3P1-IOS-2a-4

Augmenting Interaction: implementing effective interaction models for AR objects

Svetoslav Dankov Rafal Rzepka Kenji Araki

Graduate School of Information Science and Technology, Hokkaido University

Augmented Reality (AR) applications are becoming more popular with the continued miniaturization of technology. With the increasing use of smart phones, which often provide increased processing power, enhanced and open software platforms, AR has become instrumental in the way we perceive our surroundings and the information that it carries. It is now possible to implement an AR system without carrying bulky and expensive equipment. Currently, there are many systems that implement some form of AR to provide a specialized interaction to users. However, those systems usually employ expensive, immobile components with highly specialized interfaces. In this paper we present a novel approach for building interactive interfaces using AR. UIAR (User Interface through Augmented Reality) is an augmented reality framework that allows for the ubiquitous creation and dissemination of interactive user interfaces. Here we present the novel interaction schemes of UIAR alongside a discussion of their efficacy, usability and performance.

1. Introduction

In the field of collecting common sense knowledge data from a large number of voluntary users, one of the biggest challenges is to keep the users engaged, entertained, and focused so as to collect a sufficiently large amount of high-quality data. Most such projects start with a somewhat big user-base, which unfortunately dwindles in numbers and activity as the project grows older. We believe using UIAR to implement games through its interactive framework can prove beneficial in this area.

In our research, we implement an Augmented Reality system which serves as an extension to existing computer interfaces, provide enhanced user-experience, and define virtual objects and their actions in an ubiquitous way. Augmented Reality is a fairly young area of research which is currently expanding in many of the already existing fields of Human-Computer Interaction, Computer Interfaces, etc.

Augmented Reality and interfaces implemented through this technology are relevant today, and will be even more relevant in the future. Thus we believe that our focus in this direction is beneficial to the proliferation of AR as an interactive medium.

2. Related Research

2.1 Approaches to collecting common sense

With the realization of the importance of common sense to the field of artificial intelligence, considerable research has been done towards collecting and structuring this type of knowledge. The biggest research effort by far has been the Cyc project, which has already collected over a million common sense assertions in little over two decades. As the project became more of a commercial venture, a much smaller set of data is available free of charge. The work required, however, has been considerable. Common sense knowledge is manually input by experts in particular areas, who first give a complete ontological structure to the data, using a specially developed knowledge representation language called CycL, and then insert domain specific data based on their expertise [Lenat 90].

Another attempt to collect common sense data is the Open Mind Common Sense project. OMCS collects common sense statements from untrained volunteers over the Web in the form of natural language statements [Singh 02]. In the course of few years the project had already collected over 1.6 million statements.

Other systems, like Verbosity [Ahn 06] and Common consensus [Lieberman 07] identified and addressed one of the major problems with such systems - user interest. In order to consistently gather quality knowledge from a large set of volunteers, they must be given enough motivation to continue to participate, especially in light of the fact that the number of volunteers drops over the life of the project.

Yen-Ling Kuo et al. [Kuo 09] have successfully utilized social games to collect common sense and their findings provide useful suggestions for designing community-based games.

2.2 Augmented Reality, Interfaces, Frameworks

The technologies for hand and finger-based interfaces can be roughly split in two categories - sensing-based and computer-vision based. Sensing-based systems like [Rekimoto 02] are very robust but are often limited to detecting only ”touch” behavior, not able to recognize hands or other physical object that come into view. Computer-vision based systems like [Do-Lenh 09] are often limited by the lighting conditions and may not respond well to sudden changes in the field of view. However, systems like [Lee 09] have proven to be robust and accurate enough. We are using a computer-vision based system since wearing

Contact: Svetoslav Dankov, Hokkaido University, Graduate School of Information Science and Technology, Kita-ku Kita 14 Nishi 9, 060-0814, +81-11-706-7389, dankov@media.eng.hokudai.ac.jp

special hardware to enable "touch" capability reduces the mobility of the system.
There have been many Augmented Reality applications, using either multiple-camera hand and object tracking or a single camera (like a webcam). Those applications vary in both their mobility and complexity. Our project was inspired for the most part by the Sixth Sense project developed by Pranav Mistry in the MIT Media Lab [Mistry 09]. As is the purpose of [Mistry 09], we strive to provide mobility, affordability and ubiquity to Augmented Reality applications - in other words we strive to augment interactivity through AR.

3. System Overview

Interaction models with AR systems have so far been system/application dependent. Each system defines for itself how users interact with the AR objects and the interaction model cannot be extended or redefined.
Our framework (UIAR) allows developers to define how a specific AR Object will interact with the users and with other AR Objects introduced to the scene. They do so by assigning behaviors to the markers associated with the AR Object via the AR Object's URI. The URI stores an address for the server the AR object is registered to. This allows us to provide ubiquity of object presence, object persistence, and ubiquity of object interactivity; this in turn makes the framework highly adaptable to developer's needs. Our system is based on several existing technologies that allow us to perform AR overlay, QR decoding, marker recognition, tracking and handling and draw our interfaces programmatically.

We designed our AR markers to include QR codes encoding the Unique Resource Identifiers for the object that the AR marker identifies. This allows the developer to define his own AR marker patterns and objects independent of the viewer. It also allows the AR environment viewer to recognize AR markers without the need to include the patterns in the program.
The QR code can be placed either inside of the AR marker as part of the pattern or on the back of the AR marker. Note that if the QR code becomes a part of the AR marker's pattern it must do so in an asymmetrical fashion, since AR marker patterns must be asymmetrical to enable correct marker detection.

We implement two types of object interactivity - object-to-object and user-to-object. The actions are defined on a per-marker basis and are stored server-side.

- **User-Object Interaction**
  User-object interaction is implemented through control-blobs on the markers. The system can detect and distinguish between the blobs. The user activates each blob by covering it with a finger for a predefined amount of time. The action that the blob triggers is stored in the database for each individual marker. Furthermore, the actions are positionally dependent on the marker's surroundings - those actions can be directed/performed to the AR object of a different marker if that marker is close by.

- **Object-Object Interaction**
  Object-object interaction is implemented through positional dependency of the markers in the scene. The system can track the position and distance between markers which allows us to activate specific actions on the AR objects whenever they are in proximity to other objects. By detecting which marker is moving and which is stationary we implement a sender-receiver scheme where the stationary marker "receives" the action, while the moving marker "initiates" it.

4. Common sense acquisition games

We believe that using our framework can prove useful when it comes to both acquiring new volunteers and keeping the existing volunteers interested and engaged in the process. Using the UIAR framework we can implement games that are rewarding, engaging, and interactive. Moreover, those games can be targeted towards younger audiences who naturally spend more time playing games.
With UIAR, developers can assign any virtual object to an AR marker (3D models, textual and media content, etc.) which will be immersive (objects will blend in with the actual environment) and interactive (objects will be aware of the surrounding objects and react to users' input). Here we present 3 implementations of games to collect specific types of common sense inspired by such games as Verbosity [Ahn 06] and Common Consensus [Lieberman 07]. Images for those games are acquired through Google Image Search. We use Wordnet to choose related concepts when needed.

4.1 Visual sentence pattern game

One type of exercise commonly used to collect common sense is to simply fill in the gaps in a sentence pattern. While natural language sentences can often prove difficult to process, the use of different sentence patterns allows for collecting data that can be disambiguated, categorized and easily parsed.
Our example game uses templates such as: "X is a kind of Y", "X is used for Y", "X is typically found near/in/on Y", " X is the opposite of Y", "X is related to Y", etc. A user is provided with AR markers, which after being registered via their QR codes will correspond to X, Y, and a description of the template respectively. As the visual content of AR markers is dependent on the QR code only, the game can be setup so that every time the markers representing X or Y is re-reregistered, the content is changed.
For example, on first registry marker X can hold a 3D model of an apple and Y a 3D model of a tree. If the sentence pattern does not make sense, the user can re-register it until he gets the correct one (in this case, "X is typically found near/in/on Y"). To submit the entry, the user needs to arrange all three markers so that they touch each other. After submission, the game refreshes the markers with new objects and sentence patterns and the user can keep playing.

4.2 Arrange by feature game

The second type of game is for spatially oriented knowledge collection. In this case the number of markers/objects can be as many as the screen can allow. A sample exercise of this game would be to ask the user to arrange the markers in a certain order based on a certain criteria (height, length, size, etc.). As the markers are spatially aware of each other, the user will complete the exercise by putting the markers close to each other in the order needed.

For example, the user can have 4 markers. After registering each marker, he is presented with 4 different 3D (or 2D) objects which he/she will arrange by a predefined feature and submit to the system. The same game can be used to cluster individual objects in case there is more than one feature. For example, the user can be presented with representatives of fruits and animals, in which case he can group the markers together and choose their categorization.

4.3 Arrange in sequence game

The third type of game is goal oriented. Just like the "arrange by feature" game, in this game the user will be presented with two markers representing the beginning and the end of an activity, with the rest of the markers representing actions that must fit in a sequence.

For example, the user can start with 2 markers, one showing - the rising sun (or a person coming out of bed) and another a steaming cup of coffee. The rest of the markers could represent "boiling water", "mixing water in cup", "opening coffee", "pouring sugar". The user will complete the exercise by arranging the markers/activities in the right order.

Each object/activity can be represented either by a visual (a 2D or 3D model) or just text (the text being overlaid over the AR marker). In order to represent both simple physical objects and abstract concepts with more visual appeal it is better to use images or models. The example implementations given are specific to the realm of common sense acquisition. However, the system can be generalized to serve any number of language acquisition tasks.

5. Conclusions and Vision for the Future

In our research we are trying to address the need for enriching textual knowledge with interaction driven knowledge acquisition. We plan to implement an AR viewer for mobile devices using an HTC developer device running Android 2.3 OS. We plan to implement additional methods for user-object interaction, improve the overall usability of the system, and implement security schemes for the AR objects.

We are currently in the process of deploying a prototype version of the UIAR framework as an open source project. We are performing evaluations on the common sense knowledge collection games, the quality of the collected data and how our system affects user retention.

We view Augmented Reality as a technological path that will keep extending and developing with constant future software and hardware improvements. We believe Augmented Reality, as opposed to Virtual Reality and Augmented Virtuality, is the medium that will be most appealing to everyday users. The medium through which we collect, interact and distribute knowledge, will diverge from the nowadays common desktop/laptop/smart-phone solution. Data interaction will become more ubiquitous as the devices through which we perform everyday tasks become themselves more ubiquitous. Even today there are multitude of projects and proof-of-concept products that look into the future of how we interact with machines. They are only one step away from becoming as mass produced and spread as the smart phone has become in the last 5 years. This proliferation of ubiquitous devices, not restricted by common keyboard/display solutions are the reason why research into interactive AR interfaces is so important to us.

References

[Rekimoto 02] Rekimoto, J.: Smartskin: An infrastructure for freehand manipulation on interactive surfaces, SIGCHI-Human Factors in Computing Systems, (2002), pp. 113–120,

[Kuo 09] Kuo, Y., Chiang, K., Chan, C., Lee, J., Wang, R., Shen, E., and Hsu, J.Y.: Community-based Game Design: Experiments on Social Games for Commonsense Data Collection, HCOMP2009-Workshop on Human Computation, (2009)

[Do-Lenh 09] Do-Lenh, S., Kaplan, F., Sharma, A., and Dillenbourg, P.: Multi-Finger Interactions with Papers on Augmented Tabletops, The 3rd International Conference on Tangible and Embedded Interaction, Cambridge, UK, (2009), pp. 16–18,

[Lee 09] Lee, B., Chun, J.: Manipulation of virtual objects in markerless AR system by fingertip tracking and hand gesture recognition, ICCIT'09, (2009),

[Mistry 09] Mistry, P., Maes, P., and Chang, L.: WUW - Wear Ur World - A Wearable Gestural Interface, In the CHI '09 extended abstracts on Human factors in computing systems,(2009),

[Lenat 90] Lenat, D.: Cyc: Towards Programs with common sense, Communications of the ACM,pp. 30–49, (1990),

[Singh 02] Singh, P.: The public acquisition of common sense knowledge, Proceedings of the AAAI Spring Symposium on Acquiring (and Using) Linguistic (and World) Knowledge for Information Access, (2002),

[Ahn 06] Ahn, L. V., Kedia, M., and Blum, M.: Verbosity: a game for collecting common-sense facts, In Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems, volume 1 of Games, (2006), pp. 75–78,

[Lieberman 07] Lieberman, H.: Common Consensus: a web-based game for collecting commonsense goals, IUI'07, (2007)