A Framework of Adaptive Multimodal Input for Location-Based Augmented Reality Application

Rimaniza Zainal Abidin, Haslina Arshad, Saidatul A'isyah Ahmad Shukri
Centre of Artificial Intelligence Technology, Faculty of Information Science and Technology,
University Kebangsaan Malaysia, 43600 UKM Bangi, Selangor
nizazainal90@gmail.com

***Abstract*—Location-based AR is one of the most familiar mobile application currently being used. The position of the user relative to the real world will be located and digital information can be overlaid to provide information on the user's current location and surroundings. Four main types of mobile augmented reality interfaces have been studied and one of them is a multimodal interface. Multimodal interface processes two or more combined user input modes (such as speech, pen, touch, manual gesture and gaze) in a coordinated manner with multimedia system output. In the multimodal interface, many frameworks have been proposed to guide the designer to develop multimodal applications including in augmented reality environment but there has been little work reviewing the framework of adaptive multimodal input in mobile augmented reality application. This paper presents the conceptual framework to illustrate the adaptive multimodal interface for location-based augmented reality application. We reviewed several frameworks that have been proposed in the field of multimodal interfaces, adaptive interface and location-based augmented reality. We analyzed the components in the previous frameworks and measure which input modalities can be applied in mobile devices. Our framework can be used as a guide for designers and developers to develop a location-based AR application with an adaptive multimodal interaction.**

***Index Terms*—Adaptive Interfaces; Mobile Augmented Reality; Multimodal Interfaces; Mobile Sensors.**

I. INTRODUCTION

Augmented Reality is a technology that overlays virtual imaginary onto the real world. There are three characteristics of Augmented Reality: combining real and virtual images, the virtual imagery is registered with the real world, and it is interactive in real time [1]. AR is an emerging technology in many mobile applications recently. Diverse applications have been produced for the smartphones mainly to offer more convenience and innovative ideas [2-4]. There are four main types of mobile augmented reality interfaces and one of them are a multimodal interface. In multimodal interface (MMI), many frameworks have been proposed to guide the designers and developers to design a multimodal application as well as in augmented reality environment. There are frameworks which illustrate the system flow in the field of the multimodal interface, adaptive interface and multimodal AR interface. Although many of the proposed frameworks are focusing on the fields of MMI, AR and adaptive interface, there has been a little review of the adaptive multimodal interface in location-based mobile augmented reality.

In this paper, we will review several frameworks that have been proposed in the field of multimodal interfaces, adaptive interface and augmented reality. We will focus on appropriate input modalities which can be used for a location-based augmented reality application in a tourism domain. At the end of this paper, we will propose a conceptual framework to illustrate the adaptive multimodal interface in location-based AR application. This framework is expected to help developer designing their adaptive multimodal AR application. The paper starts by briefly discussing an overview of mobile augmented reality. The next section will briefly explain about the multimodal interface and adaptive interfaces. The paper then reviews the related study based on frameworks that have been proposed by previous researchers. We also analyze specific modalities that can be used as an appropriate input to the system. Finally, the paper concludes with a proposed framework.

A. Location-based Augmented Reality Application

Modern smartphones recently offer an opportunity to produce more powerful augmented reality application where the video stream of the away facing camera creates a live snapshot of the user's surrounding and supplement it with superimposed virtual content in real-time.

Location-based AR is one of the most familiar mobile application currently being used. The position of the user relative to the real world will be located and digital information can be overlaid to provide information on the user's current location and surroundings [22]. Users can naturally interact with location-based AR system using multimodal interaction through the user interface.

According to [4], mobile AR interface is defined as a medium for displaying information merged with the real-world environment mapped with augmented reality surrounding in a single view. Mobile phone's camera was used to analyze the real environment and supplement the virtual content onto it. Four types of MAR interfaces are a tangible interface, collaborative interface, hybrid interface and multimodal AR interface.

Multimodal AR is defined as a combination of real object and the system naturally in the forms of language and behaviours such as speech, touch, natural hand gestures, or gaze [17-4]. For instance, the user can see a virtual 3D model of brain system and interact with the virtual content (e.g. zooming). It is better than learning the brain system by reading a book or watching the video. It shows that an interaction in mobile augmented reality is crucial to deliver understandable information and offers more enjoyment for users.

B. Multimodal Interface

Multimodal Interface (MMI) processes two or more combined user input modes (such as speech, pen, touch, manual gesture, gaze, and head and body movements) in a coordinated manner with multimedia system output [5]. Diverse studies in MMI area have shown various possibilities in which modalities can be merged. Pioneer and well-known Bolt [9] in his "Put that There" MMI system has demonstrated the hand gestures and speech which is used in a complementary fashion. This system allows the users to move objects exhibited on a wall display. Multimodalities indicate the use of more than one modality either simultaneously or sequentially for input and output [10]. With the growth of mobile devices especially for a modern smartphone, it is equipped with various sensors such as an accelerometer, compass, camera, and proximity sensor. These sensors can provide an input that can be adapted to the systems. An advent concept in the context of mobile interaction is multimodality [10-11]. It is not only an explicit interaction (speech and gesture) but also more implicit information, which is gathered from several sensors that are currently available on mobile devices. However, using multimodal interaction might face some problems such as user needs to spend the time to learn how to use the combination of the unimodal interface. MMI will also increase mental workload where the user has to pay attention in choosing an appropriate interface type to complete the activity for a given task [8]. Particularly in mobile settings, unimodal interaction modes can suffer from limitations (e.g. small screen problem). Therefore, in this paper, we propose an adaptive interface to tackle the problems.

C. Adaptive Interface

Type of users and their preferences are heterogeneous. Many researchers have put an effort to make the interface more interactive and flexible to accomplish user needs and in a specific context conditions. Rothrock et al. [14] have defined an adaptive interface as system adapting its displays and available actions to user's current goal and abilities by monitoring user status, system task and current situations. The techniques for adaptations include what information to present, how to interact with the information and how to present the information [14]. For mobile adaptive interface, the device adapts their behaviour based on interaction context variations such as user, environment and the device itself [4]. As mentioned in the previous section, there are many sensors equipped with mobile devices that can provide an input information that can be adapted to the systems. We can use various sensors to obtain the information such as user's current location, environment factors (weather, temperature, and ambient light), number of user's steps, video, picture, sound etc. Existing sensors can also trigger an interaction between the user and mobile devices (e.g. gestures, speech, and touch). As MMI utilizes the multiple input modalities to enhance the human-computer interaction, they are suitable for adaptive interfaces.

II. RELATED WORK

Based on previous research, many frameworks have been studied to describe the design of multimodal interfaces in the field of augmented reality and adaptive interfaces. The main goal of providing a framework is as a guide for the interface designers to design their desired applications. Each of the framework's architecture usually consists of modules or components which wires together. For example, input modalities recognizer is a module in charge of processing different types of input (e.g. gesture, voice, touch) received by the user. In this section, we briefly explained each of the proposed framework's component.

A. A Framework of Multimodal System

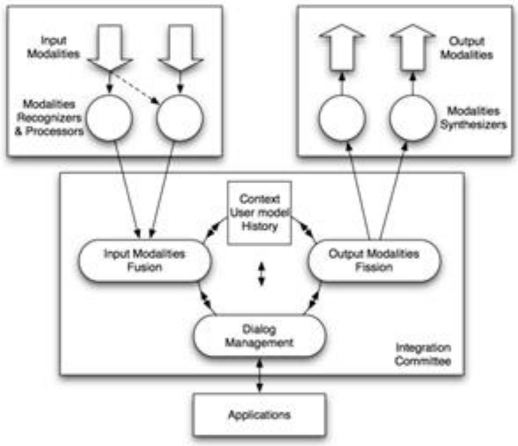

Figure 1: Architecture of multimodal system [15]

Dumas et.al [15] described machine side of multimodal interaction along with the software components for the multimodal system [15]. This framework (Figure 1) illustrates the processing flow between components (modalities fusion and fission, a dialog manager and context manager), the input and output modalities, and the client applications. Modalities Recognizer first perceived the input modalities and the results are passed to the fusion engine. When fusion engine interprets the modalities, it communicates with the dialog manager which is responsible to identify the dialog state, the transition to perform, the action to communicate to a given application, and then the message is returned through the modalities fission. The fission engine is finally responsible to return a message to the user through the most adequate combination of modalities, depending on the user profile and context of use. For this reason, the context manager is working on tracking the context, location and user profile, closely communicates any changes in the environment to the three other components so that they can adapt their interpretations [15]. This architecture is generally referred by other researchers which used to describe a multimodal interface system. However, this framework was not focused on augmented reality environment and adaptive interface.

B. Framework for Adaptive Multimodal Environment (FAME)

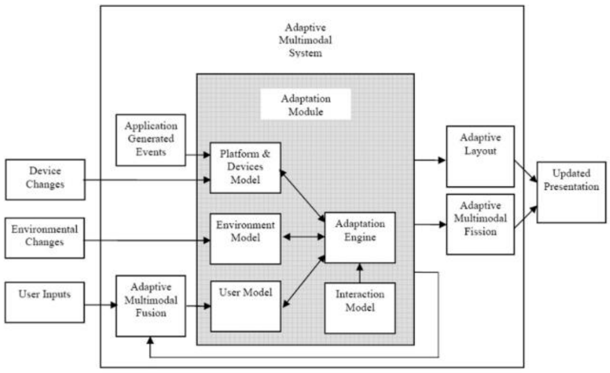

Figure 2: Framework for Adaptive Multimodal Environment (FAME) architecture [7]

Duarte et al. [7] have presented a conceptual framework called FAME (Figure 2) for designing adaptive multimodal applications. This framework illustrates the two levels of architecture which is an inner level for adaptation module and outer level for the multimodal application layer. FAME's adaptation is based on three different classes of input which is user actions, environmental and context changes and device changes. It is the same as Cameleon framework models (user, platform and environment) [16] with an additional model called interaction model. In adaptation module level, each of the modules has its own responsibility: user model is responsible to store a relevant user preferences and characteristics, platform & devices model is responsible to describe the characteristics of the execution platform and of the devices attached to it, environment model is responsible for describing the environmental characteristics that can have an impact on the presentation and interaction aspect of the application, interaction model is responsible for describing the components available for presentation and interactions. They introduced a behavioural matrix and applied this framework in a PC-based multimodal rich book player (DTB Player) application. The adaptation of output modalities is allowed by the application. Therefore, from our point of view, this framework was designed for an adaptive interface which is focused on input and output modalities for PC-based application and not in augmented reality environment.

C. *Framework for Mobile Multimodal Interaction: Top-Level Application*

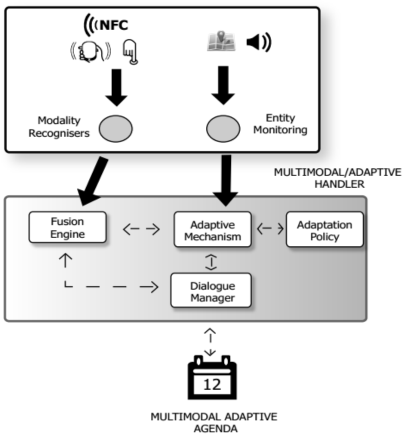

Figure 3: Mobile Multimodal Interaction: Top level Architecture [12]

Maria Solorzano [12] has designed the framework for the top-level architecture of mobile multimodal interaction (Figure 3). Multimodal and adaptive components were received from two external sources of information. The first comes from the events triggered by the user when using the supported input modalities. These events are recognized and processed by the Modality Recognizers component. The second source of information comes from the environment for instance in terms of noise or a user's location. This information is constantly tracked by the Entity Monitoring component. Then, this semi-processed information is sent to the Multimodal and Adaptive Handler. The Fusion Engine, Adaptive Mechanism, Adaptation Policy and Dialog Manager constitute this component and are responsible for each of their tasks. This framework was beneficial for developing an adaptive multimodal application based on two external sources of information but not focused on augmented reality environment.

D. *Human-Centric Adaptive Multimodal Interface Framework*

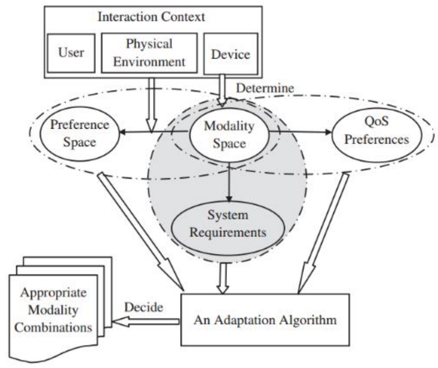

Figure 4: Human-Centric Adaptive Multimodal Interface Framework [6]

Kong et. al [6] proposed a framework based on human-centric adaptation (Figure 4). This paper quantifies the average user preference of a modality under an interaction context. For instance, a noisy environment can reduce a user's preference score of modalities related to the speech recognition. Hence, adaptation can be seen as searching for an optimal set of modalities with the highest preference score for a given scenario. The adaptation algorithm also verified that the selected modality does not exceed the system resource capacities. The adaptation algorithm is fired based on changes in the interaction context which encompasses user, device and environment properties. According to the framework, the application design and development process can be summarized in three steps: The first step is determining the tasks and available input/output for a device type. Then, the interaction scenarios should be determined as well as the interaction contexts. Last, but not least, the designers will evaluate the average preference score of a modality under an interaction context. To obtain this value, a survey with end users have been conducted. The results of the survey are used as inputs for the heuristic algorithm. In conclusion, this framework well explained the adaptive multimodal interface based on human-centric factor but not focusing on AR environment.

E. *Augmented Reality Multimodal Interfaces Framework*

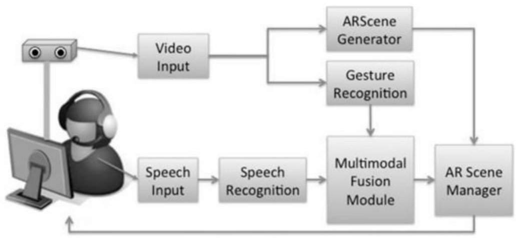

Figure 5: Architecture of Augmented Reality Multimodal Interfaces [8]

Lee et.al [8] proposed a framework of AR Multimodal Interface system (Figure 5). The components include video input for capturing gestures and speech input which recognized by the speech recognizer, Fusion Module, AR scene generation, and AR Scene Manager Module for producing the AR output and give the feedback to the user. For 3D hand gesture, they applied three gestures which are open hand (to select and drop object), close hand (to grab a virtual object) and pointing (to identify where user pointing in 3D space). They recognize all of the gestures by considering the number of fingertips. For speech interface, the user will call the colour shape of the augmented object, moving command and forward & backward command. Multimodal fusion will fuse the arrival gesture and speech input by examining the time difference then modify the AR scene [8]. This framework is a monitor-based multimodal AR with an external camera device (gestures) and headphone (speech) but it was not focused on adaptive interfaces.

F. *A Framework for M3I: Mobile Multimodal Interaction*

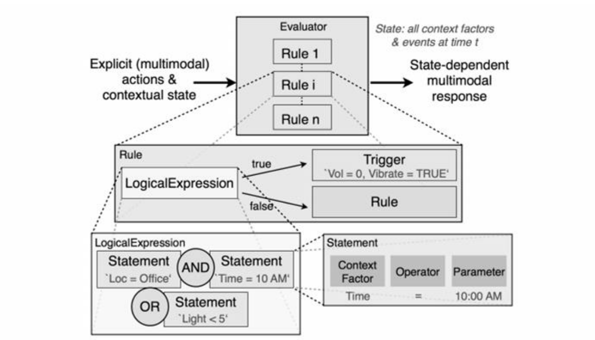

Figure 6: M3I: A Framework for Mobile Multimodal Interaction [11]

M3I framework is an extensive multimodal interaction framework for mobile devices. M3I, a rule-based framework was proposed to simplify and accelerate the creation of multimodal applications for prototyping and research (Figure 6). Explicit interaction is supported along with implicit (context-driven behaviour). The framework integrates more than 50 context factors regarding device location, ambient noise and light level, device orientation, battery information, proximity information (by using Bluetooth, NFC, or Geofence entering/leaving), availability of 3G and WLAN connections, and also for date and time. Basic activity recognition and classification routines abstracting from pure sensor readings are already integrated for example pose classification (either in a pocket or carried in hand), usage indicators, mode of transportation, vision-based detection (face) and etc. For explicit interactions such as button presses or touch, it can be intercepted and combined with implicit contextual information [11]. Hence, we can conclude that this research has proposed a framework to focus on implicit and explicit interaction by using diverse sensors available in a mobile device which allows developers to integrate various types of modalities. The researchers are focused on an adaptive interface in mobile devices but not in augmented reality area.

G. *Component-Based Framework for Outdoor Augmented Reality*

Lee et.al [21] designed a framework for outdoor augmented reality application (Figure 7). This framework consists of two main software packages, one for mobile AR application development (client) and one for building server for the application (server). Data communication layer supports interoperability between these two packages but the client package can be used for developing offline mobile AR applications with local data storage. Outdoor AR software library comprised of two layers of the component which is a lower layer (functional component layer) and service component layer. The functional component layer is vital for developing outdoor AR application. It includes scene data structure, tracking sensor, 3D graphics rendering, 3D sound rendering, and tool for UI elements, data loading and file parsing.

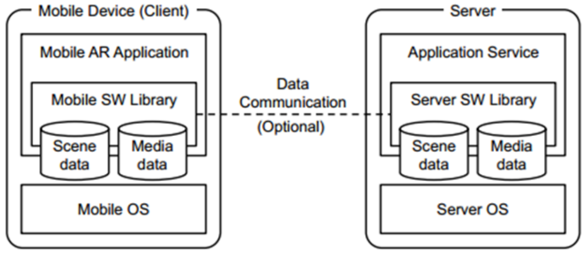

Figure 7: Outdoor augmented reality Framework [21]

Service component layer is a higher-level abstraction of user experience services commonly used in outdoor AR applications. There are three different styles of presentation of geo-located information which is an AR view, a map view, and a list view. Model-View-Controller (MVC) design pattern was used by the components in this layer. For the server, the main function is to provide an online data of scene information and media assets. They were designed as a standard web service which provides AR scene data to mobile and web client.

Therefore, we conclude that this framework was illustrated for mobile outdoor AR application without focusing on adaptive interfaces and any sorts of interaction.

III. Methodology

Based on previous research, we observed which kind of user input and context factors that can support the design of a conceptual framework for adaptive multimodal interfaces in location-based augmented reality application. We are focusing on analyzing suitable multimodalities which can make user naturally interact with the systems. We also focus on the sense of interactions for a mobile phone device. Wearable technologies are out of research scope in this paper.

Among the frameworks proposed by previous research, most frameworks were not focusing on the study of Adaptive Multimodal Interfaces in Location-Based Augmented Reality Application. For example, Kong et.al [6] proposed a framework which described an adaptation of multimodal interface based on user preferences input, QoS requirement and interaction context but not focused on augmented reality application. Duarte et.al [7], Moller et.al [4] and Maria Solorzano [12] illustrate a framework for an adaptive multimodal interface which uses context information (e.g. user behaviour, environment resource and device information) to improve user task and system presentation. They have also not implemented the framework in augmented reality application. In contrast, Lee et.al [8] proposed a framework for the multimodal interface in augmented reality with combined gesture and speech recognition to implement their task and measure the most appropriate modalities (unimodal vs multimodalities) but they are not focusing on adaptive interface and focus on wearable computing. Table 1 shows the comparison of different types of framework.

Table 1
Comparison of Different Types of Framework

Research Paper	Kong et. al. [6]	Duarte et. al. [7]	Moller et. al [11]	Lee et. al [8]	Maria Solorzano [12]	Dumas et. al. [15]	Lee et.al [21]
Multimodal Interfaces	X	x	x	x	x	x	
Augmented Reality				x			x
Adaptive Interfaces	X	x	x		x		
Input Modalities	**Hand gesture, voice, Multi-touch display etc.**	**Speech, Mouse & Keyboard Input**	**Gesture, Touch, Speech**	**Speech, 3D hand gesture**	**NFC Tag, Speech, 2D gesture**	**Speech, Gesture**	**Not Focus**
Output Modalities	Audio, Video, Graphics etc.	Visual Display and sound	Vibration, Sound, Visual	Visual	Not Focus	Not Focus	Not Focus
Task	Social Networking Application	Digital Talking Book (DTB)	3 Application Demo	MMI in Table top environment	Multimodal Adaptive Agenda (MAA)	IM2	CourierAR, CityViewAR, GeoBoids
Application Domain	Service	Entertainment	Education	Education	Service	Service	Service, Education and Entertainment

IV. Analysis of Input Modalities Factor

Most of the previous research presented their interface design architecture within their research domain. Various kind of user inputs, environmental factors and device changes have been taken into account to provide more seamless interaction between human and computer. Multimodal interaction support is valuable in various use cases such as in a car, on the street, in the library, etc.

Recently, location-based augmented reality applications have been explored and many applications are available such as Layar, Junaio and Wikitude. The main features of location-based AR application are to augment virtual objects (i.e. Point of Interest) on top of real environment relative to user's current position and altitude [19]. This technology is beneficial to the user especially tourist to explore interesting places in their surroundings. It is because tourist location-based AR will provide a useful information for the tourist regarding the hotspot places from their current location. Due to the mobility feature of location-based augmented reality, it is a viable business opportunity since users can view information anytime and anywhere [20]. Contextual information can be used to adopt certain settings automatically and the user can naturally interact with the system and can make location-based augmented reality to be more usable and efficient.

As Moller et.al [11] stated that implicit context factor and explicit user action can be fused to perform some task in the multimodal system, we are analyzing several types of sensors available in the mobile device. With emerging features of a smartphone, it provides various input modalities which are recognized by the specific sensor (i.e. accelerometer, proximity sensor). These sensors can be used to perform some system's task which provides an explicit and implicit information. For instance, GPS sensor, accelerometer sensor and compass sensor play a major role in location-based mobile augmented reality which gives an implicit information about the orientation data to the system. Table 2 shows the list of sensors available in the mobile devices which can give good impact to location-based AR system. We classify each of the sensors in two groups which can provide either implicit or explicit information.

Table 2
Type of Mobile Device Sensors

Type of Mobile Sensors	Implicit Information	Explicit Information	Description
Camera		x	Capturing real environment scene
Microphone		x	Speech detector
Accelerometer		x	Determine the current orientation of the device.
GPS	x		Allows the phone to localize itself, enables new location-based applications such as local search, mobile social networks, and navigation
Proximity		x	Measures the proximity of an object in *cm* relative to the view screen of a device.
Gyroscope	x		Works with accelerometer to detect rotation of phone
Light Sensor	x		Measures the ambient light level
Thermometer	x		Measuring ambient temperature

Nowadays, we know that people are always seeking for useful information in their surroundings. Several factors need to be taken into account in order for the user to automatically get a precise data through an application. In location-based AR application, several factors need to be considered for providing a useful information and reducing user's workload while interacting with the system.

H. Mobility

Users are always on the move while exploring their surroundings. Mobility is referred to a technology used in mobile context and in general activities and also allowing interaction anytime and anywhere [20]. It is causing a dynamically changing of the interaction context. For example, when the user is walking on the street with a high level of noise and moving action, it is difficult for users to interact with the system by user gesture and speech modalities.

I. Mobile Context Factor

The requirements for choosing a certain modality depend on the context, for instance, time, location, social setting, or security demands [11]. An example of context-driven modality setting is described when the user wants to seek information about available restaurants nearby their current location in the specific time (implicit information) such as at 2.00 p.m, the location-based AR application will automatically augment the virtual content of POI regarding the restaurants.

J. User Preference

User preference can focus on personalizing the information based on user profile and contextual data. This factor is significantly important where the user can get useful information which fit their needs. For example, location-based AR system gets information of user's current interest such as cultural exploration and shows the augmented hotspot places (POI) of cultural interesting places together with distance data.

By considering these factors, the designers and developers can set rules of adaptation to choose suitable input modalities which are available on mobile devices and provide location-based AR system with the implicit and explicit information. However, an adaptation rules are out of our focus in this study. Therefore, based on the analysis, we proposed a conceptual framework to illustrate the component needed to design an adaptive multimodal interface for location-based augmented reality application.

Our framework will focus on adaptive interfaces, multimodal interfaces and augmented reality. As mentioned in Lee at.al [8], MMI could be used in a wide range of AR applications such as mobile AR interfaces, AR navigation task and AR game applications. Moreover, to decrease the level of mental workload and learning time, an adaptive interface should be included.

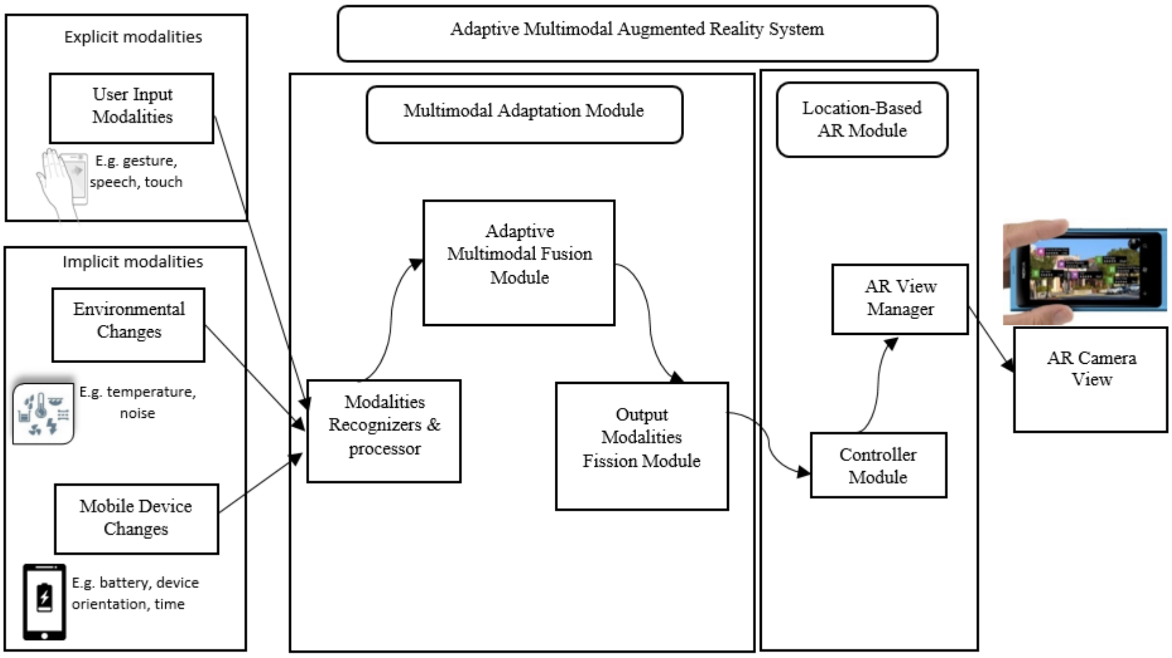

Figure 8: Conceptual framework of adaptive multimodal interface for mobile augmented reality

V. RESULT AND DISCUSSION

A. Proposed Framework for Adaptive Multimodal Interaction in Mobile Augmented Reality

Figure 1 shows our research framework for designing an adaptive multimodal interface in location-based augmented reality. This framework was produced by combining components from previously proposed frameworks in the field of augmented reality, multimodal interface and adaptive interface. It is a base for guiding the development of adaptive multimodal interfaces for location-based augmented reality application. As mentioned before, our focus is on the input modalities. Therefore, we propose an adaptive interface to perceive the input modalities based on the three-interaction context (user input, environmental changes and device changes).

User Inputs are one of the important aspects in Human-Computer Interaction. The user can interact using a hand gesture, touch and speech to give an explicit information to the system. With advanced sensors available in mobile device, the user can naturally interact with the system using different types of modalities. As location-based augmented reality features are usable for the user on the move, the user can interact with the system in an appropriate way. For example, the user can use gesture to point the AR camera around user's location and virtual point of interest will change based on the mobile acceleration.

Environmental Changes are an implicit interaction which used to input a context information about the level of noise, temperature, ambient light of surrounding, etc. For location-based AR, environmental changes can be used to notify the system about what is the current environment and automatically give the impact to the system. For example, if the ambient light is plentiful, the screen's brightness is pumped up so that user can clearly view the augmented POI.

Mobile Changes are also an implicit interaction which can provide context information such as battery capacity, device orientation, time, etc. Therefore, this information can be used to provide the system with the current status of the mobile device. For example, if the mobile device is held in a different orientation (i.e. landscape or portrait), the AR view interface will change accordingly.

Modalities recognizers and processors are responsible for recognizing appropriate input modalities before sending it to the adaptive multimodal fusion module. For mobile devices, each sensor will recognize input modalities. For instance, Gyroscope sensor will recognize the rotation of the phone by the user.

Adaptive MM Fusion module will be in charge of fusing the input modalities and interpret the modalities before sending it to the fission modalities. This module is the most important module responsible for combining one or more modalities (implicit or explicit) to make changes in the system. In this phase, adaptation rules will take place to manage the fused input modalities

The Adaptive MM Fission Module is in charge of sending synthesized adaptive information to the AR Controller module.

AR Controller module is in charge determine the adaptive information provided by adaptive multimodal interaction module.

AR View Manager is responsible to manage the AR content (i.e. POI) to display on the user interface.

Finally, *AR Camera View* will display POIs located inside the current field of view at user's specific positions.

VI. Conclusion and Future Work

An emergent of mobile device recently offers an opportunity to create more powerful augmented reality application where the video stream of the away facing camera creates a live snapshot of the user's surrounding and supplement it with superimposed virtual content in real-time. Location-based AR is one of the most familiar mobile application currently being used. Users can naturally interact with location-based AR system using multimodal interaction through the user interface. Multimodal AR is one of the augmented reality interfaces which indicates the use of more than one modality either simultaneously or sequentially for input and output. Furthermore, an existing sensor can also trigger an interaction between the user and mobile devices (e.g. gestures, speech, and touch). As MMI utilizes the multiple input modalities to enhance the human-computer interaction, they are suitable for adaptive interfaces. Several frameworks have been observed from previous research and we found that little work on producing a framework in the field of adaptive interfaces, MMI and mobile AR has been done. Therefore, we proposed a framework for Adaptive multimodal interaction in location-based augmented reality application together with the related component. For future work, we are going to developing this application by implementing our proposed framework.

Acknowledgment

This research is supported by Ministry of Education Malaysia research grant scheme of TD-2016-003.

References

[1] Azuma R, Baillot Y, Behringer R, Feiner S, Julier S, MacIntyre B, "Recent Advances in Augmented Reality." IEEE Comput Graph Appl. 2001;21(6):34-47

[2] Carmigniani J, Furht B, Anisetti M, Ceravolo P, Damiani E, Ivkovic M, "Augmented reality technologies, systems and applications". Multimedia Tools Appl. 2011;51(1):341-77.

[3] Gaved M, FitzGerald E, Ferguson R, Adams A, Mor Y, Thomas R. "Augmented Reality and Mobile Learning: The State of the Art". Int J Mob Blended Learn. 2013;5(4):43-58.

[4] Jamali S, Shiratuddin MF, "Wong K, "An overview of mobile-augmented reality in higher education. International Journal on Recent Trends In Engineering & Technology". 2014;11(2):229.

[5] Oviatt S, "Advances in Robust Multimodal Interface Design". IEEE Comput Graph Appl. 2003;23(5):62-8.

[6] Kong J, Zhang WY, Yu N, Xia XJ, "Design of human-centric adaptive multimodal interfaces". International Journal of Human-Computer Studies. 2011;69(12):854-69.

[7] Duarte C, Lu, #237, Carri s, #231. "A conceptual framework for developing adaptive multimodal applications". Proceedings of the 11th international conference on Intelligent user interfaces; Sydney, Australia. 1111481: ACM; 2006. p. 132-9.

[8] Lee M, Billinghurst M, Baek W, Green R, Woo W, "A usability study of multimodal input in an augmented reality environment". Virtual Reality. 2013;17(4):293-305..

[9] Bolt RA, "Put-that-there". Voice and gesture at the graphics interface. SIGGRAPH Comput Graph. 1980;14(3):262-70.

[10] Oviatt S, "Ten myths of multimodal interaction". Commun ACM. 1999;42(11):74-81

[11] Moller A. Diewald S. Roalter L. Kranz M, "Supporting Mobile Multimodal Interaction with a Rule-Based Framework". 2014

[12] S. Maria. Master thesis, Vrije University Brussel, Belgium, "Mobile Multimodal Interaction: An Investigation and Implementation of Context-dependent Adaptation". 2012

[13] Rothrock L, Koubek R, Fuchs F, Haas M, Salvendy G, "Review and reappraisal of adaptive interfaces: Toward biologically inspired paradigms". Theoretical Issues in Ergonomics Science. 2002;3(1):47-84.

[14] Wesson JL, Singh A, van Tonder B, "Can Adaptive Interfaces Improve the Usability of Mobile Applications?" In: Forbrig P, Paternó F, Mark Pejtersen A, editors. "Human-Computer Interaction" Second IFIP TC 13 Symposium, HCIS 2010, Held as Part of WCC 2010, Brisbane, Australia, September 20-23, 2010 Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 187-98..

[15] Dumas B, Lalanne D, Oviatt S, "Multimodal Interfaces: A Survey of Principles, Models and Frameworks". In: Lalanne D, Kohlas J, editors. Human Machine Interaction: Research Results of the MMI Program. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. p. 3-26

[16] Calvary G, Coutaz, J, Thevenin, D, Limbourg, Q, Bouillon, L, and Vanderdonckt, J. "A Unifying Reference Framework for Multi-Target User Interfaces". Interacting with Computers 15, 3 (2003), 289–308.

[17] Carmigniani J, Furht B, Anisetti M, Ceravolo P, Damiani E, Ivkovic M. "Augmented reality technologies, systems and applications". Multimedia Tools and Applications. 2011;51(1):341-77.

[18] Grubert J. Grasset R. "Augmented Reality for Android Application Development", PACKT Publishing, 2013.

[19] Geiger P, Schickler M, Pryss R, Schobel J, Reichert M. "Location-based mobile augmented reality applications: Challenges, examples, lessons learned". 2014.

[20] Irshad, Shafaq, and Dayang Rohaya Awang Rambli. "Multi-Layered Mobile Augmented Reality Framework for Positive User Experience." In Proceedings of the 2nd International Conference in HCI and UX Indonesia 2016, 21-26. Jakarta, Indonesia: ACM, 2016.

[21] Lee, Gun A., and Mark Billinghurst. "A Component Based Framework for Mobile Outdoor Ar Applications." In Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, 207-210. Hong Kong, Hong Kong: ACM, 2013.

[22] Ortman, Erik, and Kenneth Swedlund. "Guidelines for User Interactions in Mobile Augmented Reality." 2012.