International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I4P218
Volume 69 Issue 4, 125-130, April 2021

Sign Language Communication through Augmented Reality and Speech Recognition (LEARNSIGN)

Siti Nurulain Mohd Rum[#1], Braxton Isaiah Boilis [*2]

[#]*Faculty of Computer Science & Information Technology, Universiti Putra Malaysia*
43400, Serdang, Selangor

[1]snurulain@upm.edu.my, [2] braxtonboilis21@gmail.com

Abstract — *For the last few decades, Augmented Reality (AR) has become a popular technology for real-time interaction in virtual worlds. The advancement of AR is the result of an increase in the number of devices that able to unleash the potential of this technology. The primary value of AR is the way it combines the components of the artificial world with the experience of one's physical world. Studies show that this powerful feature can increase the effectiveness of the teaching and learning process in many fields. In this paper, we present the development of LEARN SIGN, an android-based mobile application for learning sign language. With the incorporation of AR and speech detection technologies, this system can also help people with disabilities, especially the deaf and mute community, to communicate with non-disabled people and vice-versa. To realize this research work, several technologies have been used, such as Unity Software and C# programming language, to develop the proposed system.*

Keywords — *Augmented Reality (AR), Sign Language, Disabled Community*

I. INTRODUCTION

Augmented Reality (AR) is an interactive real-world experience in which computer-generated cognitive feedback amplifies the real-world objects, often in a variety of sensory ways, such as visual, haptic, auditory, and olfactory[1, 2]. The beginnings of AR technology can be traced back to the 90s, when it is used primarily for training purposes in the field of aircraft engineering [3] and surgical [4]. Since then, AR has not received much attention as it requires the use of advanced devices. But today, there are many devices such as smartphones, handhelds, and other devices are available with the AR support feature to leverage the needs specifically for educational purposes, such as early education [5], engineering [6], science [7], and mathematic [8]. Recent literature review of the use of AR across disciplines also indicates the improvements in communication [9], motivation and inspiration [10], commitment [11], smart learning [12], as well as the increase of individual content memorization [13]. Mobile AR is one of the progressive research areas in AR resulted from the extensive smartphone development capabilities to support this technology. The feature offers versatile usage with a unique user interface (UI) to enhance the computing environment. Virtual data are assimilated by the mobile AR system into the physical environment so that the information can be predicted to imitate the surrounding system without requiring one to be in that physical place. The Mobile AR allows objects on computer systems to communicate directly with the physical world. People can communicate with each other to convey information, to ask and resolve questions, and even to collaborate with others. AR can also be seen as an auxiliary technology, because it has the potential to minimize the communication barrier and provide alternative methods to complete specific tasks. Typically, people with disabilities, especially the mute and deaf, have difficulty in communicating between themselves and others [14, 15]. A simple interaction between a non-disabled person and one who is deaf or mute is always more complicated since the non-disabled person is not familiar with the standard sign language (ASL)[16]. To engage with people, they often write to interact and communicate [17]. Sign language is not an easy way to use for interaction, especially in a large group of people [18]. As mentioned previously, if the receiver has not been trained to understand ASL, then communication is also not possible.

II. RELATED WORKS

AR is called Mixed Reality (MR) in general context [19], which refers to the spectrum of multi-axis areas encompassing Virtual Reality (VR), Telepresence, and other related innovation. Virtual Reality is a concept that enables users to access and communicate with the synthetic worlds in computer-generated 3D environments [20, 21]. The users can "immerse" themselves into various levels in the computer-generated world that can be a simulation of any form of reality world [10] or simulation of any complex phenomena [22]. Three aspects that must be kept in mind when developing the AR system are, firstly, it must be the combination of real and virtual worlds; secondly, it has to be in the real-time interactivity; and lastly, it has to be registered in the 3D form. The HDM (Head-Mounted Displays) can be used to present the augmented scenery [23]; however, other technologies such as holographic display, smart glasses [24], and handheld or smartphone-based can be used to support the scenery [25]. In addition to the three factors listed, one more aspect that needs to be considered is portability due to the limitation of devices, a user in the virtual environment usually unable to go around much to imitate the physical world. Portability is thus becoming an important issue if the AR system requires the user to walk through a wide area.

A. The Components of AR

There are several components required for the development of AR applications, which are described as follows:

a) AR's Scene Generator

The devices or tools that responsible for rendering in AR application is called scene generator. The rendering is not one of the big issues at present since the object that needs to be drawn are few and often not necessary to make realistic scenery image of an application [26].

b) Tracking System

One of the most critical issues of AR systems is the tracking method, primarily due to the registration problem [26]. The objects must be appropriately parallelized with the virtual world, and the illusion of these two worlds should have coexisted. However, many applications in the industry, especially in the medical field, require accurate registration [26].

c) Display

The HMD (Head Mounted Monitor) was used by many of the AR display systems, particularly in gaming, medical, aviation, and aerospace, worn on the human head or as part of a helmet. Today, with the advancement of technologies in IT gadgets, the viewing of AR images can be supported by many other instruments. The example of devices, including eyeglasses, head-up display, handheld devices, contact lenses, projector, monitors, and display systems. Two specific options are possible when integrating the physical and the virtual world, the optical and video technology. Depending on factors such as resolution, accessibility, the field of view, registration methods, and others, each of them has certain tradeoffs [27]. A device with a similar appearance to eyeglasses can render the AR displays. The version includes glasses that use a camera to intercept the real-world views and re-display the augmented views through glasses [28] and devices that able to project the AR image reflected from the surface of the lens plate. The HUD (Head-up display) displays the virtual image of the actual image and presents the information to the users without forcing them to look away from their normal point of view [29]. The VRD (Virtual retina display), also known as retina scans view or retina projector is a display technology that draws raster views directly to the retina of the eye [30].

B. AR in supporting the Sign Language.

Today, there are many mobile AR systems available for download to support sign language communication [31-33]. The work by Norziha et al. [34], for example, the functional requirements for designing the Augmented Reality (AR) book, is identified for the use of the deaf student to learn science subject specifically in microorganisms topic. The qualitative type of research is used, such as interaction and observation of the three deaf students to obtain the requirement. The study revealed that the students preferred to have the Malay and English text for the AR book. As a result, the AR-SiD [34] is developed to support two languages, Bahasa Malaysia and English. The system imitates the physical look of microorganism or protozoa (microscopic view) when students scan the marker model, and this enables the user to view it in the form of the 3D display. The MagicBook [35] by Hirokazu et al. used the virtual overlay models on real book pages for AR scene creation. The application provides an effortless feature for the user to experience the augmented reality scene to immersive virtual reality by using a head-mounted display. The main objective of MagicBook is to support one's cognitive ability improvement. A 3D virtual model will be appeared based on the user's scan point anywhere in the book. The aim of SignAR [36] is to ease the problem of the deaf community in learning English. In this system, the AR can be mapped to every word and text to replace the human interpreter. In addition, the developed mobile application utilized Natural Feature Tracking (NFT), which enables the overlay of the 3D objects to sit on the top of actual images. Fig. 1 presents the features comparison summarization with the proposed system. In this paper, we present the proposed system named LEARN SIGN with the primary objective to reduce the communication gap between non-disabled with disabled people, especially the mute and deaf people.

FUNCTIONS/ FEATURES	AR-SiD [22]	MAGICBOOK [23]	SignAR [24]	LEARNSIGN
Interface	Book	Book	Mobile Application	Mobile Application
Ways to display AR model	AR Glasses	AR Glasses	Phone's camera	Phone's camera
Natural Feature Tracking	x	x	/	/
Supported Language	Bahasa Malaysia and English	English	English	Bahasa Malaysia and English
Speech to Sign Language Conversion	x	x	x	/
Alphabet/Text to Sign Language Conversion	x	x	/	/
Target User	Deaf students for learning Microorganism topic in Science Subject	All	Deaf community	All

Fig 1. Features comparison between existing systems and the proposed system.

III. METHODOLOGY

In this section, we describe the steps required for developing the proposed system. The methodology to develop the LEARN SIGN is divided into two main phases, and the first phase is the 3D content creation. There are two activities in 3D content creation, namely the 3D model construction and the pose model creation.

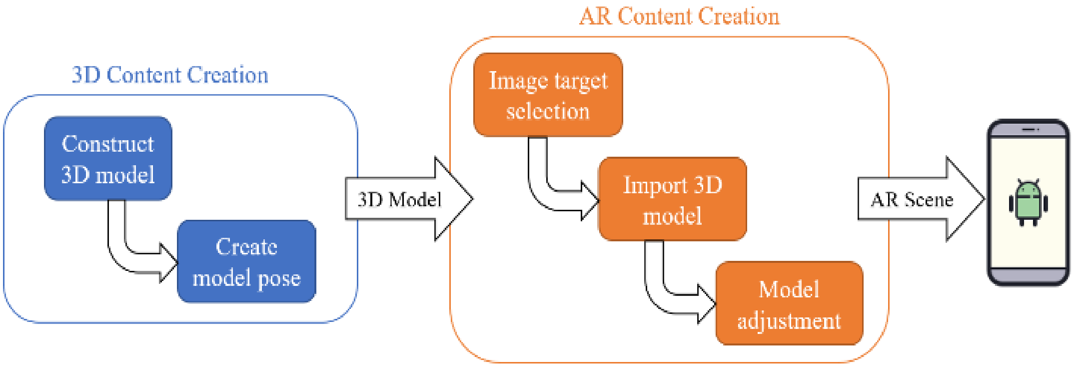

Fig 2. Methodology for LEARN SIGN development

The second phase is the AR content creation that consists of three activities, namely image target selection, 3D model import, and model adjustment. In the 3D content creation phase, the 3D model is constructed using the Blender software. This software is an open-source tool for the 3D creation suite. It supports the entire process of 3D development, from modeling to rigging, simulation, animation, rendering, motion composition, and tracking, and even video editing as well as game making. In the AR content creation phase, Unity Software, alongside its Vuforia extension, is used to implement the augmented reality feature.

A. 3D Content Creation

In this phase, the 3D human hand model is constructed using the Blender software. The steps for modeling the 3D image are presented in Fig 3. Whereby Fig 4 is the process of finalizing the 3D hand model creation. Since the Blender software is not an image-based modeling software, therefore designing the model from scratch is required. The real human hand needs to be imitated in terms of its characteristics and physical look. Since Blender is not an image-based modeling software, therefore development from scratch is required. To create the model, it is started by creating the cube, and it needs to be shaped step by step as shown in Fig 2 until it becomes the shape of the human hand. The second activity of this phase is the model pose creation. The Blender software features called armature are used to create the pose for the model once the development of the human hand 3D model development is completed. During this activity, the armature is inserted into the created model. A Blender armature can be said as identical to a real skeleton armature because an armature can consist of several bones, much like a real skeleton. By using the armature, the movement of the 3D model can be controlled to the desired gesture or pose of sign language.

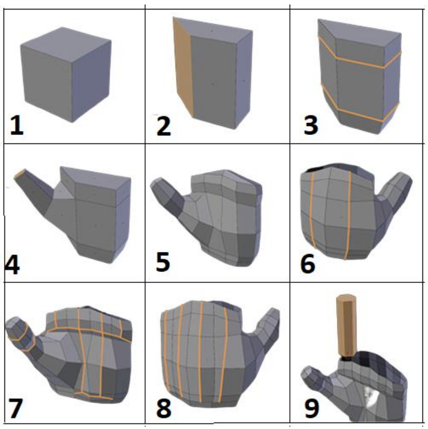

Fig 3. Steps for modeling the 3D of the human hand.

Fig 4. 3D hand model creation.

B. AR Content Creation

This phase consists of three steps, and the first step is the image target selection. The image targets reflect the images that can be detected and tracked by the Vuforia Engine. The way the engine detects the images is by comparing the natural features extracted from the image of the camera with a database of known target sources. If the image target is established, Vuforia Engine uses tracking technologies to detect the image and augment the content seamlessly. A resolution of 2.25 MB or less and a minimum width of 320 pixels is the minimum requirement of the input image. The features obtained from these images are kept in a server that is eventually accessible and packaged with the application. The Vuforia Engine will then use the index for runtime comparisons. The targeted image is captured and stored in the database. Selecting the right targeted image is important as this will determine the target 3D model to be displayed. Fig 5 is the process to associate the 3D hand model with the target image. The next step that needs to be performed is to import the created 3D model in the first phase into the augmented reality content. Both the 3D model together with the pose are imported to the Unity software. The hand models are placed on top of the image targets. This to ensure that only relevant 3D model display based on the assigned image target. The final step is to adjust the model. In this step, the process such as moving, scaling, and rotating the imported 3D model has adjusted accordingly. This to ensure that the developed model can be perfectly fit into the mobile's screen.

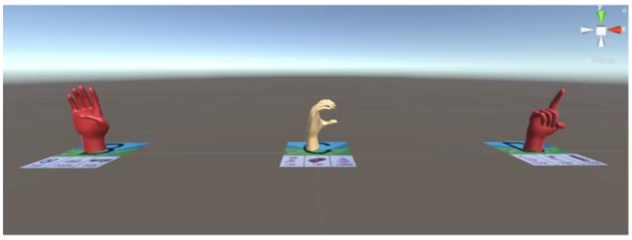

Fig 5. Associating the 3D hand model to the target image.

C. Speech Recognition

The Android Speech recognizer has been around for several years. It operates similarly to the approach from Apple since it still uses the external device to manage speech recognition. More specifically, it is used to translate voice to text and vice versa. Speech recognition in mobile applications is a feature that is in increasing demand to support natural communication. It is a very helpful tool to integrate into the mobile application. Fig 6 is the flow process of speech recognition and model matching. The process is started with the speech recording and follow by the speech analysis. During the speech analysis, the voice is converted into text. The text is then matched to the model stored in the database. The matched 3D model will then be displayed on the user screen. In this study, the Android Speech technology is incorporated into LEARN SIGN system.

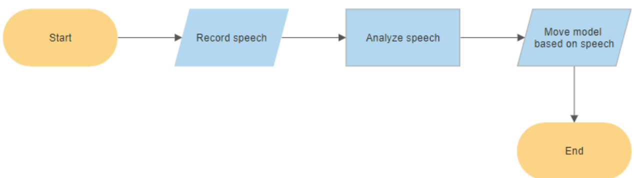

Fig 6. The speech scene flow process

D. Mobile Application Development

In this phase, the created model is integrated with the mobile application development. The Visual Studio is used as an editor to write the codes. Table 1 is the technologies used to develop the whole system.

IV. TABLE 1. THE TECHNOLOGIES USED TO DEVELOP THE WHOLE SYSTEM.

Technologies	Description
Android	The operating system of the mobile application
Unity	The software used for the end-to-end creation of the 3D mobile application and enable the AR features to be incorporated.
Visual Studio	The editor tool used to write the C# codes
Blender	To support the entire process of the 3D pipeline for model creation from modeling to the motion tracking
Vuforia Image Target Database	The extension engine provided by the Unity Software used to detect and track the camera images and compare them against the target resource database
C# programming	The programming language used to build the whole system

V. RESULT

The final look of LEARN SIGN is presented in Fig 7. The left side of the picture is the main interface of the LEARNSIGN system. To develop the system, the android device with android 4.4 or higher was used. The Unity software is the preferred technology to develop the system as it supports the android-based application platform as well as the implementation of augmented reality technology. There are three features provided in this system. The first one is the conversion of the alphabet to the sign language, and the second one is the conversion of words to the sign language, and the last one the speech conversion to the sign language. Users can select any scenes that they wanted to. The alphabet scene is where users can scan the designated target (any letter) to see the 3D model of sign language, as shown in Fig 7. If the word scene is selected, the users would be directed to the screen that presents the motion of sign language based on the scanned words. The alphabet and word scenes are almost the same as both require access to the device's camera to scan the designated target to display the sign language in augmented reality. In the speech scene, users can use the recorded voice as the input for the text conversion and then translate it into the appropriate motion of the sign language.

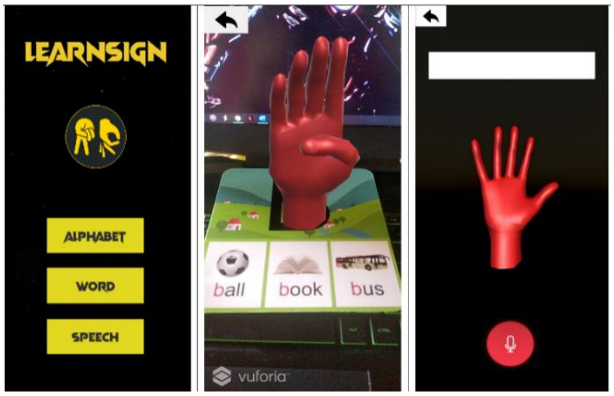

Fig 7. The final look of LEARNSIGN

The right side of Fig 7. is the speech scenes provided by LEARN SIGN application. The models used in the application are developed using the Blender software. This software grants the ability to construct the 3D hand model alongside the animations. The hand model in LEARN SIGN system is developed using the Blender software. This software allows the creation of 3D objects made easy alongside the model's animation.

VI. CONCLUSION

This paper presents the development of LEARN SIGN, a system for learning sign language through a mobile application. Behind LEARSIGN, there are several technologies used to build the whole system, such as C# programming language, Unity Software, Blender, and Vuforia Image Target Database. Typically, people with disabilities, especially the mute and deaf, have difficulty communicating between themselves and others. A simple interaction between a non-disabled person and one who is deaf or mute is always complicated, as the non-disabled person is not familiar with the standard sign language. Therefore, the primary objective of this study is to help to reduce the communication gap between non-disabled people and the disabled community, especially the deaf and mute people, by providing the LEARN SIGN as the support tool. The features provided in LEARN SIGN are simple and make it convenient to be used by the users. The users can simply scan the words or letters on any document, or they can also record their voice as input to the system. The system will then translate these inputs into sign language. The languages supported by the system are English and Bahasa Malaysia. For future development, the system can be improved by adding real-time sign language detection and increase detection accuracy. It can also be an interesting and useful application if the created model can be embedded in any movie player as an optional plugin to watch any video or movie that able to translate the text audio and display the sign language.

REFERENCES

[1] R. Aggarwal and A. Singhal, Augmented Reality and its effect on our life., in 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), (2019) 510-515.

[2] P. Schueffel, The concise FINTECH compendium, Fribourg, Switzerland, (2017).

[3] P. Thomas and W. David, Augmented reality: An application of heads-up display technology to manual manufacturing processes, in Hawaii international conference on system sciences, (1992) 659-669.

[4] M. Bajura, H. Fuchs, and R. Ohbuchi, Merging virtual objects with the real world: Seeing ultrasound imagery within the patient, ACM SIGGRAPH Computer Graphics,26(203-210)(1992).

[5] M. Z. Masmuzidin and N. A. A. Aziz, The current trends of augmented reality in early childhood education, The International Journal of Multimedia & Its Applications (IJMA), 10(47)(2018).

[6] A. Nesterov, I. Kholodilin, A. Shishkov, and P. Vanin, "Augmented reality in engineering education: Opportunities and advantages, Communications-Scientific letters of the University of Zilina, 19(2017) 117-120.

[7] D. Sahin and R. M. Yilmaz, The effect of Augmented Reality Technology on middle school students' achievements and attitudes towards science education, Computers & Education, 144(2020) 103710.
[8] S. Cai, E. Liu, Y. Shen, C. Liu, S. Li, and Y. Shen, Probability learning in mathematics using augmented reality: impact on student's learning gains and attitudes, Interactive Learning Environments, 28(2020) 560-573.
[9] X. Ran, C. Slocum, M. Gorlatova, and J. Chen, ShareAR: Communication-efficient multi-user mobile augmented reality, in Proceedings of the 18th ACM Workshop on Hot Topics in Networks, (2019) 109-116.
[10] T. Khan, K. Johnston, and J. Ophoff, The impact of an augmented reality application on learning motivation of students, Advances in Human-Computer Interaction, (2019).
[11] H. F. Hanafi, C. S. Said, M. H. Wahab, and K. Samsuddin, Improving students' motivation in learning ict course with the use of a mobile augmented reality learning environment, in IOP Conference Series: Materials Science and Engineering, (2017)(012114).
[12] G. Kiryakova, N. Angelova, and L. Yordanova, The potential of augmented reality to transform education into smart education, TEM Journal,7(556)(2018).
[13] H. Ahmad, N. M. M. Zainuddin, and R. C. M. Yusoff Augmented Operational Reality Framework to Aid Al-Quran Memorization for Hearing Impaired Students, Open International Journal of Informatics (OIJI), 6(2018) 22-32.
[14] P. Vijayalakshmi and M. Aarthi, Sign language to speech conversion, in 2016 International Conference on Recent Trends in Information Technology (ICRTIT), (2016) 1-6.
[15] A. R. Hasdak, I. Al Nur, A. Al Neon, and H. U. Zaman, Deaf-Vibe: A Vibrotactile Communication Device Based on Morse Code for Deaf-Mute Individuals, in 2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC), (2018) 39-44.
[16] C. Baker and D. Cokely, American sign language," Silver Spring, MD: TJ Publishers, (1980).
[17] R. Saha, A. Sharma, and M. Srivastava, Psychiatric assessment of deaf and mute patients–A case series, Asian journal of psychiatry, 25(2017) 31-35.
[18] M. J. Cheok, Z. Omar, and M. H. Jaward, A review of hand gesture and sign language recognition techniques, International Journal of Machine Learning and Cybernetics, 10(2019) 131-153.
[19] P. Milgram and F. Kishino, A taxonomy of mixed reality visual displays, IEICE TRANSACTIONS on Information and Systems, 77(1994) 1321-1329.
[20] A. Van Dam, A. S. Forsberg, D. H. Laidlaw, J. J. LaViola, and R. M. Simpson, Immersive VR for scientific visualization: A progress report, IEEE Computer Graphics, and Applications, 20(2000) 26-52.
[21] W. R. Sherman and A. B. Craig, Understanding virtual reality: Interface, application, and design: Morgan Kaufmann, (2018).
[22] I. Radu and B. Schneider, What Can We Learn from Augmented Reality (AR)? Benefits and Drawbacks of AR for Inquiry-based Learning of Physics, in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, (2019)(1-12).
[23] A. Petrovskaya and P. Varvak, Optimizing head-mounted displays for augmented reality, ed: Google Patents, (2018).
[24] H. M. Nguyen and E. L. Coons, Augmented reality glasses, ed: Google Patents, (2018).
[25] S. Serubugo, D. Škantárová, L. K. Nielsen, and M. Kraus, Comparison of wearable optical see-through and handheld devices as a platform for an augmented reality museum guide, in International Conference on Computer Graphics Theory and Applications, (2017) 179-186.
[26] R. Silva, J. C. Oliveira, and G. A. Giraldi, Introduction to augmented reality, National laboratory for scientific computation, 11(2003).
[27] T. Plewan, B. Mättig, V. Kretschmer, and G. Rinkenauer, Exploring the benefits and limitations of augmented reality for palletization, Applied Ergonomics, 90(2021) 103250.
[28] B. Forutanpour, S. Balasubramanyam, and L. E. Mann IV, Input method designed for augmented reality goggles, ed: Google Patents, (2015).
[29] N. Shimoda, Y. Fujita, T. Arakawa, and A. Misawa, "Head-up display, ed: Google Patents, (2020).
[30] T. A. Furness III and J. S. Kollin, Virtual retinal display and method for tracking eye position, ed: Google Patents, (2001).
[31] A. K. O. Sudana, I. G. A. A. M. Aristamy, and N. K. A. Wirdiani Augmented reality application of sign language for deaf people in Android based on the smartphone, International Journal of Software Engineering and Its Applications, 10(2016) 139-150.
[32] J. J. C. Garnica and M. A. G. Arrieta, Augmented reality sign language teaching model for deaf children, in Distributed Computing and Artificial Intelligence, 11th International Conference, (2014) 351-358.
[33] A. Vaitkevičius, M. Taroza, T. Blažauskas, R. Damaševičius, R. Maskeliūnas, and M. Woźniak, Recognition of American sign language gestures in virtual reality using leap motion, Applied Sciences, 445(9)(2019).
[34] N. M. M. Zainuddin, H. B. Zaman, and A. Ahmad, "Developing augmented reality book for the deaf in science: the determining factors, in 2010 International Symposium on Information Technology, (2010) 1-4.
[35] M. Billinghurst, H. Kato, and I. Poupyrev, The MagicBook: a transitional AR interface, Computers & Graphics, 25(2001) 745-753.
[36] B. Noor-Un-Nissah and M. H. Joseph, SignAR: A Sign Language Translator Application With Augmented Reality, Journal of Applied Technology and Innovation (e-ISSN: 2600-7304), 3(2019).