Augmented Reality Interfaces

Mona Singh and Munindar P. Singh

Abstract

A confluence of technological advances (in handheld and wearable sensing, computing and communications), exploding amounts of information, and user receptiveness is fueling the rapid expansion of augmented reality from a novelty concept to potentially the default interface modality in coming years. This article provides a brief overview of AR from the perspective of its application in natural web interfaces, including a discussion of the key concepts involved and the technical and social challenges that remain.

What is Augmented Reality?

Augmented Reality (AR) user interfaces have grown tremendously in the last few years. What is drawing great interest to AR is not only the fact that AR involves novel or "cool" technologies, but that it promises to help users overcome the information overload brought upon them by the Web. AR helps present information in a succinct manner: the information comes to the user in its "natural" home—where the user can easily benefit from and act on it.

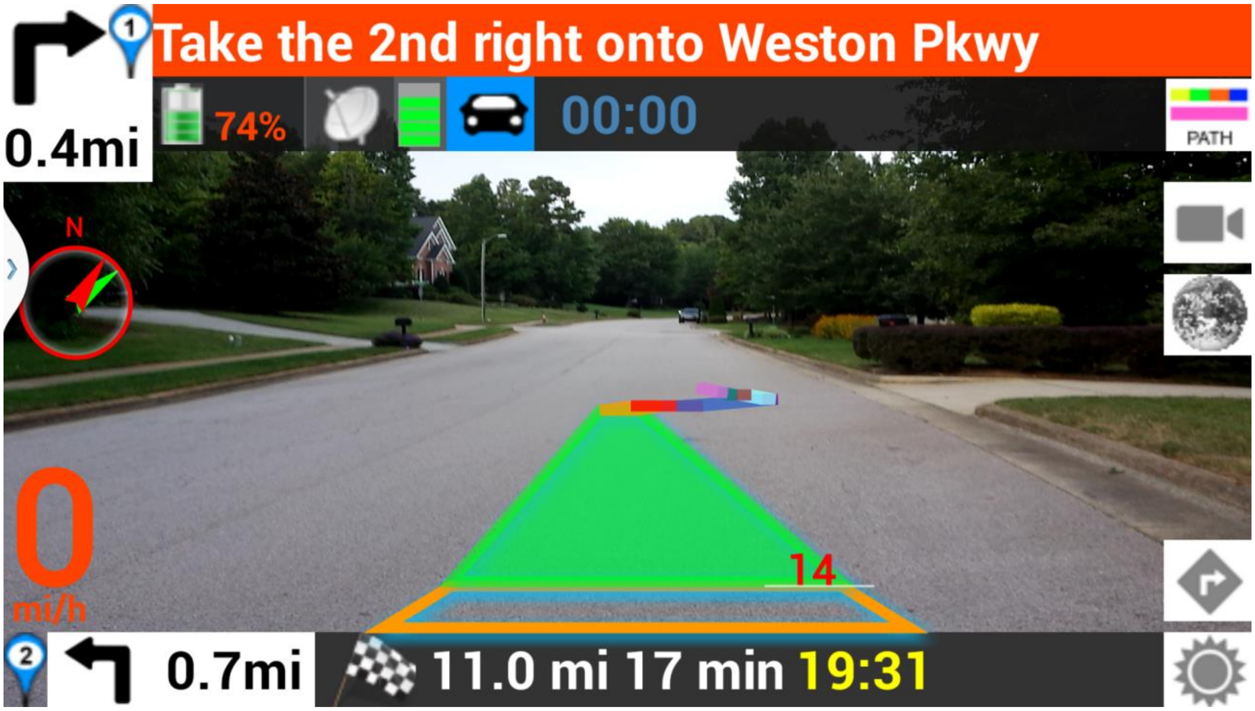
We propose the following definition of augmented reality: AR presents a view of the real, physical world that incorporates additional information that serves to augment the view. Of course, all views of the world are just that—views. An implicit intuition is that the first view is somehow direct or canonical in that it can be treated as reality itself and further augmented with additional information. The augmented "information" is information in the broadest sense and could include nonsense or false information and express any data type (text, image, video, and so on). A baseline example of AR according to the above definition would be a bird's eye view or satellite picture of a city (the "reality") overlaid with street and building names (the "augmentation").

AR is most naturally associated with settings where the aspect of reality considered is proximal to the user and is current; the augmenting information can likewise be proximal and current, or not, depending on the specific setting. Moreover, the most common settings involve visual representation (whether still images or videos), although in principle one might augment any interface modality. For example, an app may play audio signals from the environment along with commentary on the relevant sounds (such as bird calls for ornithologists or various safety warning chimes for training building occupants).

Examples and Nonexamples of Augmented Reality

We confine our attention to the uses of AR in providing natural web interfaces, and especially to phone-based AR, which is becoming widely available. Here are some example AR apps:

- Navigation. The directions a user is taking are highlighted, e.g., stating whether a turn is coming up. In vehicular displays, the appropriate highway lane or next turn may be identified. Figure 1 shows a screenshot of an Android AR navigation app (https://play.google.com/store/apps/details?id=com.w.argps).



Take the 2nd right onto Weston Pkwy

Figure 1: Screenshot from the AR GPS Drive/Walk Navigation app.

- Commerce. A common theme is presenting advertisements according to the user's location or, more specifically, regarding any object recognized in a camera view. The figures below show how the Blippar app (http://blippar.com/) progresses, beginning from the user pointing a phone camera at a grocery item. First, it recognizes the real-world object (bottle). Next, it places an interactive object (recipe book) as an augmentation.

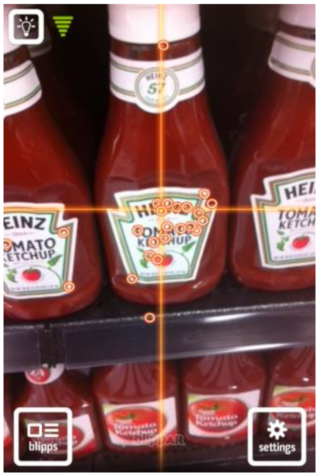


Figure 2: Blippar: identifying a product.

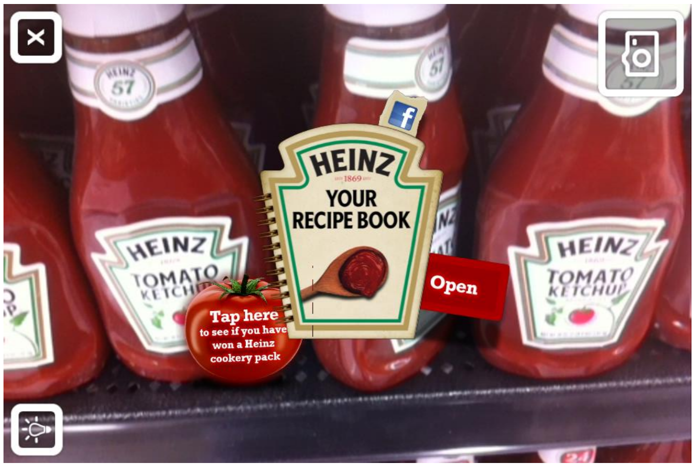


Figure 3: Blippar: augmenting a product with an interactive recipe book.

- Captioning. Generalizing from Blippar, a user would point a phone camera at a scene. The phone would display a real-time image of the scene augmented with metadata associated with scene or its salient parts. For example, a user may point a camera at a remote mountain peak and see its name, height, and current weather. Or, the app may identify landmarks in a city, or provide category descriptions (e.g., “restaurant” or “museum”) of various buildings.

Additional examples involve presenting art, education, gaming, and fashion. An example in fashion is showing how the user would appear when wearing specified apparel.

Although our definition of AR is broad, it *excludes* certain applications even though they may sometimes be described as AR.

- Immersive virtual reality (IVR). AR exposes the real world to a user though with virtual information embedded in it whereas IVR places a user in a virtual world (http://www.kinecthacks.com/augmented-reality-telepresence-via-kinect/).
- Photo editing. An example is Mattel’s “digital mirror” (http://mashable.com/2013/02/11/barbie-makeup-mirror/) wherein a user can edit a picture of herself with lipstick or glitter. Another is the Snaps iPhone app (https://itunes.apple.com/us/app/snaps!/id600868427?mt=8). There is no augmentation of reality in these cases. Were the edited pictures used in place of the original faces in a real scene, we could consider such editing as a form of authoring for an AR app.

- Augmented media. An example is the Guinness Book of World Records providing 3D animations of some world records (http://www.appsplayground.com/apps/2012/09/03/augmented-reality-sharks-star-in-guinness-world-records-2013-app/).

  The distinction between augmented reality and augmented media falls along a continuum. One would imagine pure AR as the augmentation of “natural” reality. However, all too often AR would work only when the reality has been suitably prepped. An example is the Amazon app. Here the user takes a picture of the barcode of a product of interest. The app finds the product on Amazon and presents a user interface for immediate purchase. The app relies upon a media object—a barcode—that would be embedded in the product without regard to AR. Going further, one may affix QR codes on physical artifacts specifically for AR (http://www.npr.org/2013/07/29/206728515/activists-artists-fight-back-against-baltimores-slumlords), in effect, treating the reality as less natural and more symbolic. The extreme form is, as in the Guinness example above, where the user interaction follows purely on the media object and has no bearing on the reality except to access the media object.

Architecture for AR

The figure below shows a conceptual, reference architecture of an AR app, including its essential components and some image-related annotations as examples. (AR could potentially apply to any sense, including audio.) A Reality Sensor (camera) observes a part of the reality. It passes the image it obtains along with metadata such as geolocation tags to the Trigger Matcher. The Trigger Matcher checks if its input matches the relevant app-specific trigger, such as the geolocation being near a specific landmark or the image showing the landmark. It produces matched metadata, such as its semantic category and outline. The Augmentation Selector takes the matched metadata from the Trigger Matcher and retrieves relevant information, such as the year the landmark was built. It constructs an augmenting image, such as a text bubble or a map pin placed relative to the original image, and passes it to the Reality Augmenter. The Reality Augmenter combines the images, potentially as simply as overlaying a map pin on the original image, and causes the combined image to be rendered for the user.

[AR-arch.pdf]

Realizing Augmented Reality

Enabling Technologies

The above architecture highlights the necessary enabling technologies.

First, AR needs suitable sensors in the environment and on the user’s person, including fine-grained geolocation and image recognition in order to obtain a sufficiently accurate representation of the reality.

Second, trigger matching and image augmentation require ways to understand the scene in order to determine the relevant components and display augmentations. These include techniques such as image processing (with face recognition an important subcategory).

Third, trigger matching and subsequent user interaction presume ways to determine the user's attention and immediate context, e.g., via technologies for input modalities including gaze tracking, touch, and gesture and speech recognition.

Fourth, AR presupposes a substantial information infrastructure, e.g., accessible via cloud services, for obtaining pertinent components of the user's longer term context, including intent and activities and determining what components of the real-world to augment, with what, and when.

Additionally, AR requires significant computing and communications infrastructure undergirding the above.

User Platforms

The above technologies are realized on three main types of end-user platforms, each against a backdrop of cloud services. Mobile phones are the most prevalent of these platforms today with vehicles and wearable computers to follow soon. Modern phones include high-quality cameras, geolocation capabilities, numerous other sensors, and sufficient computing and communications capabilities.

A driver in vehicle has a need for accessing information of nearby and upcoming locations. The windshield of a vehicle provides an intuitive venue for rendering augmented information. Vehicles have practically unlimited (electric) power and can support powerful computing and communications.

Wearable computers, of which Google Glass is a well-known example, are becoming viable. Like smart phones and vehicles, wearable computers provide numerous sensors and close access to a user's current environment and the user's immediate context and attention. Wearable sensors, including on the user's skin, clothing, or shoes, offer access to a user's biometric and environmental data and can thus enable smart apps. Today's wearable computers are, however, are limited in power, computing, and communications.

Toward a Taxonomy of AR Apps

The following are the essential ingredients of an AR app: Their possible ranges of varieties suggest a classification of AR apps.

- Trigger. The event or the observation upon which the augmentation occurs. Typical values are location or object recognition (which could occur at multiple levels of granularity, ranging from types of objects to faces of specific people).

A type of a location trigger is matching on GPS coordinates. For example, Nokia City Lens (http://www.1800pocketpc.com/nokia-city-lens-augmented-reality-location-app-for-lumia-devices/) provides information about places of interest nearby. It enables a user to search for restaurants, hotels, and shops, and obtain more information about them.

Blippar (Figures 2 and 3) exemplifies object recognition. Having a phone provide relevant information from a barcode is quite common. The Amazon Mobile app (https://www.amazon.com/gp/anywhere/sms/android) enables users to obtain the product description from Amazon for any UPC symbol captured in a camera. Similar apps are available from Google Shopper and eBay's RedLaser (https://play.google.com/store/apps/details?id=com.ebay.redlaser).

An example of face recognition is Recognizr, a now-defunct augmented ID app (http://www.tat.se/blog/tat-augmented-id/), which identifies a person and displays their online profile and contact details.

- Interactivity. The extent to which the user can interact with the augmented information through the app. In general, in apps where the reality is shown in a direct view there may be occasion for the user to interact only with the augmented information, not the reality.

  An example of no interactivity is road names augmented on a satellite image; an example of medium interactivity is Blippar wherein a user can request a recipe or video by selecting the appropriate marker. BMW Service's app (http://www.bmw.com/com/en/owners/service/augmented_reality_introduction_1.html) exhibits medium interactivity: it displays servicing instructions and advances the instructions whenever its user asks for the next step. An example of high interactivity is advertisement icons that open up automatically to reveal discounts when approached.

- User interface modalities. A user may interact with the augmented information through gesture, gaze, speech, and touch in addition to traditional modalities such as joysticks. Touch and speech are common these days. Google Glass provides a speech interface.
- Naturalness of view. The AR app could be triggered based on natural reality (Recognizr) or require specific features embedded in the environment or physical objects (Amazon).

Opportunities and Prospects

Modeling and applying user context remains the key challenge of realizing high-quality user experience. AR promises a way to present information and support user actions in ways that are sensitive to a user's current context.

Usability Challenges

AR faces the same core usability challenges as traditional interfaces, such as the potential for overwhelming a user with too much information and making it difficult for the user to determine a relevant action. However, AR exacerbates some of these challenges because there may be many kinds of augmentation possible at once and apps that are to be proactive run the risk of overwhelming the user.

Can the user tell the difference between reality and the augmentation? Confusion may lead to user errors by conveying an erroneous impression of the world.

Is the augmentation aligned with reality? Maintaining alignment is nontrivial because reality can change fast, especially in unanticipated ways. For example, in an AR navigation app, the traffic signal may change state or an accident may occur well before the augmented information is updated.

How can a user transition between AR and traditional apps? For example, a user searching for a product may need to move between an AR-enabled app (to identify relevant products) and a traditional app (to search and purchase). However, transitions across apps may be confusing if their underlying metaphors are incompatible.

How should the augmenting information be organized? For example, if a relevant product comes in different varieties, colors, or prices, it would help to group related products in a way that is coherent with the user's intent. An AR app that presents all the information at once may serve only to mislead the user.

Social Challenges

AR is strikingly different from previous computing technologies both in terms of what it accomplishes and in terms of its physical trappings. Just as for other new technologies, it might take years before people begin to widely adopt it except in settings where there is a pressing need or a significant immediate benefit.

Because AR is useful when the augmentations are salient given the user's context, including attributes and prior experiences, the violation of privacy of the user or those present nearby is a potential risk. For example, an advertisement would be most useful if it were for something the user wanted. However, a user upon receiving such an effective advertisement might wonder about how his or her personal information has propagated across the value chain.

Business Models

From the standpoint of business models, we anticipate that AR apps would function like traditional apps in many respects. A key difference would be in terms of who owns, i.e., controls, the AR space. Presumably, the current app (or the entity that controls it) would control the display. For example, instead of advertisements being displayed for keywords as in today's web, in AR, advertisements may be displayed for appropriate triggers, such as particular locations or patterns. However, just that apparently technical change from keywords to locations or patterns may lead to the emergence of new entities in the business ecosystem, such as those who would tackle maintaining the augmented information.

Acknowledgments

Thanks to Pankaj Mehra for comments and to the National Science Foundation for partial support for Munindar Singh under grant 0910868.

Authors

Mona Singh is an independent consultant specializing in user experience and innovative new technologies for user interaction. She has worked extensively on these topics; her previous employers include Microelectronics and Computer Technology Corporation (MCC), Dragon Systems (now part of Nuance), Ericsson, and Teradata. Mona holds a BA in Psychology from Delhi University and a PhD in Linguistics from the University of Texas at Austin. Contact her at mona.singh.n@gmail.com.

Munindar P. Singh is a Professor at NC State University. His research interests include multiagent systems and context-aware computing. He holds a BTech in Computer Science and Engineering from the Indian Institute of Technology, Delhi, and a PhD in Computer Sciences from the University of Texas at Austin. Munindar is an IEEE Fellow. Contact him at m.singh@ieee.org.