JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

An Efficient Object Augmentation Scheme for Supporting Pervasiveness in a Mobile Augmented Reality

Sung-Bong Jang* and Young-Woong Ko**

Abstract

Pervasive augmented reality (AR) technology can be used to efficiently search for the required information regarding products in stores through text augmentation in an Internet of Things (IoT) environment. The evolution of context awareness and image processing technologies are the main driving forces that realize this type of AR service. One of the problems to be addressed in the service is that augmented objects are fixed and cannot be replaced efficiently in real time. To address this problem, a real-time mobile AR framework is proposed. In this framework, an optimal object to be augmented is selected based on object similarity comparison, and the augmented objects are efficiently managed using distributed metadata servers to adapt to the user requirements, in a given situation. To evaluate the feasibility of the proposed framework, a prototype system was implemented, and a qualitative evaluation based on questionnaires was conducted. The experimental results show that the proposed framework provides a better user experience than existing features in smartphones, and through fast AR service, the users are able to conveniently obtain additional information on products or objects.

Keywords

Augmented Object Similarity, Context Awareness, Mobile Augmented Reality, Object Augmentation

1. Introduction

One innovative service that is unprecedented in mobile augmented reality (AR) is to provide users with the required information anywhere through an Internet of Things (IoT) network. This type of service can be found in many research studies and products [1]. Enterprises have begun to use these services to increase productivity and effectiveness in the industrial field [2]. Although these services contribute to decreasing time and cost, there are many issues that need to be resolved. One of them is to continuously provide optimal information (or objects) adapted to the users' changing requirements and constraints. Most current AR services provide text or information to users by using simple image overlay technology [3]. When using this scheme, unexpected text or images are displayed on the screens of mobile devices [4]. This happens given that augmented objects are not exactly changed in real time, as there are a number of things to be considered to meet users' current requirements [6]. To address this issue, many technologies have been proposed. One is to provide efficient context awareness to meet users' requirements [5,6]. The context is referred to as environmental information, which is used to characterize a place where smartphone users are located.

Other researchers attempted to address this problem by providing efficient location-based services [7,8].

To date, these solutions have contributed to improving the quality of adaptive and continuous AR experiences in mobile devices. However, there are some limitations to these studies. First, they did not consider a selection method for the optimal object to be augmented to the input video stream. To enable adaptive AR, the required objects should be changed intelligently and augmented on the screen of the mobile device in real time by adapting to user's dynamic context. Second, they did not present a scheme for handling a large number of augmented objects. The number of augmented objects is too large to store them on a single mobile device. To overcome these problems, a real-time mobile AR framework based on object similarity is presented in this study.

The remainder of this paper is organized as follows. Section 2 discusses the design and structure of the proposed method in detail, Section 3 present an evaluation of the results, and Section 4 concludes the study.

2. Proposed Framework

The basic concept behind this approach is illustrated in Fig. 1.
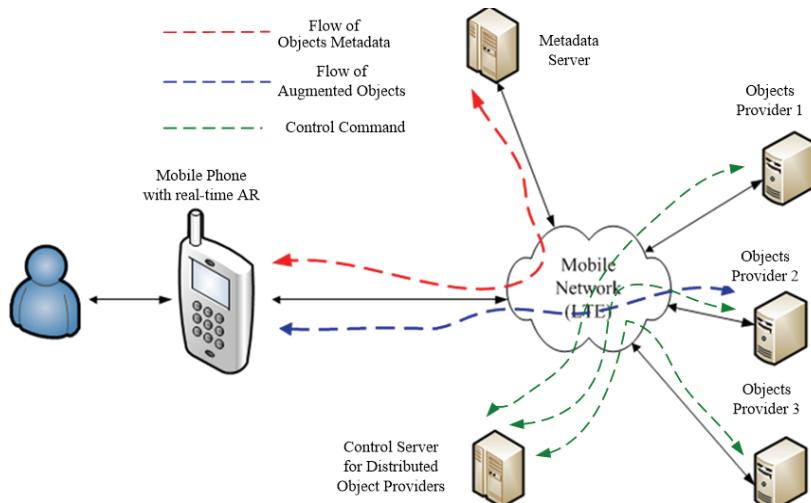
Fig. 1. Service scenario using real-time mobile AR.

The idea is to allow users acquire information using mobile devices equipped with a real-time AR framework. Whenever users capture the image of an object using their mobile camera, the AR subsystem connects to a remote server. This system has three types of servers: a metadata server, an object provider, and a control server. The metadata server contains information on each augmented object, such as its name, location, characteristics, and size. Real objects are saved in a separate server, as illustrated by the object providers in Fig. 1. Object providers can be cloud servers or third-party commercial providers. A third-party commercial provider may provide useful objects through object provision services. The third is a control server that is responsible for controlling the distribution of each object provider. The system

network architecture was extended in our previous work [9,10], by adding a control server for distribution control. To realize this idea, we propose an enhanced framework, as shown in Fig. 2. The proposed framework is composed of six blocks that include object detection, object augmentation, context awareness, rendering, a neural network for context detection, and neural networks for intelligent object selection. The architecture of the proposed system was extended from our previous work [9,10]. The object detection block detects real-world objects from input images, such as faces, soccer balls, hands, automobiles, and books. The object augmentation block performs the augmentation that overlays the images or text on the detected object area of the input images. This must be applied to all consecutive video frames and processed at a user-defined speed.
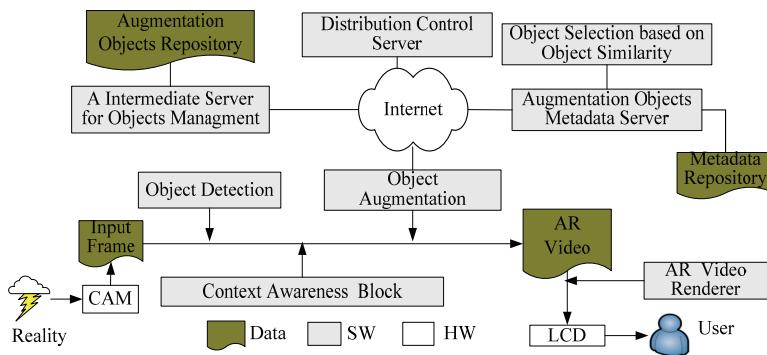

Fig. 2. Proposed framework.

The augmented data are classified into images and text. Images are processed by applying an overlay algorithm of existing image processing techniques. In the case of text, it is simply overlaid on images. The rendering block performs the role of displaying the augmented images on the screen. Finally, the context awareness block automatically searches for the user's location, place, purpose of visit, etc., and transfers that information to the augmentation block. In the architecture, machine learning techniques are used to detect context, and to choose an object intelligently. One of the issues to be resolved in this study is to search for a target augmentation object and to transmit it to mobile devices. Hence, this study presents an effective searching scheme that is based on users' context.

The scheme consists of four steps. First, the objects of interest are identified from the input video in a mobile device. The identified objects are defined using the object's name and the context information. The context information includes the user's location and the building-type being currently visited, such as school, mall, or theater. For example, suppose that a user visits a book store and captures a book on video, in the building of Kyobo. Then, the context information to be used in the augmentation object is {book, Kyobo}. Context awareness is dealt with separately. Thereafter, the name of the identified object and the context information are transmitted to a metadata server outside. The metadata server keeps the metadata of objects to be augmented and chooses an optimal object for augmentation. The server searches for the most suitable object using the received information. The search is performed in two steps. First, the server searches for all the objects, from the augmented object database, that have the same name as the requested object. The resulting objects may be more than one. Second, the server determines an optimal object for augmentation, from the objects found by analyzing the context information and the features of the input object. To select an optimal augmentation object, we propose a choice scheme that

is based on context similarity of the augmentation object. The context information contains environmental information of mobile users. This information should be based on some auxiliary information. The most frequently collected information for context reasoning is based on six principles, including "when", "where", "who", "how", "what", and "why" [11,12]. Here, "when" refers to the visiting time of the users, "where" represents the place of visit, "who" denotes the users' characteristics, "how" represents somehow, "what" represents a domain area, and "why" represents a purpose. From here, we use only "who", "when", and "where" to calculate augmentation context similarity. Let AOC_i represent the received context information for object i and AOC_j represent *the context information of augmentation object j* stored in a metadata server. Then, the context similarity between two augmentation objects can be computed using cosine similarity [13], as shown in Eq. (1).

$$Sim\left(AOC_i, AOC_j\right) = \cos\left(\overrightarrow{AOC_i}, \overrightarrow{AOC_j}\right) = \frac{\overrightarrow{AOC_i} \times \overrightarrow{AOC_j}}{\left\|\overrightarrow{AOC_i}\right\|^2 \times \left\|\overrightarrow{AOC_j}\right\|^2} \tag{1}$$

Here, AOC_i is defined using the following equation:

$$AOC_i = \{Mobile\_User\_Context_i, Location\_Context_i, Found\_Object\_Context_i\} \tag{2}$$

The mobile user context contains information about the user operating a mobile device, which includes gender, age, and job. Each context can be calculated using the following equations:

$$\begin{aligned}
&Mobile\_User\_Context_i = Gender(User_i) + Job(User_i) + Age(User_i), \\
&Location\_Context_i = Country(User_i) + City(User_i) + Longitude(User_i) + \\
&Altitude(User_i), \\
&Found\_Object\_Context_i = Objects\_Features.
\end{aligned} \tag{3}$$

The gender field may have values 1 (woman) or 2 (man), and the real age value will be assigned to the age field. For the job field, the overall domain value of jobs will be assigned. An example of a mobile user context is as follows. The location context contains the location information of a mobile device in the current location of the user. The information includes country, city, GPS coordinates, and type of place visited. The found object context represents information of the found object detected by the camera in the user's mobile device. This context contains information that includes the object's type and features. Object features include the object's type and size. An algorithm for finding an optimal object in a metadata server is as follows:

```
RAOC = Received_Augmentation_Object_Context.
for (i = 0 ; i < End_Of_Objects ; i++ )   {
        if (Sim(AOC(Objecti), RAOC) >= σ) {
            k=i; break;
        }
}
Send Information of Objectk
```

Metadata include a server's location, such as an IP address, where optimal objects are stored. In the fourth step, the user's mobile device obtains an augmented object from the server using the received location information, and it is combined with the input image. The entire process for searching augmentation objects is illustrated in Fig. 3.
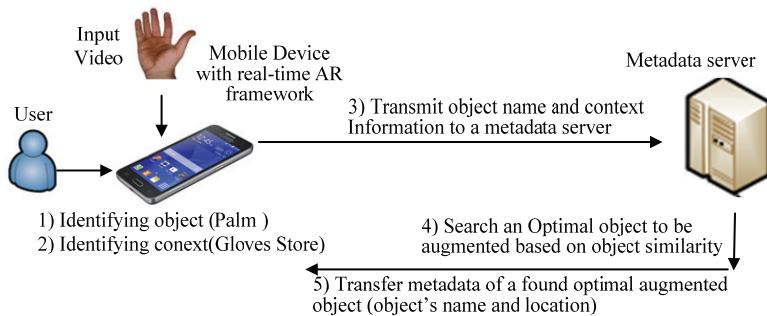
Fig. 3. Optimal object searching procedure.

3. Evaluation

3.1 Implementation Results

To evaluate the feasibility of the proposed approach, a prototype system was implemented. Additional software components were added and placed over the Android operating system. The Android operating system supports an application program interface (API) to implement applications. Table 1 presents a summary of the specifications of the prototype system.

Table 1. Prototype system specifications

Item	Specifications for experiment
CPU	2.0 MHZ Mobile Processor
Display	64.0-in. WQVA (800×480) Liquid Crystal Display (LCD)
Camera	Main-16 Megapixel CMOS
Operating system	Linux 2.9
Mobile platform	Google Android
Image processing software	Intel OpenCV Library

Fig. 4 illustrates the screen of the implementation results. In the screen, the basketball and shoe images are displayed on a camera preview of the smartphone; relevant details such as price, name, and material are displayed on each object. The real-time AR phone detects only basketballs and shoes. In future works, more objects need to be detected, and more functions should be implemented. In this version, small parts of the proposed framework were implemented to merely evaluate its feasibility and effectiveness.

Using the prototype system, a user experience evaluation was conducted by mobile users. Thirteen students, aged 21–30 years, participated in the experiment. A room was designed to resemble a store. In this room, a basketball, a book, and shoes were placed on a desk. Thereafter, two mobile phones were distributed to the students, and they were allowed to enter the room. One mobile phone was not equipped with real-time AR services, while the other phone was. Each subject tried to determine the retail price and to find similar products, by searching the Internet using a non-AR phone. Thereafter, they repeated the exercise using an AR-enabled phone. When a product was captured using the AR-enabled phone, images of similar products and their prices were automatically displayed on the screen. Subsequently, they were asked to answer each question in the test sheet, as displayed in Table 2.

Fig. 4. Screen of the prototype implementation based on the proposed framework.

Table 2. Given questions for user experience evaluation (UX)

Category	Question number	Description
User experience of mobile phone	Q1	Do you feel that the AR-enabled phone is more convenient for obtaining product information compared to a non-AR mobile phone? Answer with "yes" or "no".
	Q2	Do you feel interested when you use the AR-enabled phone? Answer with "yes" or "no".
	Q3	Do you have any intent of using an AR-based mobile phone in the future when you acquire information on a product in stores? Answer with "yes" or "no".
	Q4	Do you feel that the information acquired through the AR-enabled phone is enough to buy a product? Answer with "yes" or "no".
	Q5	Do you think that an AR-enabled phone is an innovative solution for getting information? Answer with "yes" or "no".

The results of the experiment are shown in Table 3. Clearly, majority of the 30 subjects responded with "yes" to questions Q1, Q2, and Q5. For Q3, only 11 subjects responded affirmatively, and 19 subjects responded negatively. This is because the AR-enabled mobile phone used in the experiment was not complete; it was a prototype system. Some subjects highlighted that the AR-enabled mobile phone was interesting, but they would only use it after it was completely developed.

For question Q4, most subjects responded with a "no", except for five students. This implies that more augmented objects need to be included in the provided servers. In reality, enough objects were not accumulated because the evaluation system was only intended for the prototype.

Table 3. User experience test results for the real-time AR-enabled mobile phone

Question number	Results of the qualitative evaluation	
	Answered "Yes"	Answered "No"
Q1	25	5
Q2	27	3
Q3	11	19
Q4	5	25
Q5	28	2

3.2 Qualitative Evaluation

To discuss the feasibility of the proposed approach qualitatively, we compared it with existing AR applications. A comparison between intelligent AR and existing AR is highlighted in Table 4.

Table 4. Comparison between intelligent AR service and existing AR service

Comparison category	Existing AR	Real-time AR service
Augmented object management scheme	Restricted	Metadata-based scheme
Scheme for choosing optimal augmented object	Partially provided	Optimal object selection based on object similarity comparison
Context awareness scheme	Location + user's context	User's context + location + detected object characteristics
Augmented object combination	Image overlay	Dynamic object augmentation
Information delivery	Does not provide	Information delivery based on SIP video call protocol

To provide AR services on a mobile phone, it is necessary to effectively store and search for information or various images to be added to the input images. Existing AR systems typically store augmented objects as a file on a mobile phone and read the file to add it to the corresponding image. This method is suitable when there are few augmented objects; however, it cannot be used if there are a large number of augmented objects. This work proposes a technique that can effectively store and search for a large number of augmentation objects based on a metadata server. Next, to enhance the accuracy of the information provided, the AR system should find an object as close as possible to the one required by the user, from a number of augmentation objects. However, existing AR applications combine input video with stored objects without considering users' intentions and situations. In the proposed scheme, the object to be augmented is intelligently selected by considering object similarity. Furthermore, to provide context awareness, existing AR applications utilize users' information and location. Users' information includes personal information such as age, job, gender, and location, and their GPS coordinates. In our approach, in addition to this information, one more factor is considered that includes a detected object's characteristics. The detected object represents an image or text detected from the input preview through object detection technology. To enhance the accuracy of the sensed context, the characteristics including object's name, type, and features are used. For an object combining scheme, dynamic object augmentation is proposed. In this scheme, a required object is flexibly combined with a detected object, while a stored image is simply overlaid on an input video frame in the existing AR. The process of finding and transferring the desired information during a video call is currently cumbersome and inconvenient because either the current call must be canceled or the screen must be switched to search mode. However, if users use the proposed scheme, they can immediately transfer the necessary information or images to the other party's phone without going through the aforementioned steps.

4. Conclusion

To enable adaptive AR, required objects or information should be intelligently changed and augmented in real time on a preview screen of the mobile device, by adapting to the users' dynamic contexts. To

date, many solutions have been proposed to increase this adaptability by providing efficient context awareness. However, there are limitations to existing AR solutions. First, they do not consider the selection criteria for the optimal object to be augmented on the input video stream. Second, there has been no research on how to efficiently handle a large number of augmented objects. Third, there is no scheme to integrate mobile AR with session initiation protocol SIP-based video telephony. To overcome these limitations, a real-time AR framework was proposed in this work. To evaluate the feasibility of the proposed scheme, a prototype system was implemented on an Android smartphone. Using this system, a qualitative evaluation based on questionnaires was conducted, and a comparison with existing AR solutions was performed. The experimental results show that the proposed framework provides a better user experience than existing smartphones, and users are able to conveniently obtain additional information on products or objects through fast AR services. The future work is as follows. First, more subjects need to participate in a comparison experiment to increase the reliability and validity of the evaluation. Second, an object selection scheme based on machine learning needs to be integrated to improve the accuracy of the information provided.

Acknowledgement

References

[1] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui, "Mobile augmented reality survey: from where we are to where we go," *IEEE Access*, vol. 5, pp. 6917-6950, 2017.

[2] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht, "Towards pervasive augmented reality: context-awareness in augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 6, pp. 1706-1724, 2017.

[3] J. M. Rodrigues, C. M. Ramos, J. A. Pereira, J. D. Sardo, P. J. Cardoso, "Mobile five senses augmented reality system: technology acceptance study," *IEEE Access*, vol. 7, pp. 163022-163033, 2019.

[4] G. Du, B. Zhang, C. Li, and H. Yuan, "A novel natural mobile human-machine interaction method with augmented reality," *IEEE Access*, vol. 7, pp. 154317-154330, 2019.

[5] J. Barreira, M. Bessa, L. Barbosa, and L. Magalhaes, "A context-aware method for authentically simulating outdoors shadows for mobile augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 3, pp. 1223-1231, 2018.

[6] E. S. Goh, M. S. Sunar, and A. W. Ismail, "3D object manipulation techniques in handheld mobile augmented reality interface: a review," *IEEE Access*, vol. 7, pp. 40581-40601, 2019.

[7] J. Wald, K. Tateno, J. Sturm, N. Navab, and F. Tombari, "Real-time fully incremental scene understanding on mobile platforms," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3402-3409, 2018.

[8] P. H. Chiu, P. H. Tseng, and K. T. Feng, "Interactive mobile augmented reality system for image and hand motion tracking," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9995-10009, 2018.

[9] D. S. Kim, "The augmentation objects management in a AR-based mobile video communication," in *Proceedings of the 7th International Workshop on Industrial IT Convergence*, 2017, pp. 152-154.

[10] S. B. Jang, "Object management based on metadata registry for intelligent mobile augmented reality," in *Proceedings of 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Okinawa, Japan, 2019, pp. 572-574.

[11] Y. A. Sekhavat, "Privacy preserving cloth try-on using mobile augmented reality," *IEEE Transactions on Multimedia*, vol. 19, no. 5, pp. 1041-1049, 2016.

[12] R. Shea, D. Fu, A. Sun, C. Cai, X. Ma, X. Fan, W. Gong, and J. Liu, "Location-based augmented reality with pervasive smartphone sensors: inside and beyond Pokemon Go!," *IEEE Access*, vol. 5, pp. 9619-9631, 2017.

[13] P. Fraga-Lamas, T. M. Fernandez-Carames, O. Blanco-Novoa, and M. A. Vilar-Montesinos, "A review on industrial augmented reality systems for the industry 4.0 shipyard," *IEEE Access*, vol. 6, pp. 13358-13375, 2018.

Sung-Bong Jang https://orcid.org/0000-0003-3187-6585
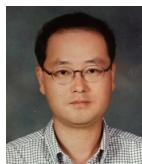
He received his B.S., M.S., and Ph.D. degrees from Korea University, Seoul, Korea, in 1997, 1999, and 2010, respectively. He worked at the Mobile Handset R&D Center, LG Electronics, from 1999 to 2012. Currently, he is an associate professor in the Department of Industry-Academy, Kumoh National Institute of Technology, Korea. His interests include augmented reality, big data privacy, and prediction based on artificial neural networks.

Young-Woong Ko https://orcid.org/0000-0002-6292-0799

He received both his M.S. and Ph.D. degrees in computer science from Korea University, Seoul, Korea, in 1999 and 2003, respectively. He is currently a professor in the Department of Computer Engineering, Hallym University, Korea. His research interests include operating, embedded, and multimedia systems.