COMMUNICATIONS IN SCIENCE AND TECHNOLOGY
Homepage: cst.kipmi.or.id

Augmented reality based mobile application for real-time arabic language translation

Henda Chorfi Ouertani*, Lama Tatwany

Information Technology Department, College of Computer and Information Sciences, King Saud University Riyadh, 11421 Saudi Arabia

Article history:
Received: 8 August 2018 / Received in revised form: 23 May 2019 / Accepted: 02 June 2019

Abstract

Augmented Reality (AR) technologies are nowadays widely and ubiquitously used to enhance our real world experience in novel and enriched ways. With the new generation of smart phones and AR technologies, we are embracing a stimulating way of Human-Computer Interaction. Many AR mobile applications were developed in many fields such as education, health, design and Translation... This paper deals with text translation. Many applications are proposed on the market but they are yet limited to a few number of languages. Arabic language is one of these unsupported ones. To overcome this shortage, an AR mobile application for real time Arabic text translation is developed. The user simply hover the device's camera on the desired text and it will be translated automatically and rapidly. This type of application consists of three main components which are: text detection, text extraction and text translation. To ensure the application is performing well, the application is tested on different datasets under different conditions. We showed that the translation took least time in all datasets. However, the recognition took more time; it is due that the job is done on the cloud over the internet. In addition, a user study test is conducted to test the usability and user satisfaction.

Keywords: Arabic translation, augmented reality, AR.

1. Introduction

Nowadays, most people around the world own smart phones/devices, and it becomes necessary in our daily life. Almost all of the smart phones/devices have cameras; so we can use them to implement real time and real world applications using Augmented Reality (AR). Azuma defined Augmented Reality as "the system supplements the real-world with virtual (computer-generated) objects that appear to coexist in the same space as the real world" [1].

Augmented Reality is about displaying immediate direct or indirect multimedia information to the user interactively on real world objects using device's camera as shown in Figure 1 below.

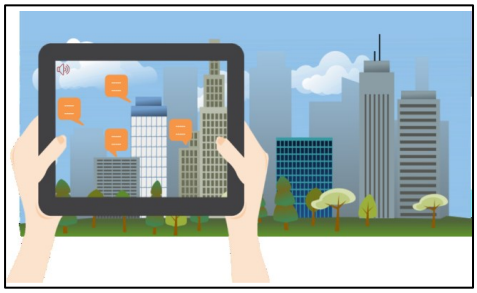


Fig. 1. Example of augmented reality application [1,2]

Mobile Augmented Reality applications are now growing increasingly and becoming more accessible due to the growth of smart phones/devices. Those applications affect many fields like education, medical, gaming, entertainments and translation to name but few. For Text translation, many applications are proposed on the market but they are yet limited to a few numbers of languages [2], [3], [4]. For further details, we refer to our paper “A Review on Using Augmented Reality in Text Translation” [5]. Arabic language is one of these unsupported ones. To overcome this shortage, an AR mobile application for real time Arabic text translation is developed and described in this paper.

2. System Architecture

The proposed application, named ARx2, consists of four main components which are: text detection algorithm, OCR, translation API, and the mobile application. Reviewing the literature [5],[6] helped us choosing the best environments and technologies to be used while developing the application. Indeed, we chose to use MSER algorithm to detect the text, ABBYY cloud SDK as the OCR and Google cloud translate API for translation. Thus, the architecture of the application is described in the figure below:

* Corresponding author. Tel.: +966 533914323;
Email: houertani@ksu.edu.sa.

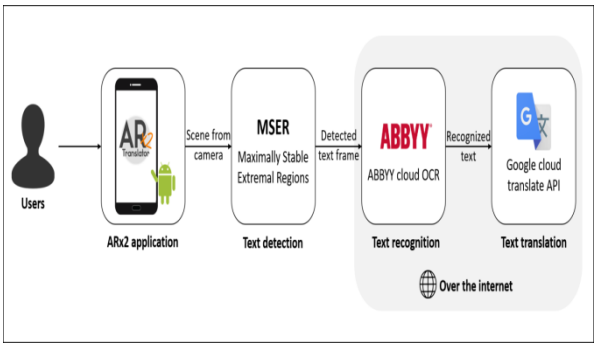


Fig. 2. System's Architecture

As shown in figure 2, A2R is an android mobile application which captures a real-world scene with the user's smart device camera, then detects the Arabic text in the scene using the MSER feature detector. Then ABBYY's cloud OCR is used to extract and convert detected text into machine readable format. After that, the text is translated to English using Google cloud translate API. Finally, the translated text is displayed to the user with the ability to hear the translation. The diagram below illustrates the system's input, processes, and output.

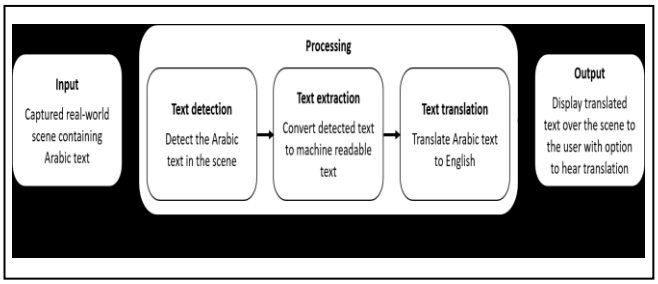


Fig. 3. Input, processes and output diagram.

- Input: is the captured real-world scene which contains the Arabic text.
- Processes: are the Arabic text detection, extraction, and translation.
- Output: is the displayed translated text to the user with the option to hear it.

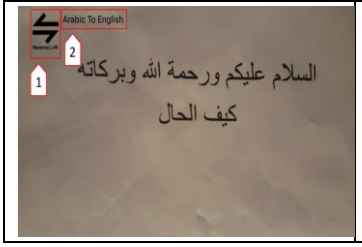
3. System Interfaces

The interface design of ARx2 translator is very simple and easy to use and could be used by wide range of people. Moreover, it was designed in respect to the accessibility requirements; so, it could be used by visually and hearing impaired people.

The user will be able to:

1. See translated text on the screen.
2. Hear the translated text by voice.

Table 1. ARX2 Interface

Interface	Description
	After the application is launched, the camera view is displayed. Point the camera on the desired text. 1. Click on it to reverse translation language between Arabic and English. 2. Display the language is translating from/to.
	When the text is detected and recognized: 1. Translated text is displayed. 2. Click on it to hear translation. 3. Click on it to copy the text in the clipboard. (a message will be displayed as shown below)
	Message which tells the user that the text is copied.

4. Results and Discussion

In order to evaluate ARx2 translator, multiple tests are conducted to ensure its usability and performance under different datasets and conditions. The details of this evaluation are described below.

4.1. Testing measurements

In order to test the application many measurements are used:

- The accuracy is calculated using the below formula:

$$Accuracy = \frac{number\ of\ images}{total\ number\ of\ images} \times 100$$

- The average time is calculated as follow:

$$Average\ time = \frac{\sum x_i}{n}, where\ x = time\ for\ each\ image\ and\ n = number\ of\ images$$

- The error rate is calculated by:

$$Error\ rate = \frac{\sum \frac{f}{n}}{4} x100, where\ f = number\ of\ failed\ images\ and\ n = number\ of\ images$$

4.2. Testing on different datasets

4.2.1. Formed dataset

A dataset collecting 159 images from Google images was formed. The images have mostly three lines of Arabic text with different background colors/images, font types, font sizes, font colors, quality, etc. to permit to test the application's capability under different conditions. The dataset contains some real-world scenes like streets signs and sentences from books to make it close to the real-world situations. Figure 4. below is showing samples from the dataset Figure 5. is showing the testing results using formed dataset. The number of success, failure and partially detected, recognized and translated images are shown in the Table 3.

Fig. 4. Formed dataset samples

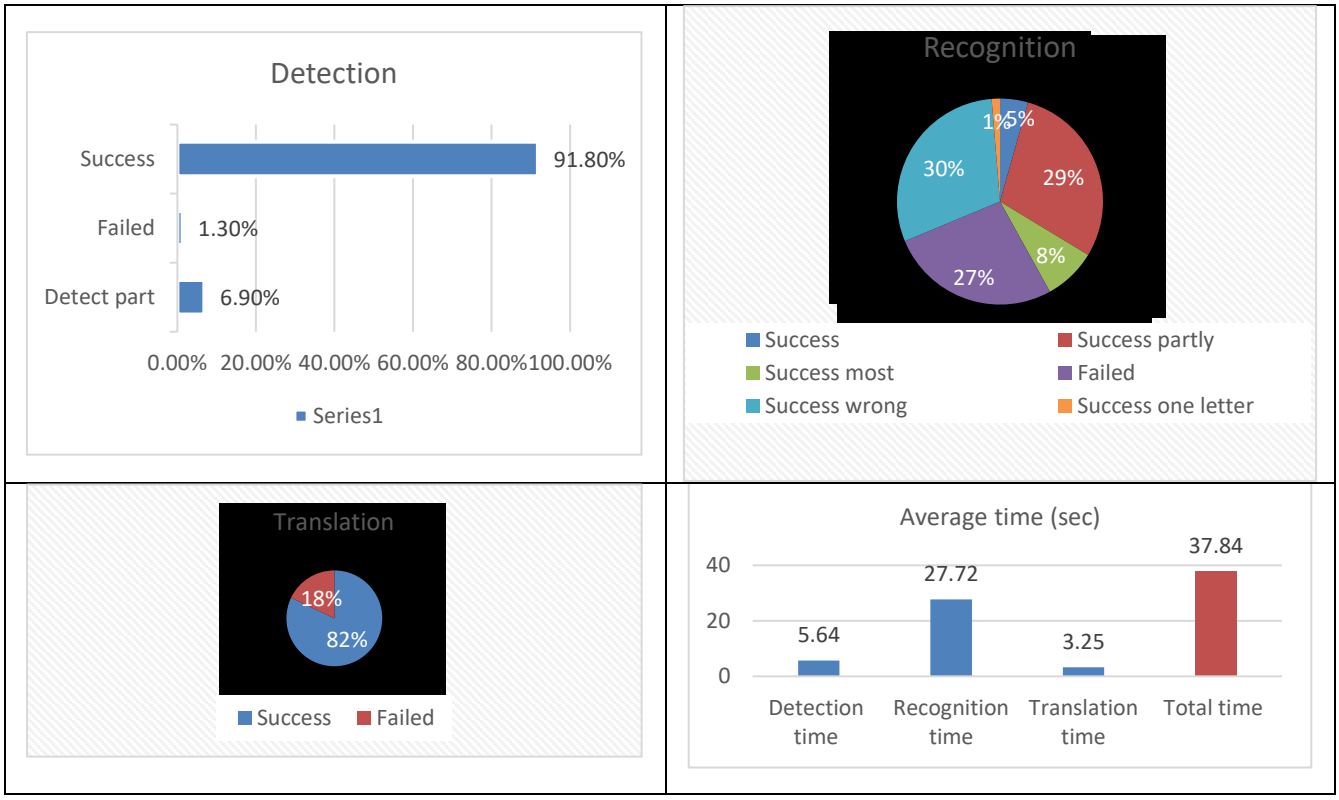


Fig. 5. Formed dataset testing results

Table 3. Number of images detected, recognized and translated in formed dataset.

	Success	***Failed***	***Part***
Text detection	146/159	2/159	11/159
Text recognition	69/157	42/157	46/157
Text translation	96/117	21/117	

4.2.2. ALIF dataset

ALIF dataset [7] is the first public Arabic text recognition in TV broadcast. It contains 6532 images of Arabic images taken from TV broadcasts. Each image, contains one sentence/word with variety of font colors, types, size, etc. and backgrounds colors, images.

One hundred and fifty random images from ALIF dataset were selected and printed out to test ARx2 translator. The dataset contains some real-world scenes like streets signs and sentences from books to make it close to the real-world situations. Figure 6. below is showing samples from ALIF dataset. Figure 7. is showing the testing results using ALIF the data set. The number of success, failure and partly detected, recognized and translated images are shown in the Table 4.

Fig. 6. ALIF dataset samples

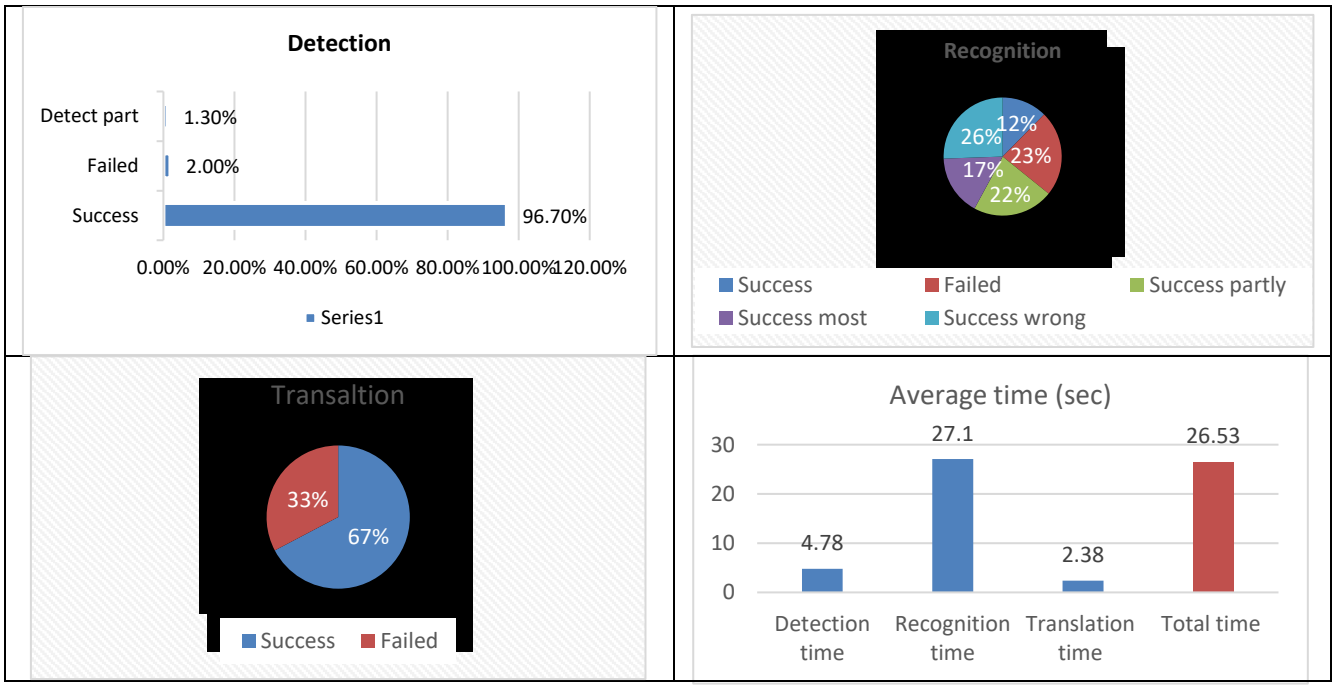


Fig. 7. ALIF dataset testing results

Table 4. Number of images detected, recognized and translated in ALIF dataset.

	Success	***Failed***	***Part***
Text detection	145/150	3/150	2/150
Text recognition	79/147	34/147	32/147
Text translation	76/113	37/113	

4.2.3. Random dataset

50 random images from the formed dataset and ALIF dataset were taken. This dataset is used to test the application from different displays in order to test the lighting effect. A tablet and a laptop as different displays are used.

Figure 8. below is showing the testing results using random dataset. The number of success, failure and partly detected, recognized and translated images are shown in the Table 5.

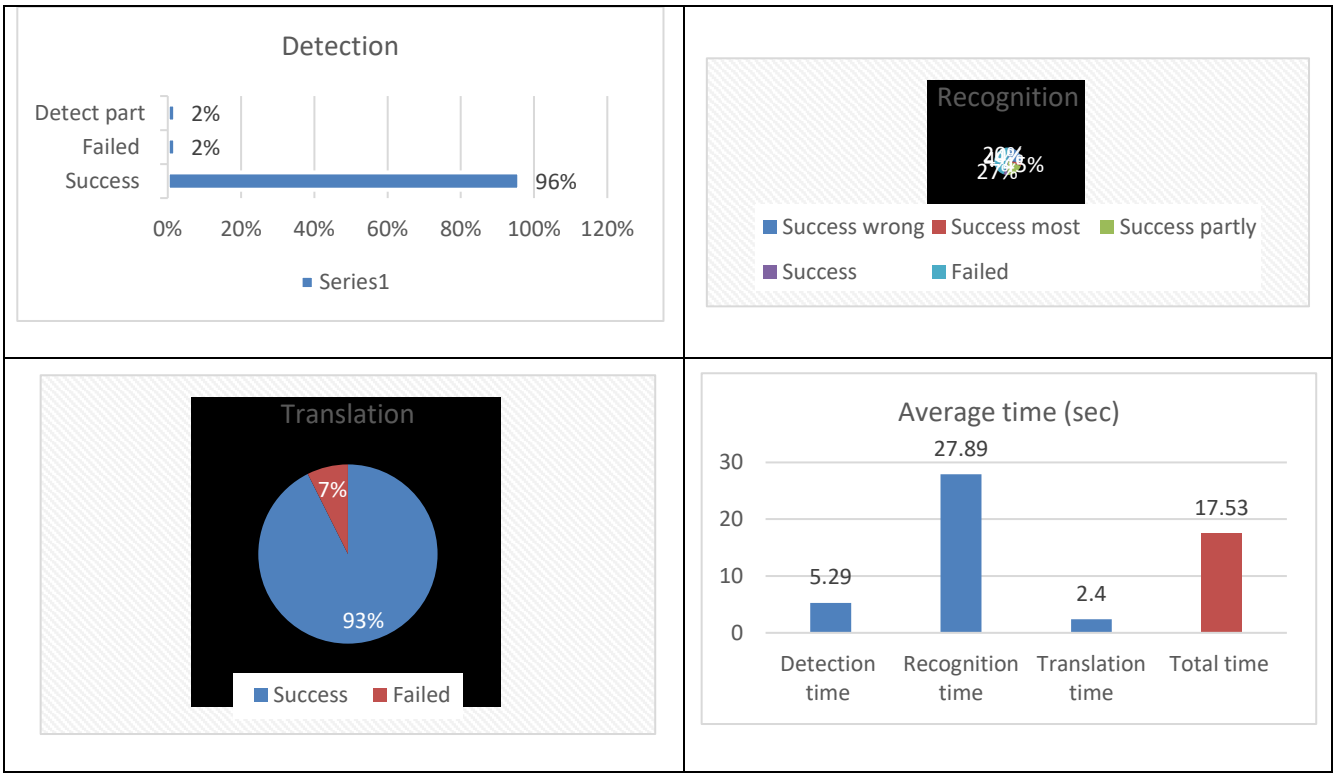


Fig. 8. Random dataset testing results

Table 5. Number of images detected, recognized and translated in random dataset.

	Success	***Failed***	***Part***
Text detection	48/50	1/50	1/50
Text recognition	17/49	22/49	10/49
Text translation	25/27	2/27	

4.2.4. Real-world dataset

In this test, we used 50 real objects with Arabic text, like: books, newspaper, species, chocolate, etc. to test real-world situations.

Figure 9 shows the testing results using random dataset. The number of success, failure and partly detected, recognized and translated images are shown in the Table 6.

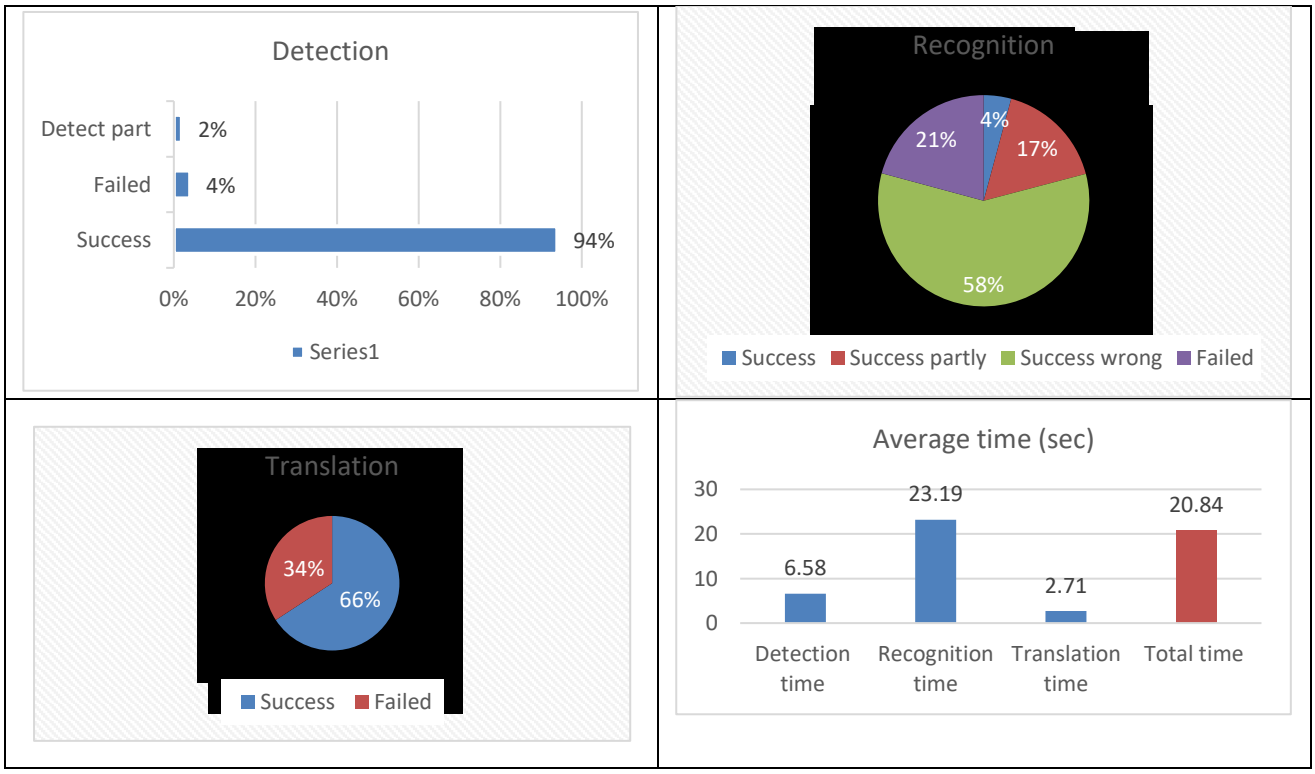


Fig. 9. Real-world dataset testing results

Table 6. Number of images detected, recognized and translated in real-world dataset.

	Success	*Failed*	*Part*
Text detection	48/50	1/50	1/50
Text recognition	17/49	22/49	10/49
Text translation	25/27	2/27	

4.2.5. Testing results discussion

In this section, testing results on different datasets and conditions are discussed. Note that the red rectangle is the detected text, blue text is the recognized text and the red text is the translation.

The cases that may happen while using the application:

Text is detected, recognized and translated completely correctly (success) as shown in the screenshots below.

Fig. 10. Successful testing

The text is not detected.

Fig. 11. Text detection failure

The text is not recognized

Fig. 12. Text recognition failure

Text will be recognized but totally wrong (success wrong).

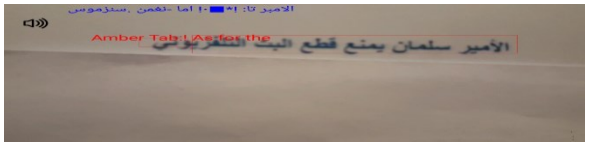


Fig. 13. Wrong success

Part of the text is recognized correctly

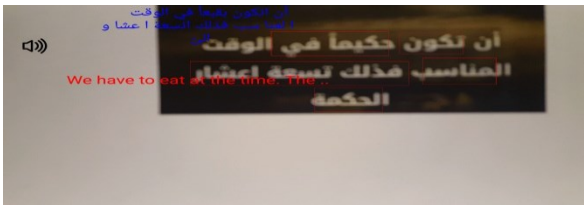


Fig. 14. Partial success

The majority of the text is recognized correctly.

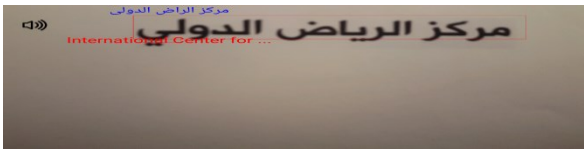


Fig. 15. Majority success

When an image contains non-Arabic text, it will be recognized but the noisy text affects the result.

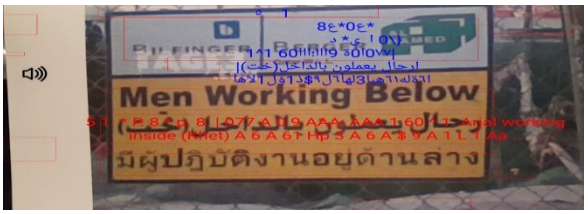


Fig. 16. Success with noisy text

Light does affect the detection and recognition results.

Fig. 17. Light's impact on results

The error rate for each translation step in all datasets is as follows:

Table 7. Error rate.

	Error rate
Text detection	1.98%
Text recognition	28.87%
Text translation	23.08%

As shown in the table above, the recognition has the highest error rate among other processes. Generally, from the chart below, we can see that the translation took least time with 2.68 seconds as an average in all datasets. However, the recognition took an average of 26.47 seconds; it is due that the job is done on the cloud over the internet.

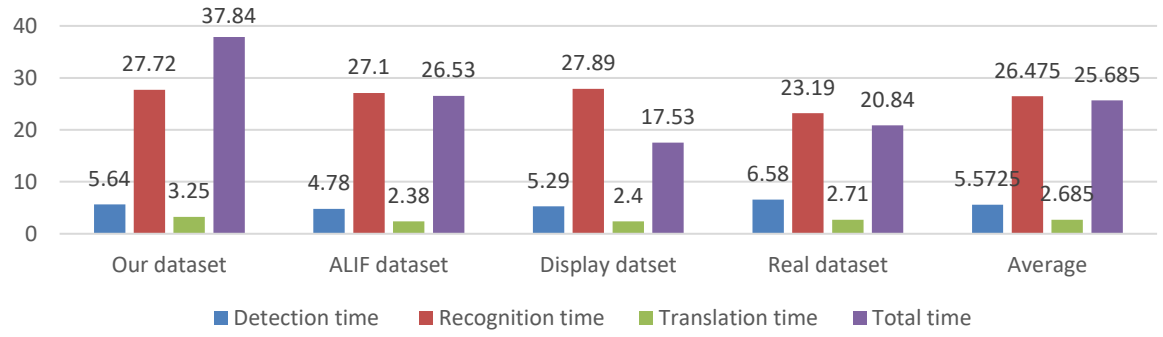


Fig. 18. Average time for all datasets

4.3. Testing with users

We have tested ARx2 translator application with different society categories, which are: normal Arabic speakers, non-Arabic speakers, visually and hearing impaired people. The tests' details and results are described below.

4.3.1. Participants characteristics

We have recruited 20 participants, all of them were from King Saud University and the test was conducted also at King Saud University. The participants' characteristics are summarized in the table below.

Table 8. Participants' Characteristics

PID	Age	Major		Impairment		Nationality
				Visually	Hearing	
1	31-40	Teacher/student	Teaching techniques			Saudi Arabia
2	25-30	Subway crew				Philippine
3	18-24	Subway crew	Midwifery			Philippine
4	18-24	Shop staff	Computer science			Pakistan
5	18-24	Student	Special education	✓		Saudi Arabia
6	41+	Sale woman				Bangladesh
7	31-40	Lecturer	English			India
8	18-24	Student	French language			Saudi Arabia
9	18-24	House keeper				Bangladesh
10	18-24	Student	Special education		✓	Saudi Arabia
11	25-30	Student	Special education		✓	Saudi Arabia
12	41+	Student	Special education		✓	Saudi Arabia
13	18-24	Student	Special education		✓	Saudi Arabia
14	25-30	Graduate	Special education		✓	Saudi Arabia
15	18-24	Graduate	Special education		✓	Saudi Arabia
16	25-30	Student	Art education		✓	Saudi Arabia
17	-	Blind writer		✓		Saudi Arabia
18	-	Blind writer	Arabic language	✓		Saudi Arabia
19	18-24	Student	Special education	✓		Saudi Arabia
20	25-30	Student	Special education	✓		Saudi Arabia

As we can see, we had two normal Arabic speakers P1 and P8, and six non-Arabic speakers' participants P2, P3, P4, P6, P7 and P9. Moreover, we had seven participants in the hearing-impaired group P10-P16. Where P12 wear hearing earphones, and the others use sign language to communicate and they cannot hear.

In addition, we recruited five participants in the visually impaired group P5, P17, P18, P19 and P20. P5 and P20 are totally blind, where P17 and P18 wear glasses and can see, P19 sees only very near things.

4.3.2. Test session

The test session was conducted individually within participant's free time at King Saud University. The whole session, took no more than half an hour with each participant. The session started by getting the participant's acceptance to participate in this study. After filling the contest form, we handed out a pre-questionnaire to take general information about participants. Then we introduced the application's idea and tried it out with them. After that, they were asked to fill out a post-questionnaire about their experience with the application.

4.3.3. Users testing results discussion

Participants were divided into four groups: normal Arabic speakers, non-Arabic foreigner, visually impaired and hearing impaired. Most of the participants are using a translation application in their phone, and most of them are using Google translate application. Surprisingly, most of them do not know about the real-time translation. On the other hand, there are five participants who uses real-time translation and all of them are from the visually and hearing-impaired participants. The others mostly use text for translating, then image and then voice. After trying the application with them, almost all of them found the applications idea very good. In addition, most of them will use this application if it was in their phones. All but one, found it enjoyable to use the application.

Most of the disabled participants found the application very helpful and usable. The two blind participants had different point of views; one stated that the application is really good and usable. However, the other one found it not that usable for them (blind people). Non-Arabic users found the application really helpful and make them understand easier and faster than the other translation ways.

5. Conclusion

Arabic language is one of the most complex languages around the world. Thus, developing an application that translates Arabic text to English using Augmented Reality will help to reduce language barrier for any users of Arabic language, for instance tourists in Arabic countries, or international students. The application developed will help users translate Arabic text faster and on real time with the minimum effort using their smart phones/devices. They only have to hover their mobile device's camera on the text to be translated.

References

1. R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, *Recent advances in augmented reality*, IEEE Comput. Graph. 21 (2001) 34-47.
2. B. Parhizkar, K. Oteng, A. H. Lashkari, and Z. M. Gebril, *Ubiquitous Mobile Real Time Visual Translator Using Augmented Reality for Bahasa Language,* Int. J. Inf. Educ. Technol. 3 (2013) 124.
3. V. Fragoso, et al., *TranslatAR: A Mobile Augmented Reality Translator* In, Proc. IEEE Workshop on Applications in Computer Vision (WACV), Kona, Hawaii, USA, 2011.
4. S. Mahbub-Uz-Zaman and T. Islam, *Application of augmented reality: Mobile camera based bangla text detection and translation,* BRAC University, 2012.
5. L. Tatwany, et al., *A Review on Using Augmented Reality in Text Translation, 6th International Conference on Information and Communication Technology and Accessibility,* 6th International Conference on Information and Communication Technology and Accessibility (ICTA), 2017, pp. 1-6.
6. D. V. Krevelen and R. Poelman, *A survey of augmented reality technologies, applications and limitations,* International Journal of Virtual Reality 9 (2010) 1
7. S. Yousfi, S. A. Berrani and C. Garcia, *ALIF: A dataset for Arabic embedded text recognition in TV broadcast*, The 6th International Workshop on Camera Based Document Analysis and Recognition, 2015, pp. 1-5.