Web we Weave", IEEE Software, Vol. 18, No.1, pp 18-21.

[3] Gerti Kappel, Birgit Proll, Seiegfried, and Werner Retschitzegger (2003), "An Introduction to Web Engineering," in Web Engineering, Gerti Kappel, et al. (eds.) John Wiley and Sons, Heidelberg, Germany.

[4] Chen, Datong, Jean-Marc Odobez, and HerveBourlard (2004), 'Text detection and recognition in images and video frames', Elsevier on Pattern recognition, pp. 595-608.

[5] Jung, Keechul, KwangIn Kim, and Anil K. Jain (2004) 'Text information extraction in images and video: a survey' Pattern recognition, Vol. 37 No. 5, pp. 977-997.

[6] Shaikh, SoharabHossain, Khalid Saeed, and NabenduChaki (2014), "Moving Object Detection Using Background Subtraction" Springer International Publishing.

[7] Stauffer, Chris, and W. Eric L. Grimson (1999), "Adaptive background mixture models for real-time tracking", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2.

[8] Tian, Ying-Li, Max Lu, and ArunHampapur (2005), "Robust and efficient foreground analysis for real-time video surveillance", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1.

[9] Ye, Qixiang (2005), "Fast and robust text detection in images and video frames", Image and Vision Computing, pp. 565-576.

[10] Raja, Arumalla, and Karumaru Dinesh Reddy (2015), "*Compact Camera Based Assistive Text Product Label Reading and Voice Output for Visually Challenged People*"', International Journal of Innovative Technologies,Vol. 3 No.4, pp. 0586-0589

[11] Aseema Mohanty and Sanjivani Shantaiya (2015), '*A Survey on Moving Object Detection using Background Subtraction Methods in Video*', International Journal of Computer Applications Proceedings on National Conference on Knowledge, Innovation in Technology and Engineering, No. 2, pp. 5-10.

[12] Kalaiselvi U.,Anish Kumar J. (2014) '*Camera based Assistive Text Reading System using Gradient and Stroke Orientation for Blind Person*', International Journal of Latest Trends in Engineering and Technology ,Vol. 4 No. 1, pp. 325-330.

[13] Priyankadeelipwagh and PatilD. R. (2014) '*Text Detection, Extraction and Removal: A Survey*', International Journal of Computer Applications Proceedings on National Conference on Emerging Trends in Computer Technology, No. 1 pp. 19-23.

[14] Yi, Chucai, and YingliTian (2011) '*Text detection in natural scene images by stroke gabor words*', IEEE International Conference on Document Analysis and Recognition, pp. 18-21.

[15] Coates, Adam (2011) '*Text detection and character recognition in scene images with unsupervised feature learning*', IEEE International Conference on Document Analysis and Recognition, pp. 18-21.

[16] Yi, Chucai, and YingLiTian (2011) '*Text string detection from natural scenes by structure-based partition and grouping*', IEEE Transactions on Image Processing, pp. 2594-2605.

[17] Gllavata, Julinda, Ralph Ewerth, and Bernd Freisleben (2003) '*A robust algorithm for text detection in images*', IEEE International Symposium on Image and Signal Processing and Analysis, Vol. 2.

[18] Chen, Huizhong (2011) '*Robust text detection in natural images with edge-enhanced maximally stable extremal regions*' IEEE International Conference on Image Processing, pp. 11-14.

[19] Chen, Datong, Jean-Marc Odobez, and Jean-Philippe Thiran (2004) '*A localization/verification scheme for finding text in images and video frames based on contrast independent features and machine learning methods*', Signal Processing: Image Communication, pp. 205-217.

[20] Kim, Kwang In, Keechul Jung, and Jin Hyung Kim (2003) '*Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm*', IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25 No. 12, pp. 1631-1639.

[21] Yi, Chucai, YingliTian, and Aries Arditi(2014) '*Portable camera-based assistive text and product label reading from hand-held objects for blind persons',* IEEE/ASME Transactions on Mechatronics, Vol. 19 No. 3, pp. 808-817.