A typology of augmented reality applications based on their tracking requirements

Jean-Marie Normand
LUNAM Université
Ecole Centrale de Nantes
jean-marie.normand@ec-nantes.fr

Myriam Servières
LUNAM Université
Ecole Centrale de Nantes
myriam.servieres@ec-nantes.fr

Guillaume Moreau
LUNAM Université
Ecole Centrale de Nantes
guillaume.moreau@ec-nantes.fr

ABSTRACT
In this paper we review existing augmented reality (AR) applications taxonomies and we propose ours, which is based on the number of degrees of freedom required for localizing the user, as well as on visualization type, allowing us to cover location-based services as well as more traditional AR applications. Other rendering modalities are also covered by the same degree-of-freedom system.

Author Keywords
augmented reality, taxonomy, degrees of freedom, multimodal

INTRODUCTION
Virtual Reality (VR) has been developing for more than two decades and resulted in numerous conferences and publications in the field [4]. But even though this research field is still very active, VR has faced from the beginning the problem of its relation to the real world which has been formalized by Milgram [9] in the reality-virtuality continuum.

The concepts, functionalities, usages and technologies of VR have been classified many times and we will not propose on a new taxonomy of VR applications, instead we will focus on trying to classify Augmented Reality (AR) applications.

AR is based on techniques developed in VR [1] and interacts not only with a virtual world but has a degree of interdependence with the real world. As stated in [5], “augmenting” reality is meaningless in itself. However, this term makes sense as soon as we refocus on the human being and on his perception of the world. Reality can not be increased but its perceptions can be. We will however keep the term of Augmented Reality even if we understand it as an “increased perception of reality”.

With the term “Mixed Reality”, Milgram [9] groups both AR and Augmented Virtuality (AV). The main difference is that AR implies being immersed in reality and handling or interacting with some virtual “objects”, while AV implies being primarily immersed in a virtual world increased by reality where the user mainly manipulates virtual objects. Nevertheless, the boundary between the two remains tenuous and will depend on applications and usages.

The main challenges of AR consist of the introduction of artificial objects at a location specified in real world coordinates. This requires determining the location of the AR interface in the real world (and not only the user position with respect to the interface as in VR) and including artificial objects in the field of view of the observer.

In the rest of this paper, we will give an overview of existing AR taxonomies, discuss their limitations and propose our own typology.

EXISTING AUGMENTED REALITY TAXONOMIES
Existing taxonomies differ in the criteria they use to classify applications. We chose to divide them into technique-centered, user-centered, information-centered and target of the augmentation taxonomies even if all of existing work will not fit into these four categories.

Technique-centered taxonomies
In [9] the authors propose a technical taxonomy of Mixed Reality techniques by distinguishing the types of visual displays used. They propose three main criteria for the classification: Extent of World Knowledge (EWK), Reproduction Fidelity (RF) and Extent of Presence Metaphor (EPM). EWK represents the amount of information that a MR system knows about the environment. The RF criterion represents the quality with which the virtual environment (in case of AV) or objects (in case of AR) are displayed ranging from wireframe object on a monoscopic display to real-time 3D high fidelity, photo-realistic objects. Finally, the EPM criterion evaluates the extent to which the user feels present within the scene.

In [8], the Reality-Virtuality continuum and some of the elements presented in [9] lay the groundwork for a global taxonomy of mixed reality display integration. The classification is based on three axis: the reality-virtuality continuum, the centricity of the type of display used (egocentric or exocentric) and the congruency of the control-display mapping.

Based on the proposal of a general architecture of an augmented reality system presented in [13], Braz and Pereira [3] developed a web based platform called TARCAST which aimed at listing and characterizing AR systems. It does not propose actual criteria but offers a long list of features for each system, hence is not really discriminative. TARCAST does not seem to be maintained anymore.

Submitted to LAMDa workshop

The technique-centered taxonomies presented here do not take into account any of the mobile AR techniques commonly used nowadays. Milgram's work was innovative at the time it was published and the authors could not predict how mobile AR would arise. Besides, we believe that presence cannot exactly be a common discriminative criterion as it does not refer to the same concept in virtual and real worlds.

User-centered taxonomies

Lindeman and Noma [6] propose to classify AR applications based on where the mixing of the real world and the computer-generated stimuli takes place. They integrate not only the visual sense but all others as well, since their "axis of mixing location" is a continuum that ranges from the physical environment to the human brain. They describe two pathways followed by a real world stimulus on its way to the user: a direct and a mediated one. In the direct case, a real world stimulus interacts through the real environment before reaching a sensory subsystem where it is translated into nerve impulses and finally transmitted to the brain. Those places are called "mixing points". In the case of AR applications, some computer graphics elements can be inserted into this path in order to combine the real world and the computer generated elements into one AR stimulus on its way to the brain. In the mediated case, the real world stimulus travels through the environment, but instead of being sensed by the user, it is captured by a sensing device (e.g. camera, microphone, etc.). Then, the stimulus might be post-processed before being merged with computer generated elements and then displayed to the user at one of the mixing point through appropriate hardware (depending on the sense being stimulated).

Wang and Dunston [14] propose an AR taxonomy based on the groupware concept. They define groupware as: computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment. The goal of groupware is to assist a team of individuals in communicating, collaborating and coordinating their activities. Based on generic groupware concepts, they isolated three main factors for classifying AR systems for construction use: mobility, number of users and space.

Hugues et al. [5] propose a functional taxonomy for AR environments based on the nature of the augmented perception of reality offered by the applications and on the artificiality of the environment. The authors divide augmented perception into five sub-functionalities: augmented documentation, reality with augmented perception or understanding, perceptual association of the real and virtual, behavioural association of the real and virtual, substitution of the real by the virtual or vice versa. The functionality to create an artificial environment is subdivided into three main sub-functionalities: imagine the reality as it could be in the future, imagine the reality as it was in the past and finally, imagine an impossible reality.

While the first axis of the taxonomy proposed by Hugues et al. covers most of the goals of AR applications, the second axis based on the creation of an artificial environment is less convincing since it does not take into account any alteration of the "present" reality. Moreover their taxonomy is limited to vision based approaches and does not handle other modalities. The groupware taxonomy of Wang and Dunston only takes into account collaborative AR and limits itself to construction-based AR applications. Finally, Lindeman and Noma propose an interesting taxonomy based on the integration of the virtual stimuli within multi-modal AR applications. Nevertheless, their proposal might not be discriminative enough, since very different methods like mobile see-through AR can be classified in the same category as a projector-based AR application. Furthermore, it only deals with each sense individually and does not offer any insight on how to merge them together.

Information-centered taxonomies

In [11], Suomela and Lehikoinen propose a taxonomy for visualizing location-based information, i.e. digital data which has a real-world location (e.g. GPS coordinates) that would help developers choosing the correct approach when designing an application. Their classification is based on two main factors that affect the visualization of location-based data: the environment model used (ranging from 0D to 3D) and the viewpoint used (first person or third person perspective to visualize the data). Based on these two criteria, the authors define a model-view number MV(X,Y) that corresponds to a combination of the environment model (X) and the perspective (Y) used. Each MV(X,Y) class offers different benefits and drawbacks and the authors suggest to choose a class depending on the final application targeted, the available hardware or sensors on the targeted devices.

In [12], Tönnis and Plecher divide the presentation space used in AR applications based on six classes of presentation principles: temporality (i.e. continuous or discrete presentation of information in an AR application), dimensionality (2D, 2.5D or 3D information presentation), registration, frame of reference, referencing (distinction between objects that are directly shown, information about the existence of concealed objects) and mounting (differentiates where a virtual object or information is mounted in the real world, e.g. objects can be hand-mounted, head-mounted, connected to another real object or lying in the world, etc.). This work-in-progress taxonomy is currently being tested with nearly 40 publications taken from ISMAR's recent conferences.

Suomela and Lehikoinen propose a taxonomy that can only be applied to location-based applications, hence is oriented towards mobile AR. Moreover they do not tackle multi-modal mobile AR applications. Tönnis and Plecher propose a complete taxonomy but they do not deal with low dimensionality (e.g. 0D in vision) in AR, nor with the multi-modality that can be used in AR applications.

Taxonomy based on the target of augmentations

Mackay [7] proposed a taxonomy which is neither based on the technology used nor on the functionalities or the application domain. The criterion used to classify AR approaches is rather simple: the target of the augmentation. Three main possibilities are listed in the paper: augment the user, when the user wears or carries a device to obtain information about physical objects; augment the physical object, the object is changed by embedding input, output or computational devices on or within it and augment the environment surrounding the user and the object. In the latter case, neither the user nor the object is affected directly, independent devices provide and collect information from the surrounding environment, displaying information onto objects and capturing information about the user's interactions with them. This taxonomy is not very discriminative. For example, one can notice that every single mobile AR technique falls into the first category, while the last category regroups only projection based methods. As in most of the taxonomies presented here, this work does not tackle the multi-modality issue.

PROPOSAL

We now propose our own taxonomy, based on three axis:

- the first axis is based on the tracking degrees of freedom required by the application and the tracking accuracy that is required. Frequency and latency of tracking can also be taken into account.
- the second axis is representing the application type, whether it is merely visualization/navigation or if it can imply interaction with the observer.
- the third axis covers other rendering modalities that go beyond visual augmented reality. It remains rather limited today but it can be taken into account by the same degrees-of-freedom system.

Tracking

The first axis can be divided into 4 classes:

1. *0D* applications: although it is questionable whether these kind of applications can be considered as AR applications, we find in this class applications that detect a marker (such as a QR-code) and display additional information about this marker. For this category of application, the displayed information has no relation with the real world position and orientation of the marker. Tracking accuracy is very limited since it only requires correct marker detection in one frame, indeed, once detected the marker is not tracked in the following frames. As a consequence of this lack of tracking, latency and update rates are no issues.
2. *2D* applications: this is the class for so-called Location-based services, i.e. applications that provide information about a given location, such as nearby restaurants, etc. Tracking accuracy is generally decametric and the tracking method is often an embedded-GPS (altitude information is not used, updates rates around 1Hz). A typical example of a 2D application is a Google Maps[1] like application which only uses a 2D map in order to help the user finding his way in a city.
3. *2D+θ* applications: this class is also for location-based services that include an orientation information which allows to show a relative direction to the user. All navigation systems are based on this principle, accuracy is most often metric. Note that a GPS alone cannot provide an orientation in static position. Orientation can be computed by differences between positions or can be given by a embedded magnetic compass as in modern smartphones. Required accuracy is also metric, update rates typically ranging from 1 to 10Hz. A typical example of a 2D + θ application is the Metro Paris[2] application which helps you locating nearby metro stations and other points of interests (restaurants, bars, etc.).
4. *6D* applications: this last class covers what is traditionally called augmented reality by computer vision scientists who usually work on tracking technologies. Several types of sensors can be used individually or all together (optical cameras, depth cameras, inertial sensors, etc.). Various precision classes exist depending on application types (e.g. marker-based vs. markerless) and on the working volume size (e.g. indoors vs. outdoors) and accuracy is relative to this size. Update rates are much more critical here, a minimum refresh rate would be around 10Hz, and can go up to 100Hz. At this point, continuous tracking must be distinguished from initial localization for which there exists fewer works.

Application type

For this second axis, we distinguish between application types. The first one is dedicated to (active) observation applications. It includes two main categories depending on the used device:

- Optical see-through applications: there are mostly found in head-up displays (HUD) where they are mostly in the 2D+θ class (for HUDs fixed to a vehicle) or in the 6D class where optical information are projected on lenses of see-through glasses (or for worn HUDs). These applications remain lab prototypes (centimetric accuracy) or can be found in the army (fighter pilots helmet based displays) where they are used to display relative position and speed of opponents as well as some navigational aid.
- Video see-trough applications where a device equipped with a back-located camera (such as a tablet or a smartphone) is filming the real environment and the video is reproduced on its display augmented with artificial images. These applications are often called *magic windows* or "video see-through" [9]. The *magic mirror* is a specific case where the camera and the screen point in the same direction.

At last, Spatially Augmented Reality (SAR) [2, 10] consists in adding information to the real world, not simply adding information onto the observer's eye. These applications have a better potential for being multi-user. They are often large scale applications where the projectors usually do not move.

Rendering modalities

Although the visual sense is by far the most important when talking about AR, some work has been carried out in order to mix the real world and computer graphics images across multiple modalities [6]. While the addition of sound in AR applications seems quite straightforward and common, it is

[1] http://www.google.com/mobile/maps/

[2] http://www.metroparisiphone.com

much more unusual to see AR applications that provide with real 3D sound. Haptic feedback integration for augmented reality is also relatively common, especially for medical or training based applications, although, for mobile AR it is difficult to be able to give the user a better haptic feedback than the one provided by a vibrator (e.g. on a mobile phone). Olfactory and gustatory senses are much more rarely used in AR applications.

Nevertheless, we believe that multi-modality should be taken into account in a typology of AR-based applications, and that their integration could also be based on our degrees-of-freedom approach. Indeed, as for sound, we stipulate that a simple monoscopic sound such as a signal represents 0D sound, stereoscopic accounts for 1D (azimuth) and binaural corresponds to location-based sound (distance and azimuth). Hence, our degrees-of-freedom based classification would take into account the audio modality. But it has to be noted that in the presence of moving objects or user, real-time feedback becomes very complex.

As for the haptic modality, we take a similar approach. A simple vibration, (e.g. provided by a mobile phone vibrator) corresponds to 0D while the use of specific devices could account for higher dimensions of the haptic modality.

Concerning the olfactory and the gustatory modalities, we assume that a non-directional stimulus (or at least a stimulus whose origin cannot be determined such as an ambient smell) is also 0D. As gustatory senses are only touch-based sensors, we limit our typology here for them. If a smell direction can be identified, it is only in azimuth and we call it 1D. Other sensors (thermal sensors on the skin for example) available in the human body could also be classified this way. At the moment, it is technically impossible to directly stimulate proprioceptive sensors, they remain absent from our classification.

As mentioned before, the integration of real multi-modal user feedback requires some extra devices that presently prevent them from being used in most mobile AR applications.

CONCLUSION

In this paper, we have briefly surveyed and discussed existing taxonomies of augmented reality applications. We have then proposed ours based on application tracking requirements, application type and rendering modalities. The originality of our proposal is that it merges location-based taxonomies, such as [11], with classical AR vision-based applications into the same classification, broadening the spectrum of applications fitting into a single taxonomy. Moreover, unlike most existing taxonomies, we included multi-modality as a classification criterion although vision remains by far the most important sense.

During the workshop, we will try to demonstrate how augmented reality applications fit into that classification through a graphical presentation and discuss its compared advantages and drawbacks with respect to other existing taxonomies.

REFERENCES

1. Azuma, R. T. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments 6*, 4 (aug 1997), 355–385.
2. Bimber, O., and Raskar, R. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A K Peters/CRC Press, July 2005.
3. Braz, J. M., and Pereira, J. a. M. Tarcast: Taxonomy for augmented reality CASTing with web support. *The International Journal of Virtual Reality 7*, 4 (Dec 2008), 47–56.
4. Fuchs, P., Moreau, G., and Guitton, P., Eds. *Virtual reality: concepts and technologies*. CRC Press, 2011.
5. Hugues, O., Fuchs, P., and Nannipieri, O. New augmented reality taxonomy: Technologies and features of augmented environment. In *Handbook of Augmented Reality*, B. Furht, Ed. Springer, 2011, ch. 2, 47–63.
6. Lindeman, R. W., and Noma, H. A classification scheme for multi-sensory augmented reality. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology*, VRST '07, ACM (New York, NY, USA, 2007), 175–178.
7. Mackay, W. E. Augmented reality: Linking real and virtual worlds: A new paradigm for interacting with computers. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '98, ACM (New York, NY, USA, 1998), 13–21.
8. Milgram, P., and Colquhoun, H. J. A taxonomy of real and virtual world display integration. In *Mixed Reality – Merging Real and Virtual Worlds*, Y. Ohta and H. Tamura, Eds. Ohmsha(Tokyo) and Springer Verlag (Berlin), 1999, ch. 1, 1–16.
9. Milgram, P., and Kishino, F. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems E77-D*, 12 (Dec 1994), 1321–1329.
10. Raskar, R., Welch, G., and Fuchs, H. Spatially augmented reality. In *In First IEEE Workshop on Augmented Reality*, IWAR'98 (1998), 11–20.
11. Suomela, R., and Lehikoinen, J. Taxonomy for visualizing location-based information. *Virtual Reality 8* (September 2004), 71–82.
12. Tönnis, M., and Plecher, D. A. Presentation Principles in Augmented Reality - Classification and Categorization Guidelines. techreport, Technische Universität München, 2011. Version 1.0.
13. Vallino, J. R. *Interactive Augmented Reality*. PhD thesis, Department of Computer Science, University of Rochester, Rochester, NY, USA, 1998.
14. Wang, X., and Dunston, P. S. Groupware concepts for augmented reality mediated human-to-human collaboration. In *Proceedings of the 23rd Joint International Conference on Computing and Decision Making in Civil and Building Engineering* (2006), 1836–1842.