A new typology of augmented reality applications

Jean-Marie Normand
LUNAM Université, École
Centrale Nantes
CERMA UMR 1563. BP92101
44321 Nantes Cedex 3,
France
jean-marie.normand@ec-nantes.fr

Myriam Servières
LUNAM Université, École
Centrale Nantes
CERMA UMR 1563. BP92101
44321 Nantes Cedex 3,
France
myriam.servieres@ec-nantes.fr

Guillaume Moreau
LUNAM Université, École
Centrale Nantes
CERMA UMR 1563. BP92101
44321 Nantes Cedex 3,
France
guillaume.moreau@ec-nantes.fr

ABSTRACT

In recent years Augmented Reality (AR) has become more and more popular, especially since the availability of mobile devices, such as smartphones or tablets, brought AR into our everyday life. Although the AR community has not yet agreed on a formal definition of AR, some work focused on proposing classifications of existing AR methods or applications. Such applications cover a wide variety of technologies, devices and goals, consequently existing taxonomies rely on multiple classification criteria that try to take into account AR applications diversity. In this paper we review existing taxonomies of augmented reality applications and we propose our own, which is based on (1) the number of degrees of freedom required **by the application**, as well as on (2) the visualization mode used, (3) the temporal base of the displayed content and (4) the rendering modalities used in the application. Our taxonomy covers location-based services as well as more traditional vision-based AR applications. Although AR is mainly based on the visual sense, other rendering modalities are also covered by the same degree-of-freedom criterion in our classification.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Artificial, augmented, and virtual realities*

General Terms

Theory

Keywords

Augmented Reality, Taxonomy, Multi-modality, Degrees of Freedom
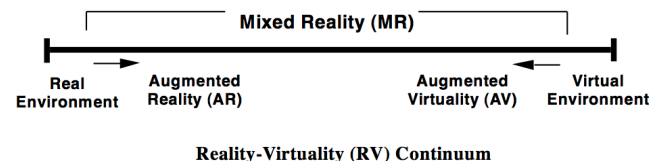
1. INTRODUCTION

Figure 1: The Reality-Virtuality continuum from [20].

Unlike Virtual Reality (VR) which only focuses on displaying and interacting with virtual environments, Augmented Reality (AR) aims at interweaving reality with a virtual world. Indeed, although AR is based on techniques developed in VR [1] the display and interaction of an AR application has a degree of interdependence with the real world. The main challenges of AR consist of the introduction of artificial, computer generated 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. Beyond the technological challenge of this collocation problem (also called *registration* by Azuma [1]), the reproduction of virtual objects, their fidelity and their consistency with the real world are still open research questions.

Milgram et al. [19, 20], defined the well-known "Reality-Virtuality continuum", cf. Fig. 1, where "Reality" and "Virtual Reality" (both being at one end of the continuum) surround "Mixed Reality" (MR), a subclass of VR technologies that involve the merging of real and virtual worlds. Mixed Reality itself is decomposed into "Augmented Reality" (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 uses.

As stated in [11], "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. We will however keep the term "Augmented Reality" even if we understand it as an "increased perception of reality".

In the remainder of this paper, we will give an overview of

existing AR taxonomies, discuss their specificities and limitations. Then, we will propose our own taxonomy, based on four criteria: (1) the number of degrees-of-freedom required for the tracking, (2) the visualization mode, i.e. the augmentation type used by the applications, (3) the temporal base of the content displayed and (4) the rendering modalities used by the AR application. Before drawing a conclusion, we will discuss the benefits and limitations of our approach and will use our typology to classify existing applications.

2. BACKGROUND

Even though a clear definition of augmented reality has not been agreed on by the community, stating whether an application uses some kind of augmented reality or not is easier to decide. What remains more difficult to achieve is to classify the different approaches or applications using AR into a meaningful taxonomy.

Existing taxonomies differ in the criteria they use to classify applications, we chose to divide them into:

- technique-centered,

- user-centered,

- information-centered,

- interaction-centered.

Each category has its characteristics, benefits and drawbacks, which we will present in the following.

2.1 Technique-centered taxonomies

In [19, 20] 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 (for example about where to look for interesting information in the image – a region of interest for tracking – or what the system should be looking for – the 3D model of an object). 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, that is how much the user experiences presence**, within the scene. As a consequence, EPM is minimal when the used display is monoscopic and maximum with high-end head-mounted displays (HMD) that can display real-time 3D graphics and offer see-through capabilities.

In [18], the Reality-Virtuality continuum and some of the elements presented in [19] 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. The idea behind the last criterion is that, depending on the means provided and the circumstances, a user can effect changes in the observed scene either congruently with, or, to varying degrees, incongruently with respect to the form, position and orientation of the device(s)

provided. Instinctively, a highly congruent control-display relationship corresponds with a natural, or intuitive control scheme, whereas an incongruent relationship will compel the user to perform a number of mental transformations in order to use it.

Based on the proposal of a general architecture of an augmented reality system presented in [30], Braz and Pereira [4] developed a web based platform called TARCAST which aimed at listing and characterizing AR systems. The six classification criteria (i.e. the six so-called classical subsystems of an AR system) used in TARCAST are: the Real World Manipulator subsystem, the Real World Acquisition subsystem, the Tracking subsystem, the Virtual Model Generator subsystem, the Mixing Realities subsystem and finally the Display subsystem. Each criterion is composed of a number of features allowing to distinguish different AR systems. TARCAST uses an XML like syntax to describe each feature for each subsystem of an AR system and offers a web interface which allowed users to browse the list of all AR systems included in TARCAST, registered users could also insert new TARCAST characterizations via a specific web-based interface. However, TARCAST does not propose actual criteria but offers a long list of features for each system, hence is not really discriminative. Additionally, TARCAST does not seem to be maintained anymore.

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 but 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.

2.2 User-centered taxonomies

Lindeman and Noma [14] 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 (a) the real environment before reaching (b) a sensory subsystem where it is translated into (c) nerve impulses and finally transmitted to (d) the brain. 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. The authors refer to the different places (a) through (d) where computer generated elements can be inserted as "mixing points". 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 points through appropriate hardware (depending on the sense being stimulated). The authors state that the insertion of computer generated elements should happen as early as possible in the pathway (i.e. at the (a) mixing point) in order to take advantage of the human sensory system which process the real

world stimulus. Based on the location of the mixing points in the process of a stimulus, the authors build their classification for each sense based on a set of existing techniques.

Wang and Dunston [31] 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. [11] 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, e.g. applications such as Sixth Sense [21] or Omnitouch [10]. 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.

2.3 Information-centered taxonomies

In [27], 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 [29], 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, often using indirect visualization, and guiding references to objects outside the field of view that might be visible if the user looks towards that direction) and mounting (differentiates where a virtual object or information is displayed 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 current work-in-progress taxonomy use nearly 40 publications taken from ISMAR's recent conferences in order to test their taxonomy based on those six presentation classes.

Suomela and Lehikoinen propose a taxonomy that can only be applied to location-based applications, thus oriented towards mobile AR. Moreover they do not tackle multi-modal mobile AR applications. Nevertheless, we found the degrees of freedom approach to be interesting and we decided to generalize it in our own proposed taxonomy. Tönnis and Plecher propose an interesting complete taxonomy but they do not deal with the multi-modality that can be used in AR applications and some of the criteria presented are somehow vague (e.g. the mounting criterion).

2.4 Interaction-centered taxonomies

Mackay [15] proposed a taxonomy which is neither based on the technology used, nor on the functionalities nor 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 directly affected, 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.

In [7], Dubois et al. propose a framework for classifying AR systems and use Computer Aided Medical Intervention (CAMI) systems in order to illustrate their classification. Their approach, called OPAC, is based on four components: the System, the Object of augmentation, the Person (the user) and the Adapters (input or output devices) and distinguish between two "main" tasks of the user depending on whether the task has to be performed in the real world (i.e. in AR) or in the virtual world (i.e. in AV). Based on this distinction and on Milgram and Kishino's [19] Reality-Virtuality continuum, the authors propose two different continua ranging respectively from Reality to Virtuality (R→V) and *vice versa* (V→R) where, along the V→R axis, they position different interaction principles proposed by Fishkin et al. [8].

In [6], Dubois et al. propose an extension, called ASUR, of their previous work, where the OPAC components are slightly modified into Adapters, System, User and Real object, where inputs and outputs adapters are more clearly distinguished in the link they create between the System and the real world (composed of the User and the Real Object). In this paper, the authors define relationships between the four components that aim at helping the developers of such systems to reflect upon the combination of the real and virtual worlds as well as the boundaries between those two worlds, while designing mixed reality applications.

The OPAC and ASUR methods presented by Dubois et al. aim at reasoning on Mixed Reality systems, thus they do not classify AR methods strictly speaking. Indeed, the components and relationships presented in their work help modeling the AR and AV systems, rather than characterizing different methods and classify them into categories.

3. PROPOSAL

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

- the first axis is based on the number of degrees of freedom of the tracking 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 represents the augmentation type, i.e. whether it consists in augmenting the whole world or whether it is linked to the user (including an artefact for some kind of mediation).

- the third axis is application-based and covers the temporal base of the content displayed by the application.

- the fourth 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. **As a consequence, this last axis should be considered, as for now, as an optional sub-axis for the classification.**

3.1 Tracking

The main originality of our taxonomy resides in this first classification axis, namely the tracking degrees of freedom. With this term we do not imply vision-based tracking as in the classical computer vision sense (e.g. marker tracking or features tracking) but rather tracking in a broader sense. In our taxonomy, tracking could be instantiated based on the applications requirements, for example tracking can be seen as user-tracking in location-based applications where the important information is the position and orientation of the user in the world, but on the other hand in a classical vision-based application tracking can be indeed seen as tracking of a marker. Hence, we want to focus on the number of freedom required for localizing the "interaction device" – which could be either the user or the camera, tablet, smartphone, etc., depending on the application – with respect to the environment.

On this first axis, we sort applications by the number of degrees of freedom they require and the spatio-temporal accuracy requirements where applicable. If we look throughout current applications, they 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 [5]) 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. A typical example for this kind of application would be to detect a QR code on an advertisement which will then open the manufacturer's web page on your mobile device. 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 [9] 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 [24] 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 [3, 26].

We believe this axis to be very important because it offers a high discriminative power in terms of applications type since tracking is a very important feature in most AR applications and we consider it can determine different classes of applications. Indeed, the tracking degrees-of-freedom we presented above allowed us to distinguish between generic types of AR applications, such as location-based, which groups a whole set of applications shar-

ing common requirements. Moreover, to the best of our knowledge this is the first time this classification criterion is proposed for a taxonomy.

3.2 Augmentation type

For the second axis of our taxonomy, which represents the use of augmentation, we distinguish between two possibilities: (a) "mediated augmentation" and (b) "direct augmentation".

The first one is dedicated to (active) observation applications. It includes two main categories depending on the device used for the mediated observation of the environment:

- *Optical see-through* (OST) applications: there are mostly found in head-up displays (HUD) where they are mostly in the $2D+\theta$ 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 so far 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* (VST) 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, computer generated, images. These applications are often called *magic windows* or "video see-through" [19]. Another metaphor called *magic mirror* is a specific case of the *magic window* where the camera and the screen point in the same direction (e.g. a front-located camera on a smartphone).

The "direct augmentation" application type, also called *Spatially Augmented Reality* (SAR) [2, 25] consists in adding information to the real world, not simply adding information between the observer's eye and the real world. This is achieved by using projectors that display the computer generated artificial images directly on top of the real world objects. These applications have a better potential for collaborative multi-user work (even if some occlusion problems might appear when a user stands in front of one of the projectors) since it is easier for the users to interact with real worlds objects since the visualization of the augmentation does not require the user to wear or to use any additional device. SAR applications are often large scale applications where the projectors usually do not move, but they can also be highly mobile applications such as Sixth Sense [21] or OmniTouch [10].

3.3 Temporal base

Our third axis is based on Hugues' work [11] and is more application-based as it deals with the temporal base of the content displayed in the application. We distinguish between:

- $< t_0$ applications that represent past situations such as archaeological applications,

- t_0 applications devoted to augmenting the world with present information,

- $> t_0$ applications that are dedicated to foreseeing the future state of a given location (e.g. a future building inserted in its environment),

- ∞ applications that represent some full imaginary applications.

3.4 Rendering modalities

The last **sub-axis** of our taxonomy refers to the modalities involved in AR applications. 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 [14]. While the addition of sound in AR applications seems quite straightforward and common, it is 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 [22].

Nevertheless, we believe that multi-modality should be taken into account in a typology of AR-based applications, and that our degrees-of-freedom approach provides for the integration of multiple modalities. 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. Nonetheless, it has to be noted that in the presence of moving sound-generating objects or user, 3D audio real-time feedback becomes very complex.

As for the haptic modality, we propose a similar approach. A simple vibration, (e.g. provided by a mobile phone vibrator) is a 0D stimulus, while the use of specific devices could account for higher dimensions of the haptic modality. For example, the use of a PHANTOM [16] device would account for 3D haptic modality (since the basic PHANTOM has 3 degrees of freedom haptic feedback).

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 of 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. **This is why we recommend using the rendering modalities criterion as a sub-axis of the taxonomy. Using this criterion could nevertheless be needed in future applications and we believe it is worth keeping it in mind.**

Collaborative AR has not yet been extensively tackled in the literature, of course some work exist on multi-user AR but so far mono-user AR is much more investigated. Mobile collaborative AR raises some interesting problems in terms of registration, update, synchronization or user interfaces of the current state of applications for users that could late-join the application.

3.5 Classifying AR applications

In this section we illustrate our proposal by creating a 3D representation of some representative AR applications within our taxonomy axis, cf. Fig. 2. In order to be able to create a representation, we decided not to take into account the multi-modal axis. As mentioned before, although multi-modality remains currently anecdotal in AR applications, we believe it may become more widely used in the future and that this axis remains valid. But for simplicity sakes of representation, we decided to focus only on the first three axis of our proposal, namely: tracking degrees-of-freedom, augmentation type and temporality.

Corresponding to our previous descriptions, those axes have respectively four (0D, 2D, 2D+θ, 6D), three (OST, VST, SAR) and four units ($< t_0$, t_0, $> t_0$, ∞). Each application classified in our taxonomy is represented by a 3D point corresponding to its position in the 3D space, a screenshot illustrating the application and its reference in our bibliography.

As shown in Fig. 2, each application is represented by its 3D position on our three axis. As for every taxonomy, many applications belong to the same category, hence would be located at the same coordinates. For example, many applications belong to the category corresponding to a 6D tracking requirement, using video see-through displays and that augment the world with present information. We chose to illustrate this category with a mobile version of the AR-ToolKit [12] library but many other applications could have been chosen instead, for example applications using marker-less tracking. This points out the fact that our classification could be extended to further analyse points in this space where there are many application candidates. We decided not to add too many applications in order to keep the figure readable, and of course as for the multi-modal axis (not represented in Fig. 2) our taxonomy could be updated when new AR applications arise.

4. CONCLUSIONS

In this paper, we have briefly surveyed and discussed existing taxonomies of augmented reality applications. We have then proposed our own typology based on application tracking requirements, application type, temporality and rendering modalities. In order to illustrate the relevance of our classification, we represented characteristic visual-based AR applications in a 3D coordinate system (where the rendering modalities axis has been removed due to relative low number of AR applications using more than the visual modality).

Our proposal does not cover every issue related to the classifications of AR applications. In particular, since our taxonomy is based on AR applications, we can classify applications that use the same underlying technology (e.g. video see-through marker-based 6D augmented reality) into different categories depending on the *goal* of the application. For example the technology used in the Nintendo AR Cards [23] application developed for the new 3DS console is very similar to the one used in ARToolKit [12] but since Nintendo's application is a video game, it falls into the ∞ section of the Temporal axis while an application using ARToolKit markers in order to test new furniture in one's house will fall into the $> t_0$ section of the same axis.

Nonetheless, we believe the proposed taxonomy overcomes some of the limitations of existing work that we detailed in Section 2, especially multi-modality which is rarely tackled in the literature. Besides the proposed taxonomy presents the advantage of offering a relatively low number of classification criteria, which allows for general categories while keeping the classification process of an augmented reality application relatively easy and straightforward.

As future work, we plan on classifying important work and projects in the field of AR into our taxonomy in order for people to be able to detect new areas of synergy and to propose new direction for future research in the area of AR. Moreover, existing AR development frameworks (such as Layar [13], Metaio [17], Total Immersion [28]) could be fitted into the taxonomy where they would represent surfaces or even volumes encompassing AR applications already characterized in the 3D space defined by our axis.

5. REFERENCES

[1] R. T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, aug 1997.

[2] O. Bimber and R. Raskar. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A K Peters/CRC Press, July 2005.

[3] N. Bioret, M. Servières, and G. Moreau. Outdoor localization based on image/gis correspondence using a simple 2d building layer. In T. Badard and S. Daniel, editors, *2nd International Workshop on Mobile Geospatial Augmented Reality*, LNGC, Québec, Canada, aug 2008. Springer-Verlag.

[4] J. M. Braz and J. a. M. Pereira. Tarcast: Taxonomy for augmented reality CASTing with web support. *The International Journal of Virtual Reality*, 7(4):47–56, Dec 2008.

[5] DensoWave. QR Code. `http://www.qrcode.com/index-e.html`, 1994.

[6] E. Dubois, P. D. Gray, and L. Nigay. Asur++: Supporting the design of mobile mixed systems. *Interacting with Computers*, 15(4):497–520, 2003.

[7] E. Dubois, L. Nigay, J. Troccaz, O. Chavanon, L. Carrat, and E. Al. Classification space for augmented surgery, an augmented reality case study. *Conference Proceedings of Interact*, 99:353–359, 1999.

[8] K. P. Fishkin, T. P. Moran, and B. L. Harrison. Embodied user interfaces: Towards invisible user interfaces. In *Proceedings of the Seventh Working Conference on Engineering for Human-Computer Interaction*, EHCI'98, pages 1–18, Deventer, The Netherlands, The Netherlands, 1999. Kluwer, B.V.

[9] Google. Google maps for mobile. `http://www.google.com/mobile/maps/`, 2010.

[10] C. Harrison, H. Benko, and A. D. Wilson. Omnitouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 441–450, New York, NY, USA, 2011. ACM.

[11] O. Hugues, P. Fuchs, and O. Nannipieri. New augmented reality taxonomy: Technologies and features of augmented environment. In B. Furht, editor, *Handbook of Augmented Reality*, chapter 2, pages 47–63. Springer, 2011.
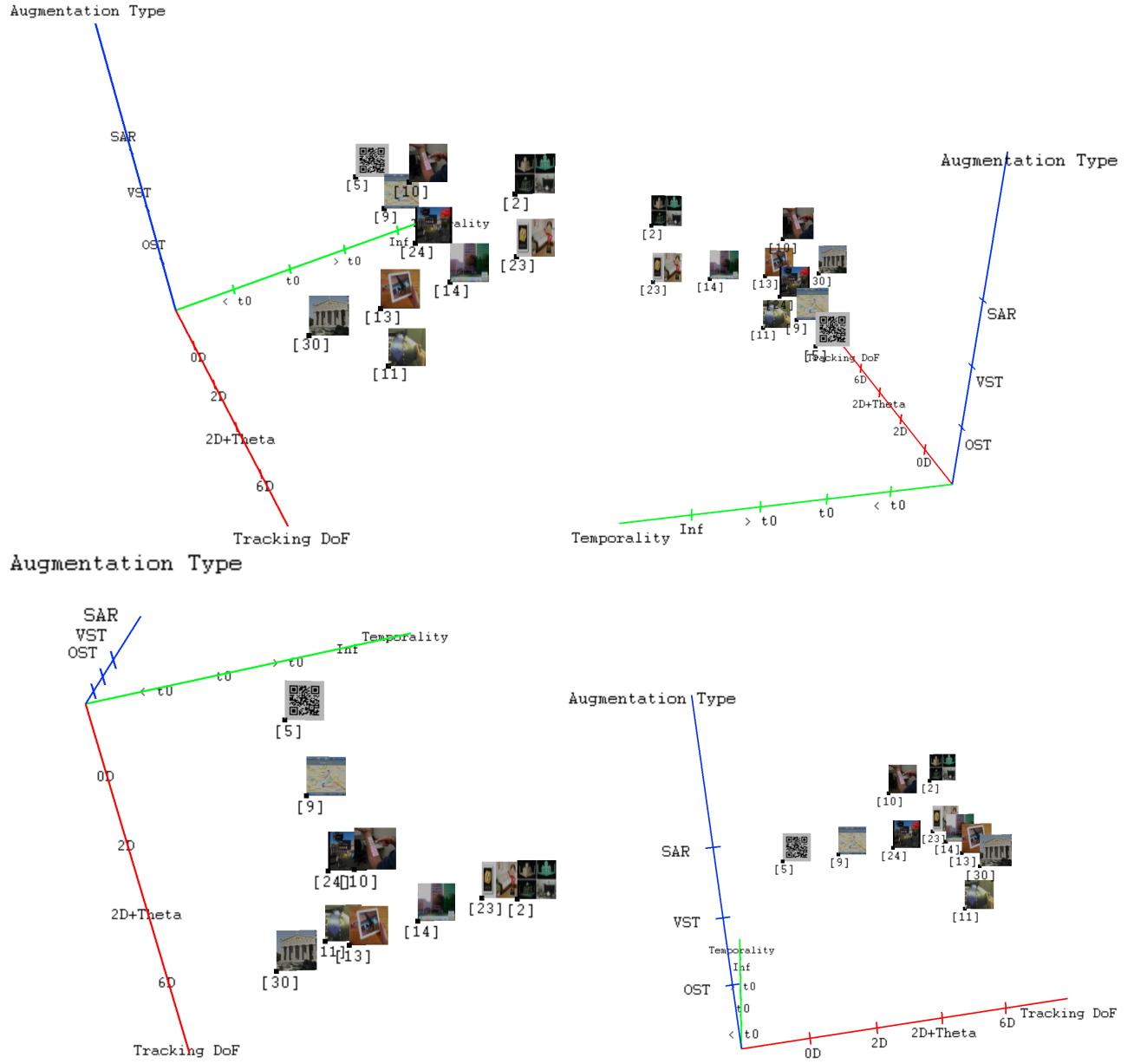
Figure 2: Classification of some AR applications using our taxonomy.

[12] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality*, IWAR'99, pages 85–, Washington, DC, USA, 1999. IEEE Computer Society.

[13] Layar. `http://www.layar.com/`.

[14] R. W. Lindeman and H. Noma. A classification scheme for multi-sensory augmented reality. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology*, VRST '07, pages 175–178, New York, NY, USA, 2007. ACM.

[15] W. E. Mackay. 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, pages 13–21, New York, NY, USA, 1998. ACM.

[16] T. H. Massie and K. J. Salisbury. The PHANToM Haptic Interface: A Device for Probing Virtual Objects. *American Society of Mechanical Engineers, Dynamic Systems and Control Division (Publication) DSC*, 55-1:295–299, 1994.

[17] Metaio. `http://www.metaio.com/`.

[18] P. Milgram and H. J. Colquhoun. A taxonomy of real and virtual world display integration. In Y. Ohta and H. Tamura, editors, *Mixed Reality – Merging Real and Virtual Worlds*, chapter 1, pages 1–16. Ohmsha(Tokyo) and Springer Verlag (Berlin), 1999.

[19] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems*, E77-D(12):1321–1329, Dec 1994.

[20] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. *Proceedings of Telemanipulator and Telepresence Technologies*, 2351(34):282–292, 1994.

[21] P. Mistry, P. Maes, and L. Chang. Wuw - wear ur world: a wearable gestural interface. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, CHI EA '09, pages 4111–4116, New York, NY, USA, 2009. ACM.

[22] T. Narumi, S. Nishizaka, T. Kajinami, T. Tanikawa, and M. Hirose. Meta cookie+: An illusion-based gustatory display. In *Proceedings of the 2011 international conference on Virtual and mixed reality: new trends - Volume Part I*, pages 260–269, 2011.

[23] Nintendo. AR Cards. `http://www.nintendo.com/3ds/ar-cards`, 2011.

[24] Presselite. Metro paris subway. `http://www.metroparisiphone.com/`, 2010.

[25] R. Raskar, G. Welch, and H. Fuchs. Spatially augmented reality. In *In First IEEE Workshop on Augmented Reality*, IWAR'98, pages 11–20, 1998.

[26] G. Reitmayr and T. W. Drummond. Initialisation for Visual Tracking in Urban Environments. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 1–9. Ieee, Nov. 2007.

[27] R. Suomela and J. Lehikoinen. Taxonomy for visualizing location-based information. *Virtual Reality*, 8:71–82, September 2004.

[28] TImmersion. `http://www.t-immersion.com/projects`.

[29] M. Tönnis and D. A. Plecher. Presentation Principles in Augmented Reality - Classification and Categorization Guidelines. techreport, Technische Universität München, 2011. Version 1.0.

[30] J. R. Vallino. *Interactive Augmented Reality*. PhD thesis, Department of Computer Science, University of Rochester, Rochester, NY, USA, 1998.

[31] X. Wang and P. S. Dunston. 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*, pages 1836–1842, 2006.