DoF-based Classification of Augmented Reality Applications

Jean-Marie Normand* Guillaume Moreau†

LUNAM Université, École Centrale de Nantes,
CERMA UMR 1563, Nantes, France.

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 tracking of the application, as well as on (2) the visualization mode used, (3) the type of tracking used in the application 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. **This article is mainly based on our previous paper published in the 3rd ACM Augmented Human AH'12 conference [29].**

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information System—Artificial, augmented and virtual realities

1 Introduction

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. [23, 24], defined the well-known "Reality-Virtuality continuum" 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 [14], "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 cannot 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 way frames of references are linked together, (3) the type of tracking used in the application and (4) the rendering modalities used by the AR application. Before drawing a conclusion, we will discuss the benefits and limitations of our approach, will use our typology to classify existing applications and propose possible extensions to our typology.

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 [23, 24] 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,

*e-mail: jean-marie.normand@ec-nantes.fr
†e-mail: guillaume.moreau@ec-nantes.fr

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 [22], the Reality-Virtuality continuum and some of the elements presented in [23] lay the groundwork for a global taxonomy of MR display integration. The classification is based on three axes: 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 [39], Braz and Pereira [5] 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 [19] 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 [41] 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. [14] 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 [25] or Omnitouch [12]. 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 [37], 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 [38], 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 [20] 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 [8], 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 [23] 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. [10].

In [7], 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). They aim at helping the developers of MR applications to reflect upon the combination of the real and virtual worlds as well as the boundaries between those two worlds.

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

With the recent democratisation of smartphones, tablets and mobile computing, Ubiquitous Computing (seamless integration of computer in everyday life), first proposed by Weiser [42] in the 1990s seems to be more and more related to AR technologies. Although those two notions do not address the same problems, they obviously are related. Based on the Milgram continuum, Newman et al. [28] proposed the so-called " Weiser continuum", a 1-D axis where Ubiquitous computing would appear at one end of the axis while monolithic mainframe-based computing would lie at the other end (classical personal computers – PC– would then lie in the very middle of this continuum). The authors then combine those two continua into the "Milgram-Weiser" 2D continuum in order to relate VR and AR to Ubiquitous computing. This 2D space allows the authors to classify Mixed Reality and Ubiquitous computing applications and highlight gaps in the existing technologies. This taxonomy served as a basis for defining a middleware architecture for Ubiquitous Tracking (or "UbiTrack", see also [31]).

3 Proposal

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

- 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 relations that may exist between the various frames of reference that are involved in the augmentation process (the user, the sensor(s), the display system(s) and the real world).
- the third axis is application-based and covers the type of tracking used by the application.
- the fourth axis covers other rendering modalities that go beyond visual augmented reality. Strictly speaking, this last axis would rather be declined as multiple axes, each of them being dedicated to a specific rendering modality: audio, haptic, gustatory and olfactory. It should be noted that, as of today, the use of non-visual rendering modalities remains rather limited today, but they can nevertheless be taken into account by the same degrees-of-freedom system. As a consequence, this last axis (or axes) should be considered, as for now, as optional.

3.1 Tracking

The main originality of our taxonomy lays 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. On the other hand, in a classical vision-based application, tracking can be seen as tracking of a marker, i.e. the position of the camera with respect to the marker which represents the position information required by the application. Tracking requirements can also depend on the display device, whether it is rigidly linked to user position or not, cf. Section 3.2. 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 [6]) 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 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 [11] 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. Every navigation system is based on this principle, and for those applications, 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 less than metric, update rates typically ranging from 1 to 10Hz. A typical example of a 2D + θ application is the Metro Paris [30] 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/depth cameras, inertial sensors, etc.). Various precision classes exist depending on application types and on the on the technology used 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 (i.e. recovering the new position from the former one with a small motion assumption at high frequency) must be distinguished from initial localization for which there exists fewer works [4, 33]. In the early years of AR, many other (non-visual) technologies were used to deliver 6D tracking. Although nowadays the vast majority of AR applications rely on visual and/or a combination of inertial sensors, which present the advantage of being wireless and can be used outdoors, one can also rely upon technologies commonly used in VR, such as electromagnetic, acoustic or mechanical tracking, cf. [34, 43].

It should be noted that the 2D+θ notation has been chosen knowingly, and we prefer it to the 3D notation, mainly because the latter could be confused with applications requiring three dimensional positional information (i.e. X, Y, Z coordinates).

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 applications classes. Indeed, the tracking degrees-of-freedom we presented above allowed us to distinguish between generic types of AR applications, such as location-based ones, which group a whole set of applications sharing common requirements. To the best of our knowledge this is the first time this classification criterion is proposed for a taxonomy.

DOF	**Precision classes**	**Typical Update Rates**
0D	N.A.	N.A.
2D	$\sim$ 10m	$\sim$ 1Hz
2D + θ	$\sim$ 10m	$\sim 1-10$Hz
6D	$\leq$ 1cm	$\sim 10-100$Hz

Table 1: Summary of degrees of freedom versus metric accuracy and update frequency.

3.2 Degrees of Freedom between frames of reference

For the second axis of our taxonomy we chose to model the links between the user frame of reference and the frame of reference of the device used to display the AR content. Indeed we believe that there is always a device involved when talking about AR technologies. Those devices differ depending on the "type" of augmented reality used:

- for Optical See-Through (OST) applications, devices are mostly head-up displays (HUD), e.g. for HUDs fixed to a vehicle or see-through glasses (or for worn HMDs) where optical information is projected onto special lenses.

- for Video See-Through (VST) applications, we distinguish between:
  - HMDs : where cameras attached to a head-mounted display film the scene and are used to compute the position and orientation of the information to be displayed on top of the video stream,
  - devices equipped with a back-located camera (such as a tablet or a smartphone) that films the real environment and for which the video is reproduced on its display augmented with artificial, computer generated, images. These VST applications are often called *magic windows* or "video see-through" [23]. Another metaphor called *magic mirror* is a specific case of a *magic window* where the camera and the screen point in the same direction (e.g. a front-located camera on a smartphone).

- for *Spatially Augmented Reality* (SAR) [3, 32] applications, which consist 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. Here again w distinguish between:
  - fixed projector applications: these applications are often large-scale applications and 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.
  - projectors attached to the user: these are typically highly mobile applications such as Sixth Sense [25] or OmniTouch [12], where the projector is physically and rigidly linked to the user. There also exist some

augmented reality applications based on head-mounted laser projectors that can display information directly on the environment, see for example [36].

Based on these devices, we distinguish between four types of relations between the several frames involved (the user frame, the display frame, the sensor frame and the world frame):

1. tight relation between display device and the user: this is the case for OST and HMD-based VST as well as head-mounted projectors that can be found in some Spatially Augmented Reality applications where the spatial transform between the eyes of the user and the display device is fixed over time. These devices may also include a fixed transform to some sensors (one or several video cameras in most of the cases).

2. loose relation between the display device and the user: this is the case for *magic-window* based VST applications: where there is a relation between the frame of reference of the user and the one of the device but this relation is relatively loose. For example, when holding a smartphone, the user can move his head and keep the phone fixed, or the other way around. But he can also move the device as he moves his head. In this case, the sensors (video camera, inertial sensors...) are rigidly linked to the display device.

3. no relation between the display device and the user: this is the case for fixed-projector(s) based SAR applications. In this case, there is no relation between the frame of reference of the display device and the one of the user. The user's movements are independent of the frame of reference of the projector. The latter is rigidly linked to the word frame of reference.

4. both frames of references are "merged": this is the case for Location-Based Services such as GPS localization for which the user frame of reference does not play any role, only the device is taken into account. At least we can assume that the various class of precision used are different enough to allow this merge: in the case of GPS-navigation, the position and orientation of the car and the driver are fairly similar with respect to a complete trajectory that can be much longer.

We illustrate the relations between user and devices frames of reference in Fig. 1 and summarize these relations in Table 2.

Relation between user and device frames	Augmented Reality Device type
Tight	OST, HMD VST, SAR with projectors attached to the user
Loose	Handheld VST
Merged	LBS, LBS + orientation
No Relation	Fixed-projectors SAR, QR-Codes

Table 2: Summary table for relations between frames of references

3.3 Tracking Type

Our third axis aims at differentiating between the multiple tracking technologies that are used in AR applications. Indeed, as we mentioned in Section 3.1 it is possible to achieve the same level of tracking (e.g. 6D) by very different means. Hence, we decided to dedicate an axis of our taxonomy to further discriminate the classification of AR applications. We tried to organize this axis based on the artificial nature of tracking, i.e. whether the environment must be prepared or whether the tracked features are natural or artificial (whole images or small units such as segments). Combination of tracking methods is put aside to avoid checking several marks on an axis. The units on this axis are the following:

- Marker-based tracking (Marker),
- Natural Feature (NF),
- 2D or 3D Template Matching (TM),
- Optical + Sensor Fusion (SF).
- Non optical (NO),

We decided not to distinguish between the non-optical tracking types (e.g. GPS, compass, etc.) in order not to multiply the number of marks on this axis, especially since it is nowadays relatively rare to rely upon inertial, mechanical or acoustic tracking alone. Sensor Fusion on its part corresponds to a combination of data coming from different sources/devices (e.g. accelerometers, gyroscopes, GPS, etc.) and an optical source (camera) allowing to perform tracking. This axis is relatively easy to understand but will nonetheless prove useful in discriminating further AR applications that require the same degrees of freedom for the tracking but that do not rely on the same tracking technology.

3.4 Non-visual Rendering modalities

The last optional axis of our taxonomy refers to the non-visual 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 [19]. 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 [26].

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 [21] device would account for 3D haptic modality (since the basic PHANTOM has 3 DoF haptic feedback).

As of today, visuo-haptic is by far the most popular and researched combination of modalities in AR, especially because it offers a very interesting potential for training applications where the user can learn specific gestures and movements that require tools. For example Sandor et al. [35] developed a visuo-haptic painting application while Bianchi et al. [2] focused on developing AR systems capable of giving realistic haptic feedback from real or virtual objects in real-time. Medical applications and particularly surgery training appear to be a prime target for visuo-haptic AR and research is currently being carried out in order to develop such applications, see for example [16, 15].

Figure 1: Frame of references. An orange arrow denotes a tight link whereas a blue denotes other links.

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 non-visual rendering modalities axes to be optional, each modality could be represented by a single axis. Using these criteria could nevertheless be needed in future applications and we believe it is worth keeping them in mind.

Collaborative AR has not yet been extensively tackled in the literature, of course some work exist on multi-user AR (especially thanks to SAR) 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, see 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 tracking type.

Corresponding to our previous descriptions, each axis has four units: {0D, 2D, 2D+θ, 6D} for the Tracking DOF; {Tight, Loose, Merged, No Relation} for the relation between frames of reference; {Marker, TM, NF, SF or NO} for the tracking type.

As shown in Fig. 2, each application is represented by a 3D position which coordinates corresponds to its characteristics on each of our three axes. Actual coordinates are given in Table 3. 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, having loose relations between frames of references (which would correspond for example to video see-through displays on tablets or smartphones) and that superimpose the real world with present computer generated information. We chose to illustrate this category with a mobile version of the ARToolKit [17] library but many other applications could have been chosen instead, for example applications using markerless tracking.

In order to overcome this limitation, we decided to represent two frameworks directly in Fig. 2. Indeed, AR toolkits may give rise to many applications represented at different locations in our 3D coordinate system. As a consequence, AR toolkits can be modelled as volumes in our taxonomy. We added ARToolKit and Layar frameworks in Fig. 2 in order to illustrate this. Many applications belong to the ARToolKit framework volume meaning that every one of them could also be developed based on this toolkit.

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 non-visual multi-modality axes (which are not represented in Fig. 2) our taxonomy could be updated when new AR applications arise. Anyway, new applications trends including augmentation by acoustic feedback (for car parking), head-up displays in cars or applications helping sportsmen to improve their gestures with acoustic feedback fit our classification.

Table 3: Classification examples

Application	Tracking DoF	Frames Relations	Tracking Type
Archeoguide [40]	6D	Loose	2D TM
Insitu (Outdoors) [18]	6D	Loose	SF
ARMAR [13]	6D	Tight	Marker
VW SAR [1]	6D	No Relation	Marker
Omnitouch [12]	$2D+\theta$	Tight	SF
Metro [30]	$2D+\theta$	Merged	NO
Gmaps [11]	2D	Merged	NO
QR codes [6]	0D	No Relation	Marker
ARToolkit [17]	6D	Loose	Marker

3.6 Possible extensions

In addition to the previous classification criteria, we think that two additional ones could be integrated in a taxonomy.

3.6.1 Integration of virtual and real worlds

We believe that the integration of virtual objects could be use to discriminate between applications and/or toolkits. Indeed, integrating virtual objects within an AR scene gives rise to the following problems and challenges: realistic color integration of virtual objects in the scene, realistic lighting integration of virtual objects in the scene, realistic shadowing of the scene by virtual objects, realistic depth integration of virtual objects in the real scene by occlusion management, interaction of the user with virtual objects, interaction between virtual objects and the real world.

Those criteria help at developing more realistic applications by aiming at a seamless and transparent integration of virtual objects with the real scene, either filmed by a camera or directly seen through see-through lenses. Of course all those criteria are not necessarily met at the same time, and some applications do not meet

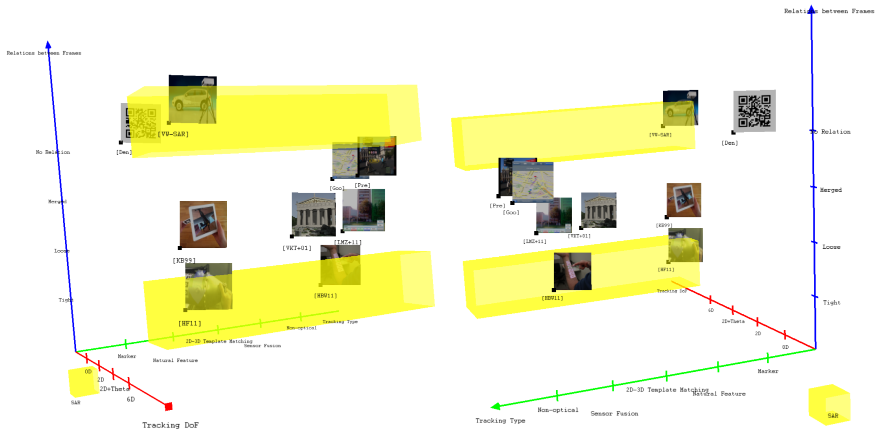

Figure 2: Classification of some AR applications using our taxonomy.

any of them. Nevertheless they may be desirable for some applications, e.g. for an interior design application getting realistic shading and lighting on some virtual furnitures can be a real advantage.

It is worth noticing that some of these criteria represent, as of today, technical challenges, especially for mobile AR where both central (CPU) and graphical (GPU) processors suffer from limited (even if constantly improving) computing power. Moreover mobile AR also faces specific problems such as rapid movements due to direct manual handling of the device by the user as well as easily cluttered information (due to relatively small size of displays). Lastly, the labelling/annotation problem also faces specific challenges in AR beyond the traditional legibility and occlusion issues. Labels also have to be linked to the objects they inform about but also require a specific attention with respect to the colors they use in order to be readable with respect to the background they are displayed on which is most often the unknown real world.

However, integration of virtual and real worlds does not impact every AR application in the same way. For example, Location-Based AR (e.g. Layar-based applications) does not typically use more than the GPS sensors of a mobile device. As a consequence it seems impossible to work on most of the criteria listed above (except for the interaction one), since they require some real-time analysis of the (unknown) real world in order to modify the virtual objects accordingly. Moreover some criteria might require knowledge of a 3D model of the environment (typically the last criterion, cf. [27] for example) which further complicates the computation.

3.6.2 Level of Abstraction of Virtual Objects

Our last comment raises the question of the display of information for AR applications, especially for mobile AR. Indeed, choosing the right abstraction for information presentation seems a fundamental problem problem for Augmented Reality. Some studies such as [9] focused on the level of abstraction for landmark-based user navigation and differentiate between six levels, cf. Fig 3.

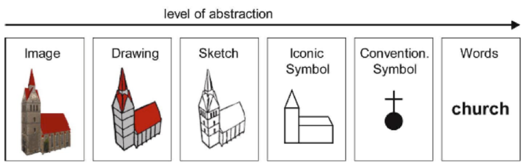

Figure 3: Abstraction levels according to [9].

These levels of abstraction should be investigated for other application types in order to clarify how information should be presented to the user. This is indeed an interesting issue for AR since real world information (on which synthetic information is superimposed) is by nature image data of better than photo-realistic quality. Therefore, adding virtual objects onto this may require a trade-off: high quality rendering may stress out the issues raised in the former paragraph about integration of both worlds whereas introducing more abstraction could be less demanding in terms of integration and even as far as tracking accuracy is involved.

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, tracking type and rendering modalities. In order to illustrate the relevance of our classification, we represented characteristic visual-based AR applications, cf. Table 3, in a 3D coordinate system (where the rendering modalities axes have been removed due to relative low number of beyond visual-AR applications).

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. 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.

References

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

[2] G. Bianchi, C. Jung, B. Knoerlein, G. Szekely, and M. Harders. High-fidelity visuo-haptic interaction with virtual objects in multi-modal ar systems. In *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, ISMAR '06, pages 187–196, Washington, DC, USA, 2006. IEEE Computer Society.

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

[4] 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.

[5] J. M. Braz and J. M. Pereira. TARCAST: Taxonomy for Augmented Reality CASTing with Web Support. *The International Journal of Virtual Reality*, 7(4):47–56, Dec 2008.

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

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

[8] 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.

[9] B. Elias and V. Paelke. User-Centered Design of Landmark Visualizations. In L. Meng, A. Zipf, and S. Winter, editors, *Map-based Mobile Services*, Lecture Notes in Geoinformation and Cartography, chapter 3, pages 33–56. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[10] 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.

[11] Google. Google maps for mobile. `http://www.google.com/mobile/maps/`.

[12] 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.

[13] S. J. Henderson and S. Feiner. Augmented reality in the psychomotor phase of a procedural task. In *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality*, ISMAR'11, pages 191–200, 2011.

[14] 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.

[15] S. Jeon, S. Choi, and M. Harders. Rendering virtual tumors in real tissue mock-ups using haptic augmented reality. *EEE Trans. Haptics*, 5(1):77–84, Jan. 2012.

[16] S. Jeon, B. Knoerlein, M. Harders, and S. Choi. Haptic simulation of breast cancer palpation: A case study of haptic augmented reality. In *Proceedings of the 9th IEEE International Symposium on Mixed and Augmented Reality*, ISMAR 2010, pages 237–238, 2010.

[17] 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 85–94, Washington, DC, USA, 1999. IEEE Computer Society.

[18] T. Langlotz, S. Mooslechner, S. Zollmann, C. Degendorfer, G. Reitmayr, and D. Schmalstieg. Sketching up the world: in situ authoring for mobile augmented reality. *Personal and Ubiquitous Computing*, pages 1–8, 2011. 10.1007/s00779-011-0430-0.

[19] R. W. Lindeman and H. Noma. A classification scheme for multisensory 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.

[20] 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.

[21] 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.

[22] 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.

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

[24] 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.

[25] 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.

[26] 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.

[27] R. A. Newcombe, S. Lovegrove, and A. J. Davison. Dtam: Dense tracking and mapping in real-time. In *Proceedings of the IEEE International Conference on Computer Vision*, ICCV 2011, pages 2320–2327. IEEE, 2011.

[28] J. Newman, A. Bornik, D. Pustka, F. Echtler, M. Huber, D. Schmalstieg, and G. Klinker. Tracking for distributed mixed reality environments. In *Proceedings of IEEE VR 2007 Workshop on 'Trends and Issues in Tracking for Virtual Environments'*, Charlotte, NC, USA, Mar. 2007. IEEE, Shaker Verlag, Aachen, Germany.

[29] J.-M. Normand, M. Servières, and G. Moreau. A new typology of augmented reality applications. In *Proceedings of the 3rd Augmented Human International Conference*, AH '12, pages 18:1–18:8, New York, NY, USA, 2012. ACM.

[30] Presselite. Metro paris subway. `http://www.metroparisiphone.com/`.

[31] D. Pustka, M. Huber, C. Waechter, F. Echtler, P. Keitler, J. Newman, D. Schmalstieg, and G. Klinker. Automatic configuration of pervasive sensor networks for augmented reality. *IEEE Pervasive Computing*, 10(3):68–79, July 2011.

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

[33] 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.

[34] J. P. Rolland, L. D. Davis, and Y. Baillot. A survey of tracking technology for virtual environments. In W. Barfield and T. Caudell, editors, *Fundamentals of Wearable Computers and Augmented Reality*, chapter 3, pages 67–112. Routledge, 2001.

[35] C. Sandor, S. Uchiyama, and H. Yamamoto. Visuo-haptic systems: Half-mirrors considered harmful. In *Proceedings of the Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, WHC 2007, pages 292–297. IEEE Computer Society, 2007.

[36] B. Schwerdtfeger, D. Pustka, A. Hofhauser, and G. Klinker. Using laser projectors for augmented reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, VRST '08, pages 134–137, New York, NY, USA, 2008. ACM.

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

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

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

[40] V. Vlahakis, J. Karigiannis, M. Tsotros, M. Gounaris, L. Almeida, D. Stricker, T. Gleue, I. T. Christou, R. Carlucci, and N. Ioannidis. Archeoguide: first results of an augmented reality, mobile computing system in cultural heritage sites. In *Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage*, VAST '01, pages 131–140, New York, NY, USA, 2001. ACM.

[41] 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.

[42] M. Weiser. Some computer science issues in ubiquitous computing. *Commun. ACM*, 36(7):75–84, July 1993.

[43] G. Welch and E. Foxlin. Motion tracking: No silver bullet, but a respectable arsenal. *IEEE Comput. Graph. Appl.*, 22(6):24–38, Nov. 2002.