

Zero-Effort Search and Integration Model for Augmented Web Applications

Ryong Lee and Kazutoshi Sumiya

School of Human Science and Environment, University of Hyogo,
Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan
{leeryong, sumiya}@shse.u-hyogo.ac.jp

Abstract. Due to the rapid advancements of mobile web access environments, there are constantly increasing requirements for web searches of real-world spaces. Nevertheless, search methods available for mobile devices are not much different from those currently available in indoor computing environments. Quite often, it is inconvenient to manipulate such small devices in outdoor environments. In this paper, we propose a zero-effort search model composed of zero-query search and zero-synthesis integration. With zero-query search, there is almost no need for users to express their queries explicitly with cumbersome manipulations. In order to reduce the user's efforts to generate search queries with the least amount of interaction, users' intentions are analyzed from sensing information and map databases. Applying zero-synthesis integration, users can browse an interpolated form, requiring less effort to understand. To realize this convenience, media contents will be displayed in an overlaid form over the real world image. With this proposed intuitive platform model, many various web search and browsing applications will be easily realizable in all the stages from design to actual implementation.

Keywords: Zero-Query Search, Augmented Web Space.

1 Introduction

With the advancement of mobile computing environments, today's smartphones or cell-phones are already web-accessible from anywhere. It is universally possible to get information on the spot through mobile web browsers. However, there are still many difficulties in searching for and integrating various kinds of information from mobile devices in outdoor environments. Even though connectivity to the Web is available with mobile devices, there is still work involved in the physical and mental processes of searching, arranging, and matching the information with real-world objects on a small screen with keypads.

For example, in order to find a restaurant in an unfamiliar downtown area, a few words need to be inputted into a cell-phone to begin the process of investigating the town's information from the Web, using an embedded browser. If the phone is equipped with an apparatus to estimate the physical location of the target, the search guides the user directly to a page including a map showing restaurants near the desired location. Once one of the restaurants is chosen, the best route to the restaurant

needs to be determined. With the help of a map showing the user's current position, the user is able to navigate toward his goal. This is the type of short mission that is often experienced in everyday life in urban areas.

In order to reduce the physical and mental efforts of mobile web searches, we are developing an Augmented Web Search System based on Augmented-Reality (AR) technology. AR systems [1] have so far been introduced in see-through devices, such as the head-mounted display [3], or other camera-embedded mobile gadgets [7, 8, 9]. These devices can concurrently display real and virtual worlds in a mixed form; virtually created graphics or information are superimposed over outer-world images corresponding to real-world objects or spaces on a see-through screen. Those combined worlds are called mixed reality or augmented reality. Such augmentation from virtual existence to the real world can provide many innovative applications from personal activity assistance by wearable computing systems to front-glass based automobile navigation systems [2, 6]. Recently, certain smartphones, such as iPhone [9] and G-Phone [8], have shown one of their possible usages as an AR system. However, AR systems are still in their infancy. Most of the work done has been in relatively closed platforms under different system requirements to visualize graphic contents in order to achieve various purposes in augmented worlds.

We think that these emerging systems are candidates for the next generation of web browsing systems. Future web browsers do not necessarily need to show web contents pages as they currently do. Rather the contents should be displayed adaptively to suit a given user environment. In mobile systems, it would be better to show the web search results in a mixed form with real-world images that have been drawn in most AR systems. It is a new challenge for browser developers to show page contents and real-world space together. In such cases, if a new kind of media format is not available, it is necessary to fragment the page contents, for an enhanced or more natural display of photos or real-time videos. AR developers need to consider the interoperability issue that web browser developers have been pursuing for the last decade.

In order to extend AR functionality onto current web browsers and to make it possible to share content interoperability between heterogeneous AR systems, we are developing an open and interoperable browsing platform for augmented media that can be easily shared and browsed on any augmented system. Before discussing this common media share platform, the common specifications for AR systems should be considered. For constructing media for AR worlds, in Güven et. al [4], a structured document has been proposed, but there is no concrete or standardized format or system for general web search purposes. However, from most research and the fundamental philosophy of the first work by Azuma et al. [1], the common criteria required would at least be that the platform should combine the real and virtual world and be interactive in real-time. In other words, the media should be strongly related to a real-world space, and occasionally it should be able to react to the user's controls on the spot. Based on this simple policy, a media space where every user has a mobile device to search and browse augmented media can be imagined, assuming that current web spaces are specified as the media and a web browser can display both worlds concurrently, as in an AR system. In this paper, this type of media is denoted as Augmented Web Media, the browser as the Augmented Web Browser, and the share environment as the Augmented Web Space as shown in Fig. 1.

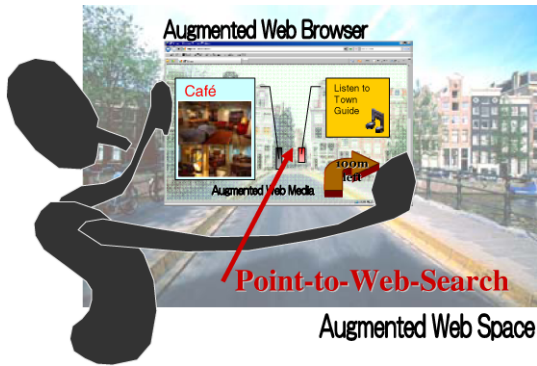


Fig. 1. Concept of Augmented Web Space

With the augmented browser, the search work described above can be greatly improved by reducing unnecessary effort by the user. First, it can be supported by sensing devices and map databases. For example, in order to know a current location, GPS-based latitude and longitude data must be available and used to determine the place name using map databases. Furthermore, if the name of a building in a distant location is required, simply pointing to the geographic object would suffice (the direction data is usable by digital compass and the system searches for the visible object using the map and returns the object name.) In a strict sense, it is not a perfect zero-query. However, this kind of human interaction can reduce efforts to generate queries. It is much easier to just point to target objects than to manipulate the small keypads. The user's effort and time are nearly reduced to zero. For this reason, such a query generation function is called zero-query search model or platform. Although the use of keypads is still required to express users' queries precisely, the targeted goal is to support the query generation by making the most of available context information about the users and their surrounding environments.

The zero-query search model consists of three major components: (1) user motion recognition, (2) environment sensing, and (3) query resolution. In this paper, the main focus is on the 'pointing' action from various human gestures and motions. Especially, the authors investigate which geographic objects are pointed at, and how they can be represented as the next step in high-level queries. We also describe the methodology used to extract visible geographic objects using a digital map rather than simply selecting the nearest geographic objects, which is the most-used method by current location-based systems. Furthermore, with an overlaid display on real-time videos with attached contents, users can understand the captured objects better, with less matching efforts, compared to current browsing methods in normal web browsers. Hence, this function is called zero-synthesis integration support. In order to exactly display searched results or contents from the Web on the screen, the corresponding region on the displayed image should be known. The support of these two zero-effort functions would accelerate application development work for augmented web applications, and would enable performance on a high abstract level for all designs and implements.

In the remainder of this paper, Section 2 introduces emerging augmented web application scenarios and describes a layer-based media integration platform to integrate the web browser and AR systems. Section 3 describes the zero-query search platform using spatial sensing devices and map databases with user-intention reasoning. Section 4 explains the zero-synthesis integration platform for mapping searched results to correspondent regions on displayed images. Lastly, Section 5 concludes the paper with future work.

2 Augmented Web Applications

The advantage to using the Web as a fundamental platform for AR systems is twofold: first, the current open Web platform can be held and its contents managed on the Web servers; second, anyone can participate in the contents service. In other word, there is no restriction to creating and publishing augmented contents. In this section, we describe emerging scenarios by an augmented web browser based on an open media platform for media sharing in real-world space. The assumed media has various types: it can be multimedia, as in photo, audio, or video files; or it can be for applications such as navigation or games with interactive 3D characters. Each media is managed and accessed by a unit of layer. In this study, it is called augmented layer. As in the concept of layers used for map data management, augmented layers can be overlaid to concurrently show a multiple number of layers. For example, a layer showing blogs about a shop can be displayed with another advertisement layer. It is possible to make every augmented layer over a structured file and stored in a uniquely identified location on the Web.

2.1 Layer-Based Augmented Media

Unlike the current web browsers, augmented web browsers need to display web contents over real-time video images. Thus, in order to harmonize both images on a screen, it is necessary to superimpose web contents over a video image, as shown in Fig. 2. This means that the web page cannot be displayed in the author's intended original form. Furthermore, just listing search results in rows is not an effective way to display them. To solve these tangled requirements, we propose a layer-based media that is extracted from searched pages and is able to display the information, together with other results.

The layer-based media we present is quite different from the HTML-based document that is the basic unit for current web browsers. The media included in the layer is somewhat atomic rather than a document. It can be only a photo or audio; but sometimes it also consists of a complex document having these primitive media, which can be drawn on real-world video images. Compared to recent web browsers, it seems that layer-based media shows multiple HTML documents on one screen. This augmentation might reduce the backward or forward operations or the efforts to compare or integrate multiple numbers of contents together. The alignment and accordance of the layers should be considered. For major media formats, the following layer usage scenarios are possible:

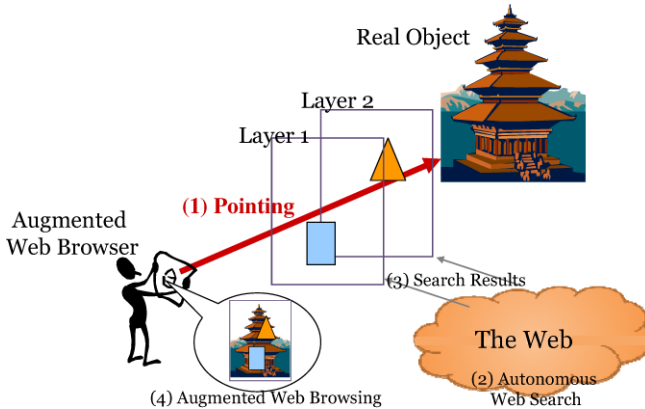


Fig. 2. Layer-based Media for Augmented Web Browsing

- *Photo Layer*: A virtual advertising poster can be attached to a geographical object. While it is accessible only through an augmented media player, these new kinds of media can be beneficial to both advertisers and potential customers. Advertisers can easily and frequently update their content, whereas customers have access to the most recent information.
- *Audio Layer*: A voice can be recorded in a physical space. In [5], they introduce an interesting scenario, using voice guides to warn people of poisonous plant hazards in an unfamiliar forest area.
- *Blog Layer*: With primitive media formats, complex formats, such as blogs, are possible on augmented web space. A blog page, including the primitive media and an opinion article, can be attached to a geographic object. For example, users may find an appraisal on the front door of a shop, attached by another prior visitor.
- *Game Layer*: Interactive applications, such as games, are also realizable in real-world space [10]. For interactivity, operational codes and stable connections are required for synchronization. It would surely require to recognize human behaviors and to give feedback to have people indulge in the best reality in the mixed world.

These are only a few examples of the wide range of possibilities in media applications. As described above, media searched from the Web are able to be concurrently displayed using the layer concept. It surely provides better understanding with less effort, especially for users who have little time to search the Web while engaged in outdoor activities.

3 Zero-Query Search Platform

As described in Section 1, various human gestures and motions are regarded as a query for our augmented web browser, especially, ‘pointing.’ For each pointing action, it is necessary to find what geographic objects can be pointed at with the help of sensing devices and a map database.

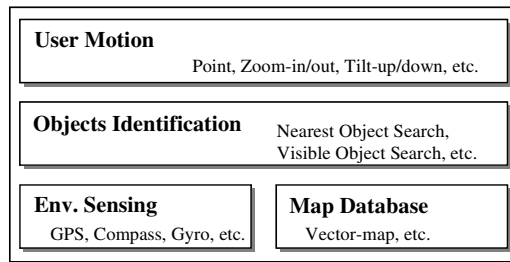


Fig. 3. Zero-Query Generation Platform

The zero-query search platform that we designed and actually implemented is composed of four critical components in three layers, as shown in Fig. 3. In the figure, users' motions previously registered in 'User Motion' are activated by monitoring critical changes of sensing data. For example, if a 'pointing' action is registered, the 'Env. Sensing' component starts to monitor 'direction,' actually azimuth, to determine which direction the augmented browser device is facing (here, a two-dimensional direction is considered, ignoring tilting or rolling of the device.). As the direction is received in real time, the positioning sensor is also activated to determine the current location.

These two sensed data are now converted to an explicit pointed region in real space. Furthermore, the 'Object Identification' layer translates the region query into actual geographic object names with the support of 'Map Database,' for the purpose of searching for the relevant information from the conventional web search sites.

In this section, the process performed by 'Object Identification' is explained; that is, how the user intention is established from the pointing motion with the sensing data and map database. For practical use of the query generation function, it should be performed in real time. In other words, if a user changes his direction, the system needs to detect it as soon as possible to respond in real time.

3.1 Reasoning User Intention from Pointing Action

In order to understand the difference between using location only and using location and direction in AR systems, a representative example is shown. To simplify the discussion, let us assume that we want to automatically tag photographs in the context of location. In order to specify spatial information in a photograph, additional sensors are necessary, since camera devices generally take photographs only in the visual domain.

To accomplish this, integration with GPS has been the approach in many applications; however, this primitive integration is very restricted in its representation of geographic areas of photographs. For example, as shown in the upper-right section of Fig. 4, a photograph is taken from *Kobe Harbor-Land Park* in Japan; fortunately, the camera used here was embedded with GPS and could acquire the exact location coordinates. Hence, it was a simple task to orient the photo to the corresponding location point on the map and the user was able to share the photograph with others over the Web through a photo-sharing site, such as Google Earth or Flickr. The appearance of the *Oriental Hotel* in the background of the photo may provide a good resource for a tourist who wants to visit the hotel later; however, a photograph indexed only by the

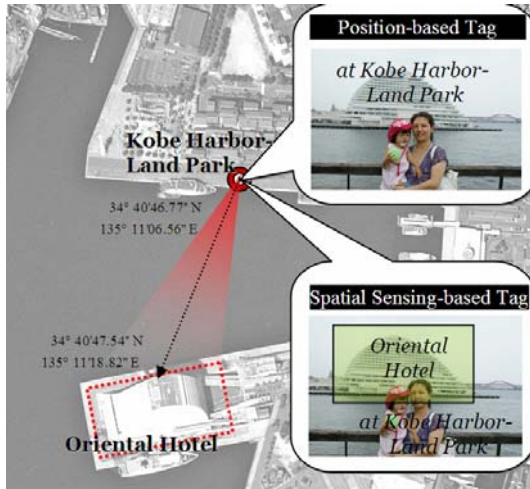


Fig. 4. Photo Tagging based on (1) Location-only and (2) Location and Direction

photo-taking position could not be easily found by simple word-based or map-based photo searching. The major problem is that the photograph was not properly tagged to represent the architecture appearing in the background. If such geographic objects in the content of a photograph are ‘taggable,’ there is a good possibility that object can be searched and the usability of photographs over the Web improved.

In the above example, the greatest limitation in identifying the building is in the difference between the photo-taking location and the location of the hotel. In order to solve the mismatch problem by location-only sensing, we need to apply additional direction information from the photo-taking location to target objects. Direction is usually available using compass and gyro sensors; in 2D real-space, it corresponds to azimuth, and in 3D, it has three attributes—roll, tilt, and yaw. If the photo-taking location and the direction toward the target object are known, it is possible to guess what geographic objects were captured by the camera with a digital map. Furthermore, if other spatial sensing data are available, identification of geographic objects within the camera view range becomes easier, although it would require some additional computational efforts.

4 Augmented Web Browser: Zero-Synthesis Integration

In this section, we describe the later parts of presenting searched results, which augmented browser displays all the searched contents on the real-time video images of the background. To accomplish this, the Zero-Synthesis Integration Model was designed, as shown in Fig. 5. With the zero-query search platform, we could extract the visible object names and their visible angles inside of the camera view angles could be extracted as the set of $\{<name, (SA,EA)>*\}$ (SA: starting angle, EA: ending angle, managed by the ‘Object Mapping Mgt.’ component). Based on real-time

calculation, we calculated which part of the displayed images was correspondent to each pair of the set by the ‘Drawing Object Mgt’ component with the results of {<name, drawing_rectangular region>*}. It is then possible to superimpose the searched results over the correspondent part using the ‘Content Display’ component. Furthermore, each part on the screen is allocated to a user operation, so that selecting or touching it makes it possible to go for further detailed information (by the ‘User Interaction Mgt.’ component). In practice, these requirements would be a general interface for construction of augmented web applications. The superiority of current web browsers can be adopted for navigating other pages by following links, with the support of user interaction.

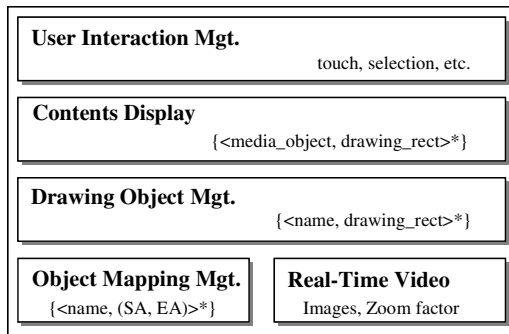


Fig. 5. Zero-synthesis Integration Platform

4.1 Mapping to Display Region

To draw searched results onto the correspondent position of the real-time video image plane, the system uses the angular rate of each object compared to the whole-view angle

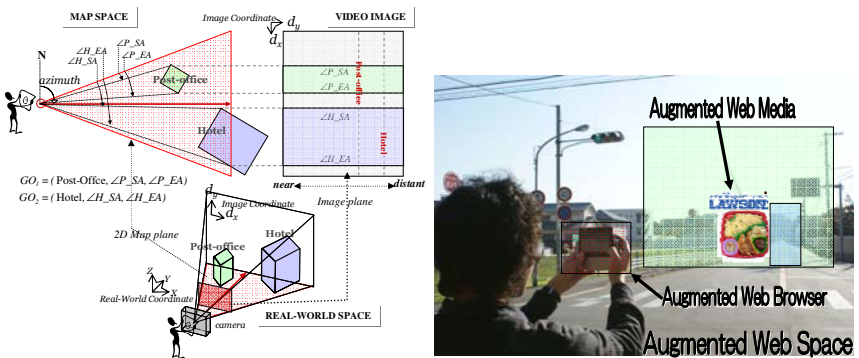


Fig. 6. Mapping geographic object names to the image plane

of the camera to draw position in x_way , and the distance from the $view_point$ to the nearest point of each object to draw position in y_way as follows and shown in Fig. 6:

$$drawing_position_{x_way} = width_of_image_plane \times \frac{visual_angle(object_i)}{view_angle}$$

$$drawing_position_{y_way} = height_of_image_plane \times \frac{distance_to_object}{distance_of_query_range}$$

In order to investigate practical problems and deliver new emerging application scenarios, we are developing a new type of web browser. The only action required by the user is to select a layer and to point to a geographic object. Then, it is able to calculate what spatial area is being targeted with the Zero-Query Search Component. In the browser, this mapping should be computed continuously to respond to the user's movements by changing the direction with Zero-Synthesis Integration.

5 Conclusion and Future Work

In this paper, we introduced the development of a future web browser for outdoor activities, where inputting a query and arranging search results are still an encumbrance. In an absolute sense, it is impossible to predict users' general query intentions. However, there are still many contexts in which unnecessary users' efforts can be reduced by the use of users' profiles and surrounding environmental sensing data. In the present study, our contribution is two-fold: First, we have tried to reason the user-pointing action with position and direction sensing on the map and we have discovered what users want to direct with a visible object search. Such identified geographic object names are used for web search queries. Second, synthesizing web search results with real objects often required when we find paths in urban areas is removed by visualizing searched results directly onto the real video images. In future work, we will study a conversion method from the current web contents to the media drawn in our browser, and also study the accordance between conflicting layers having heterogeneous types of media.

Acknowledgments. This research was supported in part by Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1. Azuma, R.T.: A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6(4), 355–385 (1997)
2. Barfield, W., Caudell, T. (eds.): *Fundamentals of Wearable Computers and Augmented Reality*. Lawrence Erlbaum, Mahwah (2001)
3. Feiner, S., MacIntyre, B.T., Holler, A., Webster: A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment. In: *Proc. Int. Symp. on Wearable Computers*, pp. 74–81 (1997)

4. Güven, S., Feiner, S.: Authoring 3D Hypermedia for Wearable Augmented and Virtual Reality. In: Proceedings of the 7th IEEE international Symposium on Wearable Computers. ISWC, October 21 - 23, p. 118. IEEE Computer Society, Los Alamitos (2003)
5. HP MSCAPE: <http://www.hpl.hp.com/mediascapes/>
6. Hu, Z., Uchimura, K., Lu, H.: Fusion of Realities for Vehicle Navigation. In: Int. Symposium on Computer Vision, Object Tracking and Recognition, Beijing (2004)
7. Kähäri, M., Murphy, D.J.: MARA - Sensor Based Augmented Reality System for Mobile Imaging. In: ISMAR 2006 (October 2006)
8. Mobilizy: <http://www.mobilizy.com/wikitude.php>
9. Sekai Camera: http://tonchidot.com/index_info.html
10. Thomas, B., Close, B., Donoghue, J., Squires, J., De Bondi, P., Morris, M., Piekarski, W.: ARQuake: An Outdoor/Indoor Augmented Reality First Person Application. In: 4th International Symposium on Wearable Computers, Atlanta, October 2000, pp. 139–146 (2000)