AR Glossary

Every domain has specialized terms to express domain-specific meaning and concepts. Many misunderstandings and errors can be attributed to improper use or poorly defined terminology. The Augmented Reality space suffers from this issue due to its immaturity as well as the wide variety of people, from vastly different domains, who are interested in the topic and wish to communicate about their contributions to its advancement.

The goal of the present contribution is a universal glossary of terms with clear definitions which are widely recognized and accepted by practitioners as well as “visitors” to the subject of Augmented Reality. The present document is contributed by its authors to the AR Standards Community as input to the Fifth meeting in Austin, TX on March 19-20, 2012 for the purpose of discussion. This is a work in progress.

We seek to include and expand the glossary to meet the needs of the AR industry and any Standards Development Organizations seeking to work in the area of Augmented Reality. Inputs from the community of AR Standards experts, as well as experts of other fields, are welcome and invited on all aspects including but not limited to the selection of terms to be included, the definitions of terms proposed in the present version.

Terms

- Animated Objects ................................................................................................................. 4
- Auditory AR Experience ........................................................................................................ 4
- Augmentation .......................................................................................................................... 4
- Augmented Reality Application ............................................................................................... 4
- Augmented Reality Browser .................................................................................................... 4
- Augmented Reality Content Management System/Platform .................................................. 4
- Augmented Reality Experience ............................................................................................... 5
- Augmented Reality Marker .................................................................................................... 5
- Authoring ............................................................................................................................... 5
- Behaviour ............................................................................................................................. 5
- Camera View .......................................................................................................................... 6
- Composed Scene .................................................................................................................... 6
- Digital Asset / Digital Object / Virtual Object ......................................................................... 6
- Geo[spatial]-based Augmented Reality ..................................................................................... 6
- Haptic AR Experience ............................................................................................................. 7
- Interaction .............................................................................................................................. 7
- List View ............................................................................................................................... 7
<table>
<thead>
<tr>
<th>Author</th>
<th>Version</th>
<th>Date modified</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ben Butchart, EDINA, University of Edinburgh</td>
<td>1.0</td>
<td>19th December 2011</td>
<td>The terms used are based on documents from AR standards meeting “Notes from June 16th Session 12” (pdf) and AR Vocabulary Spreadsheet. I have some issues with some of the terms myself but would like to elicit comment from others. I have not attempted to give definition of “registration and tracking” but think it would be good to include these terms as they appear often in writing about AR technology.</td>
</tr>
<tr>
<td>Timo Engelke, Dep. Virtual and Augmented Reality, Fraunhofer IGD, Germany</td>
<td>1.1</td>
<td>11.01.2012</td>
<td>Just extended and added some terms.</td>
</tr>
<tr>
<td>Christine Perey, PEREY Research &amp; Consulting</td>
<td>1.2</td>
<td>12.01.2012</td>
<td>Minor grammatical edits. Added terms</td>
</tr>
<tr>
<td>Ben Butchart, EDINA, University of Edinburgh</td>
<td>1.3</td>
<td>20.01.2012</td>
<td>Incorporated new terms and rewriting these to maintain consistent style – hopefully without harm to intended meaning. Added hyperlinks for cross reference. Added some new terms that might be included. Come notes and questions to be considered.</td>
</tr>
<tr>
<td>Christine Perey, PEREY Research &amp; Consulting</td>
<td>1.6</td>
<td>10.02.2012</td>
<td>Alphabetical order, edited AR Browser and AR Application, and AR Marker definitions</td>
</tr>
<tr>
<td>Christine Perey, PEREY Research &amp; Consulting</td>
<td>1.8 and 1.9</td>
<td>13.03.2012</td>
<td>Insert of the introductory information, minor formatting of front matter.</td>
</tr>
<tr>
<td>Sreejumon Purayil, Nokia</td>
<td>2.0</td>
<td>16.03.2012</td>
<td>Fixed Composed Scene definition, added definition of six degrees of freedom</td>
</tr>
<tr>
<td>Ben Butchart, EDINA, University of Edinburgh</td>
<td>2.1</td>
<td>02.05.2012</td>
<td>Changes from workshop added to “Augmentation” and “Composed Scene”. Rolled definition of “Digital Asset”, “Virtual object” and “Digital object” into one. Cascading changes to other definitions.</td>
</tr>
</tbody>
</table>

1 Fourth International AR Standards Meeting, Basel, October 2011-12-19
Animated Objects

**Animated Objects** are digital objects in a composed scene that have been given (by the developer of the object) a dynamic property (changing over time). The behavior (or dynamic property) can begin upon appearance or be the result of user’s interaction with the digital object. There are also digital objects (e.g., virtual humans) that can move within a scene in a fixed trajectory.

Auditory AR Experience

An **Auditory Augmented Reality experience** is the result of a system providing the user digital audio as a result of detecting some trigger in the user’s proximity. The trigger may be visual (computer vision), auditory (natural language), geospatial or detection of other environmental conditions such as radio signals.

Augmentation

An augmentation is a relationship between the real world and a digital asset. The realization of an augmentation is a composed scene.

An augmentation may be formalized through an authoring and publishing process where the relationship between real and virtual is defined and made discoverable.

**Note:** The term “augmentation” is sometimes used as a verb, used loosely to describe the composed scene and the related process of detecting the pose of the real world observer relative to a real world object (cf. registration) to realize a convincing rendering of digital asset in a real world setting.

Augmented Reality Application

An AR application or AR app is any service that provides augmentations to an AR-ready device or system. Sometimes it is useful to make a distinction between a single purpose (or narrowly themed) “app” and an “AR Browser” that can offer a user access to augmentations using the browser provider’s content authoring system. In this distinction, an AR application would normally only consume content from a single source content provider or a restricted set of trusted providers whereas a “browser” would support content from many sources.

Augmented Reality Browser

The term **AR Browser** refers to a class of AR applications that offer a wide variety of AR experiences and themes from more than one content provider. Browser vendors will typically offer a publishing platform and will either host content themselves (in the browser provider’s [content management system]) or offer a mechanism for others to host content that can be served to the browser on demand. At the moment, the distinction between an “AR Browser” and “AR App” is fairly loose, as the industry lacks standards required to implement compliant browser applications.

Augmented Reality Content Management System/Platform

An Augmented Reality Content Management System is a database with well defined interaction types that a content provider can use to produce AR Experiences. Augmented Reality CMS normally
offer data hosting and provide a **Graphical User Interface** with the ability to specify locations for **Points Of Interest** on an map, upload and process **reference images**, add actions to **digital assets** and preview and publish the content. Some CMS offerings can prepare content to publish to several different **AR Browsers**. Sometimes the CMS is provided to support a particular AR Browser or application, allowing content providers to add digital assets and **reality anchors** to a database controlled by the application provider. Occasionally, the term CMS is used to refer to a **Software Development Kit** that enables developers to rapidly create a hosting environment on their own servers.

**Augmented Reality Experience**
An augmented reality experience (also an AR User Experience) is that which is produced as a direct result of combining, in real time, one or more elements of the physical world, one or more **augmentations** and related **user interactions**.

**Augmented Reality Marker**
An augmented reality marker is a 2D (frequently black and white and square in shape) symbol that looks like a 2D barcode and serves as a **trigger** for an **augmentation**. It is defined within an AR **authoring** platform and is unique for each augmentation. There is no defined standard for “AR Marker”, however, many applications with remote or embedded computer vision algorithms are capable of recognizing an AR marker.

**Authoring**
**Authoring** is the process of creating a link (an **augmentation**) between a **digital asset** and the real world. The author must define how a **digital asset** will be rendered and how it is linked to a real world environment. The author can specify a **reference object** in the real world (a **Point of Interest**, a **reference image**, or **marker**) to anchor the **digital asset** in a **composed scene**. Authoring may also involve the specification of **behaviours** that can apply to the **digital asset**. The concept of authoring differs from the term **modelling** which describes the creation of 3D scenes, although 3D modelling (or preview) might be incorporated into an authoring platform. The output of authoring is often some form of **mark up** which provides a structured format for describing the **augmentation**. Authoring also involves the specification of **styling** and formatting options. **Categorization** of **virtual objects** falls into authoring where the purpose is to provide a presentational **filter**. The same categorization of **digital assets** can assist **discovery** and is therefore part of the **publishing** process. In the **publishing** context, categorization is sometimes called **tagging**.

**Behaviour**
A **behaviour** is a feature of a **digital asset** that enables the user to manipulate or visualize the object in a variety of ways. Example behaviours include rotating the virtual objects relative to the users’ point of view, animations, calling a phone number or opening a URL associated with the augmentation in a web browser. A behaviour can be activated by **user interaction** or by sensors. For example, a behaviour could be activated when a GPS sensor detects the user has arrived at a specified location or the camera (light sensor) detects a predefined pattern. Behaviours are often referred to as **scripts**.
Camera View
Camera View is the term used to describe the presentation of information to the user (the augmentation) as an overlay on the camera display.

Composed Scene
A composed scene is produced by a system of sensors, displays and interfaces that creates a perception of reality where augmentations are integrated into the real world.
A composed scene in an augmented reality system is a manifestation of a real world environment and one or more rendered digital assets. It does not necessarily involve 3D objects or even visual rendering.
The acquisition of the user (or device) ‘s current pose is required to align the composed scene to the user’s perspective.
Examples of composed scenes with visual rendering (AR in camera view) include a smartphone application that presents a visualisation through the handheld video display, or a webcam-based system where the real object and augmentation are displayed on a PC monitor.
Note: A composed scene is different from the terminology used in 3D modelling, where the 3D scene simply describes camera angle and lighting but is often divorced entirely from a real world environment.

Digital Asset / Digital Object / Virtual Object
A digital asset is data that is used to augment users’ perception of reality and encompasses various kinds of digital content such as text, image, 3D models, video, audio and haptic surfaces.
A digital asset is part of an augmentation and therefore is rendered in a composed scene.
A digital asset can be scripted with behaviours. These scripts can be integral to the object (for example, a GIF animation) or separate code artefacts (for example, browser mark up).
A digital asset can have styling applied that changes its default appearance or presentation.
Digital assets are sometimes referred to as content, but this is more general in its use whereas digital assets are understood as components in an augmentation.
Note: A digital asset is normally understood as a single entity from the user perspective, even if it is technically composed of several artefacts. So textures, materials and scripts would be bundled together as part of the same object even if they are physically separate files. A digital asset is a broader concept than model as it incorporates a variety of content types - not just 3d models and scenes.

Geo[spatial]-based Augmented Reality
Geo- or location based AR refers to augmented reality experiences based on the user ‘s location and orientation in a geographic coordinate space. Therefore the registration and tracking system relies principally on geo positioning techniques. Most frequently, the user’s position is approximated from the location of the user’s device based on one or more sub systems such as GPS, WiFi or cellular geo positioning. Sometimes the user enters a location manually or scans a LLA marker. The user’s orientation is approximated from the movement of the device using sensors such a digital compass, accelerometer and/or gyroscope. Together with a location fix, the orientation sensors can provide enough information to approximate the user’s 6 degree of freedom pose. Geo-based registration sometimes provides a first approximation for obtaining a user pose which is then refined using
computer vision techniques. Another class of system, called *external tracking systems*, use external cameras to detect and track the position of a user relative to the known camera position.

**Haptic AR Experience**
A Haptic Augmented Reality experience is the result of a system providing the user a vibration, temperature change or introduction of another sign detectable by the user’s sense of touch as a result of detecting some trigger in the user’s proximity. The trigger may be visual (computer vision), auditory (natural language), geospatial or detection of other environmental conditions such as radio signals.

**Interaction**
Defines how users interact with *digital assets*, how *augmentations* are presented to the user, how the user can provide input to an *augmentation*, actions such as search and filtering that the user can perform. *Behaviours* are a subset of user interactions that relate to how the user interacts with *digital assets*. Interactions also describes how *digital assets* react to external events and changing condition in the real world. (i.e. event not initiated by users)

**List View**
List View is the term used to describe the presentation of relevant information to the user in a list organized by alphabetical order, relevance score, date or another filter.

**Longitude Latitude Altitude Marker**
A Longitude, Latitude Altitude Marker is a planar symbol, a particular type of [AR Marker](#), that, when detected by an application which is in communication with a [content management system](#), receives an absolute or relative user (device) position.

**Map View**
Map View is the term used to describe the presentation of relevant information to the user (the *augmentation*) in a geospatial coordinate system depicted on a map.

**Markup**
Markup is the use of encoding of augmentations, triggers and any other information to create the composed scene.

**Point of Interest**
A *Point Of Interest* (POI) provides a geospatial anchor for an *augmentation*. Typically the Point Of Interest specifies geographic coordinate (e.g. WGS84 longitude, latitude, altitude) or a set of points representing area of interest or other *geographic feature*. The Point of Interest links the geographic location to the *augmentation* and is used in an extended sense to include metadata (address, feature type etc.), *styling* and *behaviours*. The term Point of Interest is in common use by geographers to link a spatial geometry to any kind of geo referenced data, not just to augmentations. There is a close similarity in function between a Point Of Interest and a *reality object* in that they both provide the “reality” part of an augmentation. Sometimes the terms *Feature of Interest* or *anchor* are used to capture both geographic and reality objects in a single term. It is not uncommon for developer to use the acronym POI to describe any anchoring of data to a real world,
even if the link is not a geographic feature. In this use **POI** [pronounced “po-y”] is used as a synonym for an augmentation.

**Pose / Six Degrees of Freedom Pose**

A real object in space can have three components of translation - up and down (z), left and right (x) and forward and backward (y) and three components of rotation - Pitch, Roll and Yaw. Hence the real object has six degrees of freedom.

**Provenance**

The **provenance** of a virtual object used in an augmentation is its source prior to the use in an authoring / publishing process or use in an end user AR experience. An augmentation’s provenance can include information about the content creator, the date of the virtual object’s publication or other metadata.

**Publishing**

**Publishing** enables an augmentation to be discovered. This includes the provision of metadata, the formatting of digital assets and the transfer of data to one or more servers to make the link discoverable by search engines, crawlers and AR clients. Publishing is closely linked to authoring and often the two processes are supported by the same content management system. **Authoring** focuses on the creation of augmentations whereas publishing concerns the **discoverability** of augmentations. The term **search** describes the user interfaces and APIs that are used to discover both the augmentation and related metadata. The term **filtering** applies to the presentation and styling of information so is usually related to authoring rather than publishing.

**Real Object / marker / anchor**

A **real object** or a **marker** is a feature or artefact in the real world that is used to **anchor** an augmentation in a composed scene. This includes **natural features** such as buildings and landmarks, artefacts such as posters, book covers or pictures, and **markers** such as barcodes, 2d matrix codes and other machine readable patterns. A distinction (**marker vs. marker-less**) is frequently made between objects that are part of the everyday environment (posters, book covers etc.) and objects that have been created specifically for the purpose of an augmentation (markers, barcodes). In computer vision, the term **feature** is used to describe a pattern that image recognition algorithms can use to identify an object. This **feature extraction** class of algorithms is central to image recognition approaches to **registration and tracking**. Typically a **real object** would have several distinctive features that detection algorithms can identify. A **real object** is sometimes called a **target, anchor** or **trigger** [computer vision]. A real object is usually represented internally in the system as a **reference image** – a digital image with additional metadata or formatting that assists feature extraction.

**Reference image / reference object**

A reference image is a representation of a **real object** (usually an image of the object) that is used by image recognition algorithms to match a frame from the composed scene so that an anchor point for the augmentation can be identified. Often the **authoring** process creates a version of the original image encoded in a format more efficient for the image processing algorithms. In the context of a **visual search**, a reference images refers to one of the images used by the search engine as a **search index**.
Registered Scene
A registered scene is a representation captured after the registration system has detected a real object and is used subsequently as a reference point for tracking. This registered representation is optimized for tracking algorithms and might differ from composed (rendered) scene presented to the user.

Registration / Detection
Originating from computer vision, the term registration (also known as Detection) – describes a system for providing an initial 6 degree of freedom pose relative to a real object or previously registered environment. The pose usually represents the physical location and orientation of the viewing device or its camera relative to a known point or object. The registration system can obtain a pose using either sensor data or and computer vision techniques. In the case of computer vision, the registration system will involve a classification step, where a visual search operation detects a reality object using a set of pre-defined reference objects. In sensor based registration, sometimes called location based registration, the geo positioning sensor is used along with orientation sensors such as a magnetic compass, gyroscope and accelerometer to approximate an initial 6 degree of freedom pose. It is common to combine both vision and location based registration techniques in a single registration system.

Styling
Styling specifies how the digital asset will render in the composed scene. This includes the specification of colours and fonts, specifying the size of symbols, defining what symbols should be used to represent different categories. Styling is an optional part of the process of authoring that overrides the default rendering of an augmentation. Often styling options are incorporated into the same mark up that describes the augmentation. However, a style sheet can be logically and cleanly separated from structural content.

Tracking
Tracking describes a subsystem for providing a 6 degree of freedom pose relative to previously registered real object or previously registered scene. In contrast with registration, tracking uses the previously known pose for generating a new one based upon the frame or location fix that came before. Typically, computer vision algorithms exploit features extracted from the reference object to perform tracking. The system can either use previously extracted features or generate features from the reference object on-the-fly. Where a system’s registration process is too slow to be used in frame-to-frame mode, tracking is unavailable or sporadic. If an algorithm is fast enough to register an object in real time at a reasonable frequency (i.e. 25Hz (reference)) the method usually is called tracking by detection. In the case of location based tracking, the tracking system often obtains a new location fix based on the last, using delta measurements taken from fast, low power sensors.

Trigger
A trigger is the condition(s) that will cause an augmentation to be sent to the user. When a reality object is detected, and the trigger conditions are met, a system pushes one or more augmentations and any associated interactions to the composed scene. In computer vision, the term trigger often refers to the salient attributes of the real world object or marker (or sound) that are necessary to facilitate rapid detection. As a result of a match with a trigger, the digital asset, and any embedded or associated interactions, is rendered by the device output and display system (including visual,
haptic or auditory experience). The term trigger is a common coinage in computer vision and is close in use and meaning to terms such as reality object or real world object, but specifically emphasizes the representations used for computer vision techniques and the actions and behaviours that a visual match will generate.

**User Pose**
The User Pose is the position of the user relative to the target for augmentation and other parts of the real or digital environment. It can be used as an input to registration and also as input to a composed scene.

**User Query**
A User Query is a user-driven (user initiated) request for a digital object or digital asset as part of the AR Experience. It may be communicated by the user via gaze, touch, pointing, speech, text or another input method.

**Visual AR Experience**
A Visual Augmented Reality experience is the result of a system providing the user a digital object displayed in the user’s field of view in response to detection of one or more triggers in the user’s proximity. The trigger may be visual (computer vision), auditory (natural language), geospatial or detection of other environmental conditions such as radio signals.

**Visual Search**
Visual Search involves obtaining information about a real world artefact by submitting a digital image or any subset of the image to a visual search engine. Visual search does not, by itself, constitute AR. It can be very valuable detecting the trigger for an augmentation. Clearly a link is being made between the real world and some digital content but the user experience does not greatly enhance of users’ perceptions of reality (in real time) so it is not clear if a composed scene is part of a visual search experience. Visual search can be used for discovering augmentations, for example, a barcode, 2d matrix code, logo or package design could be used to discover an augmentation applied to a box of cornflakes. In computer vision, Visual Search is typically used for classification of an object or retrieval of information to be used for further registration and tracking.
Domain Model