Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 176 (2020) 602–611

Procedia Computer Science

www.elsevier.com/locate/procedia

24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

A New Architecture based AR for Detection and Recognition of Objects and Text to Enhance Navigation of Visually Impaired People

Imene OUALI[a*], Mohamed Saifeddine HADJ SASSI[b], Mohamed BEN HALIMA[a], Ali WALI[a]

[a]REGIM-Lab: Research Groups in Intelligent Machines, University of Sfax, National Engineering School of Sfax,BP 1173, Sfax, 3038 Tunisia
[b]Digital Research Center of Sfax, LT2S Laboratory, Tunisia

Abstract

Navigation in unfamiliar places is a big challenge for partially sighted and visually impaired people. Improving visual information on the location and content of objects such as drugs can help navigation in unfamiliar environments. There are several existing navigation solutions capable of helping these people. However, navigation solutions are rarely adopted and implemented in reality. In order to optimize the perception of digital information of objects from sensors and to naturally interact with the pervasive computing landscape, Augmented Reality (AR) equipment has to be seamlessly integrated into the user's environment. For this purpose, we develop an architecture of text and objects recognition based on AR in order to assist partially sighted and visually impaired people. Such architecture makes navigation easier in the environment by helping users to find drugs and to verify the number of pills using speech. The proposed architecture uses context-aware mobile computing and shows great potential to integrate AR for object recognition through AR engines.

Keywords:
Augmented reality, Vuforia, Usability, Visually impaired, Objects, Text, Detection, Recognition.

1. Introduction

Real-time detection and recognition of objects and text in natural scenes has aroused more and more interest of researchers in recent years due to its necessity and importance in our life. In fact, human beings should have good vision to recognize objects and words in their environment. However, finding and identifying drugs, pills, or medications by color, shape, or imprint is a difficult task for blind and low vision people. The visual impairments

E-mail address: Imene3ouali@gmail.com

10.1016/j.procs.2020.08.062

prevent them from understanding and exposing themselves to the environment. The number of visually impaired and low vision people is very large. These people find difficulties to satisfy their daily needs their daily needs relying on their eyes and visions. Augmented Reality (AR) provides an ideal interface to smart devices in order to show information, in which 3D virtual objects are integrated into a 3D real environment in real time. The new technology of AR offers many advantages for digital architectural design and construction fields. While AR is already being considered as new design approach for architecture, indoor AR is another practical application that can take advantage of new wearable computer equipment (Head-mounted display (HMD), position and orientation sensors, and mobile computing) to superimpose virtual graphics of the environment. In this context, we develop an architecture that helps visually impaired people to find objects (drugs) and know exactly the number of pills through speech. Our system starts by selecting the name of the object (drug in our case) in the user's smartphone using his voice. The system informs the user of the availability of the object in his environment and help him to find his target using the camera. Our system allows to rewrite the pills names bigger and clearer. Furthermore, our system calculates the number of pills remaining in the drug tablet. Finally, it displays the number found. This solution allows visually impaired people to do his daily needs by the detection of text and object at the same time, checking the availability of the drug and specifying the location and the number of pills left. The source date of our system is divided into two parts. These two parts can be found in the AR engine database. The object part, it contains all the image of drugs of the user. The text part contains all the names of the drugs which is stored in the engine database. This system can detect any style of letters on any background, no matter the color, size, style through the proposed engine. This paper is organized as follows:

- Several works in the related to text and object detection and recognition recently published by the research community have been introduced (Table 1). The problem of detection and recognition has been studied for many years. Scientific researchers indicates that small fonts and complex background and adjustments like color contrasts are big problems for detection and recognition tasks.

- We provide an architecture that deals with detection and recognition problems using AR. Furthermore, we validate and evaluate our work with partially sighted people.

Section 2 defines the related works. Open research issues are described in section 3. We present an overview of our architecture in section 4. Moreover, we explain how to deal with limitation through appropriate existing technologies and AR engine. Sections 5 and 6 respectively define the evaluation through usability and subjective test, and the discussion of our architecture. Finally, concluding remarks are presented in section 7.

2. Related work

In this section, we present several related works with an emphasis on the development of new correspondence methods and systems. Table 1 shows a comparative study of different methods and systems for detection and recognition of object and text. Navigation is important for people's life cycle and it mainly depend on human vision. Besides, navigation is essential and necessary to carry out different daily tasks easily and independently. Vision less People need help with orientation and navigation. Technologies that may help visually impaired and partially sighted people to navigate can have a significant impact on the quality of their lives. Although existing aids such as GPS technology and canes can help the visually impaired in outdoor orientation and avoid obstacles, their applicability to orientation and indoor signage is limited. To solve this problem for the visually impaired and to extend navigation aids, several researchers have proposed methods and developed tools that work with mobile technologies, such as smart glasses or smartphones, to provide position information and indications of orientation indoors as shown in [1] and [2]. In particular, AR can provide a specific information without improving the user's natural vision. Furthermore, other recent work has explored to help users with a specific task (searching for text or objects), and has found that these tools allow participants who have reduced vision and the visually impaired to considerably reduce their research time. Several works which tries and seek to find solutions for detection and recognition from video such as [3, 4, 5] and other allow the detection and recognition of text from images such as [6, 7, 8, 9, 10, 11, 12, 13]. Other approaches for object detection and recognition such as [14], [15], [16], [17], [18], [19] and [20]were proposed. However, these techniques suffer from lack of textual information during detection or the collection and storage of information for the recognition of these objects. Another application study that reported text detection and recognition for smartphones to help read text in objects such as [21], [22], [23] and [24]. The main advantages and drawbacks are described in Table 1.

Table 1: Comparison between Detection and Recognition (D/R) of objects and text systems

	Ref	Year	Advantages	Limitations
Text D/R	[14]	2019	Text detection using double SWT will give good accuracy.	Text can not be detected in images with a dark background and a light background or that contain side-by-side clustered text.
	[15]	2019	This system offers a recognition network with overall attention.	It does not process text images with non-alphanumeric characters and images less than 3 characters.
	[16]	2019	End-to-end method for multi-oriented scene text detection.	It has good performance but not high efficiency.
	[17]	2019	MORAN can read both regular and irregular scene text.	Result with some error for recognition of difficult image text.
	[18]	2019	Improves OCR accuracy by adding several prepossessing techniques: improving resolution, locating documents, locating text.	Accuracy and speed are bad when the GPU is not available.
	[19]	2019	This technique effectively improves the distortion effect caused in the synthetic text image due to the loss of information.	This technique works well only when applied to foggy and dense natural text images (indoor / outdoor).
	[20]	2019	Decrease in their overall reading time for text passages.	This method is only theoretical and is not practically tested.
Objects D/R	[21]	2020	This system could be reused, adapting new knowledge and free three-dimensional (3D) models on the Web.	This dedicated application only for medical students.
	[22]	2019	Attaching markers to the back of pedestrians allows them to be recognized even when they look the other way.	This method allows a maximum of three pedestrians to be recognized.
	[23]	2019	This augmented reality application demonstrates easy handling and strong interactivity with the user.	The setting of PID controllers in industrial processes is an issue.
	[24]	2018	This method combines real and synthetic data for learning the segmentation of semantic instances and object detection models.	Synthetic objects can only be placed on real images and therefore cannot be partially occluded by real objects during the data generation model.

2.1. Objects detection and recognition

Several methods of object detection using AR such as Alhaija, et al [24] present a method to detect and recognize of objects using AR. In addition, Romero, et al. [23] define a method to recognize the equipment via a Smartphone and to detect the characteristic points of objects and image recognition, 3D modeling via CAD software and integration into a multi platform, incorporation of animations and mathematical modeling of industrial processes which allow the development of closed loop control algorithms. Furthermore, Kim, et al. [22] has proposed a pedestrian detection method using marker recognition. Several pedestrians are detected and then followed according to tags fixed at the rear. Once the marker is recognized, the unique character associated with this marker is displayed as a 3D object. Moreover, Reyes-Ruiz, et al. [21] describe a simple and AR-based method that create appropriate learning environments that allow students to feel motivated, encouraged and eager to continue learning; With the support of ARs, systems can be generated to facilitate the learning of abstract or hard to perceive knowledge; Medicine manages models whose appearance and shape help and reinforce learning, which can be represented with three-dimensional (3D) entities;

The interaction with AR and multimedia materials which are added to physical reality allow to gradually stimulate the sensory senses of the human being, and in particular of the pupils, thus allowing the pupil to learn in an audition, visual and kinesthetic.

2.2. Text detection and recognition

Several systems of text detection using AR were proposed such as [19] in this system, Ansari et al. improves accessibility to dyslexic disorders thanks to AR using a smartphone camera. Several methods of text detection using deep learning such as [17], [16], [15] and [18] were proposed. Luo,et al. [17] define a system MORAN: A multi-object rectified attention network for the general recognition of scene text. The multi-object rectification network is designed to rectify images containing irregular text. It decreases the difficulty of recognition and allows the attention-based sequence recognition network to read irregular text more easily. Cheng, et al. [16] show an end-to-end method for multi-oriented scene text detection by introducing position-sensitive segmentation in the direct regression method. The proposed method involves three prediction tasks: the first classification task performs a sub sampled segmentation to locate the text regions, the second regression task regresses the text areas and the third classification task performs position-sensitive segmentation text to get a more refined text location. Gao, et al. [15] describe a text recognition system in natural images offers a recognition network with overall attention. Akopyan, et al. [18] define a system that allows recognition of text on images from social networks. Ansari, et al. [19] present a methodology for scrambling a single image followed by a new method for generating text proposals based on characters. Shahana, et al. [14] propose a method to detect a text in Malayalam language from images of natural scenes. Gupta, et al. [20] describe a system based on AR using a smartphone camera to overcome these issues by permitting the user to adjust the contrast of the text background and to personalize the text in real life.

3. Open research issue

Several limitations in text detection and recognition methods such as in [14] text is undetectable in images with a dark background and a light background or containing grouped text side by side. In addition, for [15] does not process text images with non-alphanumeric characters and images less than 3 characters. Furthermore, for [16] this method has good performance but the efficiency is weak. Moreover, in [17] results are with high error level in recognizing difficult image text. Besides, in [18] accuracy and speed are poor when the GPU is absent. In addition, this technique [19] only works well when applied to foggy and dense natural text images (indoor / outdoor). Furthermore, this method [20] is only theoretical and is not practically tested. According to all these limits we propose a system that overcomes all these problems and other problems text detection and recognition. In addition, for object detection and recognition limits in methods such as in [24] synthetic objects can only be placed on real images and therefore cannot be partially masked by real objects when of the data generation model. In [23] the installation of PID controllers in industrial processes is a problem. Furthermore, this method [22] allows to recognize a maximum of three pedestrians. This application [21] is dedicated only to medical students. Based on the reviewed related works, we define a novel approach for people who have limited vision in order to detect objects such as drugs and pills. Thus, it is a great opportunity for those kind of suffering people to help them with orientation to provide improved visual information about the objects, the location and content of the objects.

4. Proposed architecture

In this section, we present a new architecture for detecting and recognizing text and objects. Based on our previous work [3], we detect the drug by using smart device features such as a camera (Figure 1). Besides, the collected data will be extracted and recognized through a tool based on AR engine. Then, the application detects the object (the pills or the drug) using a camera. Then, it checks the existence of the object needed through the name of pills (text detection). Finally, if the needed drug is found, our system computes the number of the remaining pills (Figure 3). This tool basically establishes connections between images of pills in the stored object data such as the picture of the pills and drugs (Figure 4).

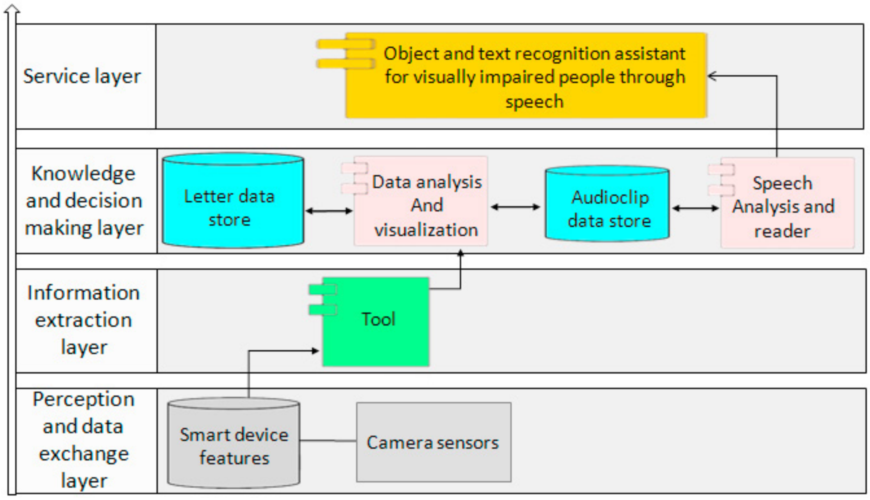

Fig. 1: Layers of our architectures.

Within existing AR technologies proposed in the related work section. First, we need to collect data from the camera sensor of smart devices. As a result, we define a program for objects and texts detection from a data engine (Figure 4). Android system offer high efficiency in object and text detection thanks to its availability. Apart from being a platform for mobile phones, Android is a complete platform for smart devices such as smart watch, smart glasses, etc. Vuforia AR software development kit is adopted in our study for accessibility, data storing (pills name data) and rapidity (data recognition in a few seconds) reasons. Vuforia detects and tracks text elements by scanning the camera display called the region of interest (ROI). This technology uses two different ROIs, one for detection and the other for monitoring. In terms of performance, detecting too many words in a track will result in longer detection times and a slow system response. When text detection is activated, Vuforia tries to match the words found in the word lists. Word lists can be loaded from files and can be upgraded to store additional words in necessity cases. Word lists can also be filtered using a filter list to exclude certain words from detection (using a blacklist filter), or only by allowing certain words to be detected (using a white list filter). It is a collection of basic words that Vuforia can use as references to appropriate text elements that appear in the camera's field of view. The pill image lists and word lists are extracted from binary files encoded in the Vuforia datastore which contains our data set of pill images and letters. Vuforia uses computer vision technology to identify and track planar images (target images) and simple 3D objects such as boxes, in real time. Image recording features allow developers to adjust the position and orientation of virtual objects, such as 3D models and other media, relative to real-world images. The virtual object then follows the position and orientation of the image in real time so that the user's perspective on the object matches that of the target image. Thus, the virtual objects can be displayed in the image in the real world. Vuforia SDK supports different types of targets. There are many mobile applications to detect, recognize or read text from natural images, but they suffer from low efficiency and correctness because of flaws. So, the probability of obtaining wrong information about the meaning of the text is high. AR technology offers a real time object detection. It performs detection, recognition and reading in real time of objects and texts as well using a good engine. Vuforia can detect the words included in the list of predefined words. Vuforia allows also to specify a list of special words and filters that prevent word recognition. In our system we have added a database for the language Latin. Text recognition is useful for applications that need to recognize individual and series of words. Text recognition can be used as a standalone feature or in combination with a target. Vuforia relies on UTF-8 character encoding standards and can recognize any listed character.

4.1. Specification of our architecture

We identified architectural requirements which are interdependent with the Location of Drug in the Environment (LDE). Therefore, in this section, we illustrate an algorithm to localize the drugs Dr_j and the number of pills. The

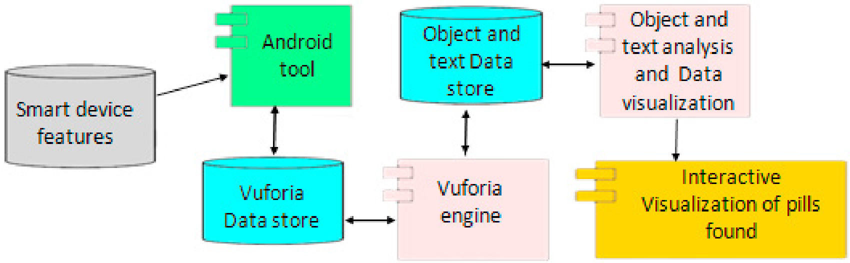

Fig. 2: Oriented system components of our architecture.

new architecture is defined as:

$$LDE = \{L_1, ..., L_{26}, Dr_1, ..., Dr_j\} \tag{1}$$

Each letter L_i can be lowercase Lo_i or uppercase Lu_i and is composed by a set of symbols S_i:

$$L_i = \{Lo_1, ..., Lo_{26}, Lu_1, ..., Lu_{26}, S_1, ..., S_6\} \tag{2}$$

Each drug Dr_m contains a set of pills Pl_j: that was created in the Vuforia database:

$$Dr_m = \{Pl_1, Pl_2, ..., Pl_j\} \tag{3}$$

After giving the specification of our architecture, we define an algorithm to localize the drug in the environment.

Algorithm 1 Location of Drug in the Environment (LDE)

Require: $exist = 0; message = ""; i = 1$
Ensure: $String\ result$ //message
  while $exist = 0$ and $i < length(BD) + 1$ **do**
    if $Dr_{name} == BD(i)$ **then**
      $exist = 1$
      $message = Dr_{name} + "FOUND"$
    end if
    $i++$
  end while
  if $exist = 0$ **then**
    $message = Dr_{name} + "Not\ FOUND"$
  end if
  $return\ message$

Algorithm 1 is used to find the place of the drug in the environment. The output is the message FOUND or NOT FOUND to verify whether the drug found or not (Figure 3.a). The Boolean variable EXISTE defines the existence of this drug in the environment. EXIST=1 when the drug is found and EXIST= 0 when otherwise. First, before starting to execute this function, it is initialized to 0 (ie the drug does not exist). The message variable is initialized to empty string. If the name of the drug Dr_{name} exists in the list of names of the drugs in the created database

(BD), then the variable EXISTS receives 1. Thus, the drug is founded in the environment. In addition, it displays a message can be capital Lu_i, lowercase Lo_i letters or symbols: $Dr_{name} + FOUND$. If not, the message displayed is $Dr_{name} + NOTFOUND$.

Fig. 3: Drug detection and recognition (a). Number of pills detection (b)

The existence of the drug in the user's environment is detected through speech and sensors. The user should move the camera in different sides of his environment to allow the system to detect the drug. Thus, the process of detection of pills existence may be easily executed.Our system can detect and recognize objects and items using Vuforia engine (Figure 5). It is also capable of running as a real-time app. It works with user smartphone via text-to-speech the name of object that they are looking for in his room e.g the name of the pills (for example, ALGESIC, AUGMENTIN, etc.) which are stored and collected in the datastore. Then, the algorithm identifies the object name and search for its position in the room. Finally, it indicates the number of pills left in the drug (Figure 3.b).

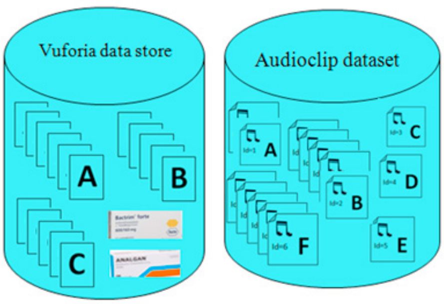

Fig. 4: Used databases

Word lists can be loaded from files and can be extended with additional words specific to the application use cases. The user can also filter word lists using a filter list to exclude the detection of certain words (using a black list filter) or allow only the detection of certain words (using a white list filter). Word lists represent a basic set of words that Vuforia can use as a reference to the appropriate text elements appearing in the field of view of the camera.

Our system detects all objects presented in front of the camera. The basic types of the synthesis system are: training, concatenation and pre-registration. The concatenate synthesis is based on the concatenation of the speech segments recorded from the audio clip database (Figure 4). Generally speaking, reading the word is the most difficult task, especially when it comes to pronouncing a word.

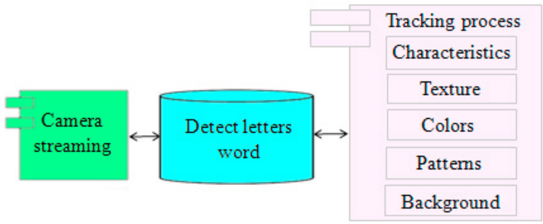

Fig. 5: Detection process method using Vuforia engine.

5. Evaluation

In this section, we assess the contribution of our proposed approach. First, we explain the experimental configuration for the evaluation of the proposed system. Then, we focus on the evaluation of the proposed system (Figure 6). The object if of the this evaluation is to test the potential of an AR device for the general public to improve the functional vision of the visually impaired and partially sighted people. The experiment is based on two groups. The first group is with the proposed AR system and the second group is without voice instruction. The results show that the first group give better results compared to the other group. During the evaluation of the performance of the proposed AR application, the participants have given positive comments and have confirmed their satisfaction in real life. In this part, we will detail the configuration steps to evaluate our proposed system. We tested our system with 20 participants. The latter were visually impaired peoples. In this evaluation, different levels of visual impairment which varies depending on the person were taken. The participants are 8 women and 12 men of different ages (between 45 and 75). Participants were identified as P1-P20, respectively. Each participant tested the application and after answering a short questionnaire (Q1-Q4).

- Q1: Is the application useful?
- Q2: Is the overall user interface easy to use? and are human-machine interactions easy to learn and use?
- Q3: Are the responses in real time?
- Q4: Are the application responses correct?

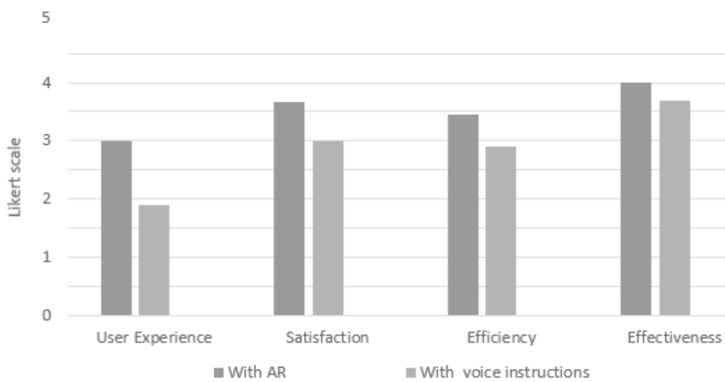

Fig. 6: Mean values per question for measuring usability on Likert scale .

The participant must give a mark to each question from N = 1 to N = 5, (very good, good, average, bad, very bad), respectively. We are doing this work to determine the effectiveness of the proposed system. Subjective assessments indicated that visually impaired participants with using our AR system found the application useful. Our analysis of participants' feedback helped to better understand the usefulness, expressiveness, usability and User-friendliness. Aimed at blind or visually impaired users, our system must be efficient, easy to learn and easy to use. Overall, the user-friendliness of the application was appreciated by the participants on the basis of the result of the tests (85% said positive). More specifically, based on the results of the participants responded positively on the main functionality of our system. Almost all participants said that the application was useful and it makes their lives easier (Q1 = 95%). The majority of participants (16 participants) found that the user interface design is easy to use (Q2 = 80%) and assessed the human-machine interactions as an easy system to use and also to learn. Most of the participants also estimated that the responses of the application are in real time (Q3 = 80%). Most of the participants were impressed by our approach: "it is much better to find the pills" (P15). According to them, our system is intuitive and effective but also "the sting that I like in this application that I do not need anyone, I do it all alone" (P8). As for effectiveness, there is a room for improvement with regard to correct responses. Regarding the score of participants in Q4, (85%). Nine participants pointed out that the inability to take the camera of the smartphone in front of the pills makes the proposed application fail, to allow to read the name of the pills to answer the questions whether the pills exists or not is exactly and finally how many pills left. We proposed to add voice input methods in the future to find the pills. Almost all participants were satisfied with our proposal.

6. Discussion

Based on the analysis performed on the test of object detection, the results show that our proposed system give good performance against different text styles, as long as the text still has a standard pattern on paper. The complex background, orientation and the size are solved through the Vuforia engine and the detection phase as well. We created another database that can deal with recognition phase of different drugs and pills. Our system based on Android-based Smartphone in real-time unlike [22] and [21] systems. The used architecture is similar to [23]. However, we use a different method to recognize texts and objects and add an audioclip dataset to read vowels. As we mentioned, we used two datasets that can handle the different letters and solve the problem of writing style that can be found in [14], [15] and [16] systems. A detection and recognition method based on ontology such as [25] is proposed as a Future work.

7. Conclusion and future work

In this article, we have presented an architecture that helps the visually impaired to find their drugs and determines the number of remaining pills. Hence, we have studied and analyzed the existing architecture under two categories based on the recognition of objects and text. This helped us to define the appropriate technologies for our architecture. Furthermore, we have described the open research issues and the most important limitation of existing works. Accordingly, we have designed an AR architecture for object recognition assistant for visually impaired people. Besides, we have mapped between schema architecture and suitable existing technologies. This allows dealing with the constraints and solve open research issues of existing works. Our architecture provides a straightforward mobile application that uses Vuforia as an AR engine to bring element with intuitively comprehensible interfaces in real time. This paper focused on solving the limitations of existing works for both text and object detection and recognition by taking into consideration visually impaired people needs. In the future, we will try to explore the prospects of extending the data store by improving the system for a bigger data store of drugs. In addition, we intend to implement Microsoft Hololens display system. This could be used to improve the AR experience to be more comfortable.

References

[1] Jonathan Huang, Max Kinateder, Matt J Dunn, Wojciech Jarosz, Xing-Dong Yang, and Emily A Cooper. An augmented reality sign-reading assistant for users with reduced vision. *PloS one*, 14(1), 2019.

[2] Jonathan Huang. A hololens application to aid people who are visually impaired in navigation tasks.

[3] Hanene Elleuch, Ali Wali, Anis Samet, and Adel M Alimi. A real-time eye gesture recognition system based on fuzzy inference system for mobile devices monitoring. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 172–180. Springer, 2016.

[4] Hanene Elleuch, Ali Wali, Anis Samet, and Adel M Alimi. Interacting with mobile devices by fusion eye and hand gestures recognition systems based on decision tree approach. In *Ninth International Conference on Machine Vision (ICMV 2016)*, volume 10341, page 103410D. International Society for Optics and Photonics, 2017.

[5] Slim Abdelhedi, Ali Wali, and Adel M Alimi. Human activity recognition based on mid-level representations in video surveillance applications. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 3984–3989. IEEE, 2016.

[6] Houssem Turki, Mohamed Ben Halima, and Adel M Alimi. Text detection in natural scene images using two masks filtering. In *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, pages 1–6. IEEE, 2016.

[7] Abdelkarim Ben Ayed, Mohamed Ben Halima, and Adel M Alimi. Mapreduce based text detection in big data natural scene videos. In *INNS Conference on Big Data*, pages 216–223, 2015.

[8] Houssem Turki, Mohamed Ben Halima, and Adel M Alimi. Scene text detection images with pyramid image and mser enhanced. In *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, pages 301–306. IEEE, 2015.

[9] Zied Selmi, Mohamed Ben Halima, Ali Wali, and Adel M Alimi. A framework of text detection and recognition from natural images for mobile device. In *Ninth International Conference on Machine Vision (ICMV 2016)*, volume 10341, page 1034127. International Society for Optics and Photonics, 2017.

[10] Houssem Turki, Mohamed Ben Halima, and Adel M Alimi. A hybrid method of natural scene text detection using msers masks in hsv space color. In *Ninth International Conference on Machine Vision (ICMV 2016)*, volume 10341, page 1034111. International Society for Optics and Photonics, 2017.

[11] Zi-han Wei, Xiao-ping Du, and Lei Cao. Research on the text detection and recognition in natural scene images. In *Eleventh International Conference on Graphics and Image Processing (ICGIP 2019)*, volume 11373, page 113730F. International Society for Optics and Photonics, 2020.

[12] Saad Bin Ahmed, Muhammad Imran Razzak, and Rubiyah Yusof. Progress in cursive wild text recognition. In *Cursive Script Text Recognition in Natural Scene Images*, pages 85–92. Springer, 2020.

[13] Zhenpeng Wang and Yongli Wang. Frca: High-efficiency container number detection and recognition algorithm with enhanced attention. In *Eleventh International Conference on Graphics and Image Processing (ICGIP 2019)*, volume 11373, page 1137306. International Society for Optics and Photonics, 2020.

[14] IL Shahana and HS Thaha. Malayalam text detection from natural-scene images. 2019.

[15] Hongchao Gao, Yujia Li, Xi Wang, Jizhong Han, and Ruixuan Li. Ensemble attention for text recognition in natural images. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.

[16] Peirui Cheng, Yuanqiang Cai, and Weiqiang Wang. A direct regression scene text detector with position-sensitive segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[17] Canjie Luo, Lianwen Jin, and Zenghui Sun. Moran: A multi-object rectified attention network for scene text recognition. *Pattern Recognition*, 90:109–118, 2019.

[18] MS Akopyan, OV Belyaeva, TP Plechov, and DY Turdakov. Text recognition on images from social media. In *2019 Ivannikov Memorial Workshop (IVMEM)*, pages 3–6. IEEE, 2019.

[19] Ghulam Jillani Ansari, Jamal Hussain Shah, Muhammad Sharif, and Saeed ur Rehman. A novel approach for scene text extraction from synthesized hazy natural images. *Pattern Analysis and Applications*, pages 1–18, 2019.

[20] Tushar Gupta, Mudita Sisodia, Schezeen Fazulbhoy, Mitali Raju, and Supriya Agrawal. Improving accessibility for dyslexic impairments using augmented reality. In *2019 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–4. IEEE, 2019.

[21] Gerardo Reyes-Ruiz and Marisol Hernández-Hernández. Augmented reality as a new and innovative learning platform for the medical area. In *Mixed Reality*. IntechOpen, 2020.

[22] Chanyong Kim, Hyojin Kim, Huiseung Son, Muhammad Tanseef Shahid, Hyun Chul Song, Francesco Piccialli, and Kwang Nam Choi. Marker based pedestrian detection using augmented reality. In *Proceedings of the 2019 3rd International Conference on Advances in Image Processing*, pages 19–22, 2019.

[23] Jonathan A Romero, Washington D Quero, Jorge S Sánchez, and Víctor H Andaluz. Training assistant for industrial processes through augmented reality. In *Proceedings of the 2019 11th International Conference on Education Technology and Computers*, pages 308–315, 2019.

[24] Hassan Abu Alhaija, Siva Karthik Mustikovela, Lars Mescheder, Andreas Geiger, and Carsten Rother. Augmented reality meets computer vision: Efficient data generation for urban driving scenes. *International Journal of Computer Vision*, 126(9):961–972, 2018.

[25] Imene Ouali, Faiza Ghozzi, Raouia Taktak, and Mohamed Saifeddine Hadj Sassi. Ontology alignment using stable matching. *Procedia Computer Science*, 159:746–755, 2019.