INVITED PAPER *Special Issue on Networked Reality*

A Taxonomy of Mixed Reality Visual Displays

Paul MILGRAM†, *Nonmember and* **Fumio KISHINO**††, *Member*

SUMMARY This paper focuses on Mixed Reality (MR) visual displays, a particular subset of Virtual Reality (VR) related technologies that involve the merging of real and virtual worlds somewhere along the "virtuality continuum" which connects completely real environments to completely virtual ones. Probably the best known of these is Augmented Reality (AR), which refers to all cases in which the display of an otherwise real environment is augmented by means of virtual (computer graphic) objects. The converse case on the virtuality continuum is therefore Augmented Virtuality (AV). Six classes of hybrid MR display environments are identified. However, an attempt to distinguish these classes on the basis of whether they are primarily video or computer graphics based, whether the real world is viewed directly or via some electronic display medium, whether the viewer is intended to feel part of the world or on the outside looking in, and whether or not the scale of the display is intended to map orthoscopically onto the real world leads to quite different groupings among the six identified classes, thereby demonstrating the need for an efficient taxonomy, or classification framework, according to which essential differences can be identified. The 'obvious' distinction between the terms "real" and "virtual" is shown to have a number of different aspects, depending on whether one is dealing with real or virtual objects, real or virtual images, and direct or non-direct viewing of these. An (approximately) three dimensional taxonomy is proposed, comprising the following dimensions: Extent of World Knowledge ("how much do we know about the world being displayed?"), Reproduction Fidelity ("how 'realistically' are we able to display it?"), and Extent of Presence Metaphor ("what is the extent of the illusion that the observer is present within that world?").

key words: *virtual reality (VR), augmented reality (AR), mixed reality (MR)*

1. Introduction—Mixed Reality

The next generation telecommunication environment is envisaged to be one which will provide an "ideal virtual space with [sufficient] reality essential for communication."* Our objective in this paper is to examine this concept, of having both "virtual space" on the one hand and "reality" on the other available within the same visual display environment.

The conventionally held view of a *Virtual Reality* (VR) environment is one in which the participant-observer is totally immersed in, and able to interact with, a completely synthetic world. Such a world may mimic the properties of some real-world environments, either existing or fictional; however, it can also exceed the bounds of physical reality by creating a world in which the physical laws ordinarily governing space, time, mechanics, material properties, etc. no longer hold. What may be overlooked in this view, however, is that the VR label is also frequently used in association with a variety of other environments, to which total immersion and complete synthesis do not necessarily pertain, but which fall somewhere along a *virtuality continuum*. In this paper we focus on a particular subclass of VR related technologies that involve the merging of real and virtual worlds, which we refer to generically as *Mixed Reality* (*MR*). Our objective is to formulate a taxonomy of the various ways in which the "virtual" and "real" aspects of MR environments can be realised. The perceived need to do this arises out of our own experiences with this class of environments, with respect to which parallel problems of inexact terminologies and unclear conceptual boundaries appear to exist among researchers in the field.

The concept of a "virtuality continuum" relates to the mixture of classes of objects presented in any particular display situation, as illustrated in Fig. 1, where *real environments*, are shown at one end of the continuum, and *virtual environments*, at the opposite extremum. The former case, at the left, defines environments consisting solely of real objects (defined below), and includes for example what is observed via a conventional video display of a real-world scene. An additional example includes direct viewing of the same real scene, but not via any particular electronic display system. The latter case, at the right, defines environments consisting solely of virtual objects (defined below), an example of which would be a conventional

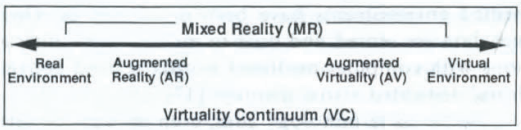

Fig. 1 Simplified representation of a "virtuality continuum."

Manuscript received July 8, 1994.
Manuscript revised August 25, 1994.

† The author is with the Department of Industrial Engineering, University of Toronto, Toronto, Ontario, Canada M5S 1A4.
†† The author is with ATR Communication Systems Research Laboratories, Kyoto-fu, 619-02 Japan.

* Quoted from Call for Papers for this IEICE Transactions on Information Systems special issue on Network Reality.

computer graphic simulation. As indicated in the figure, the most straightforward way to view a Mixed Reality environment, therefore, is one in which real world and virtual world objects are presented together within a single display, that is, anywhere between the extrema of the virtuality continuum.

Although the term "Mixed Reality" is not (yet) well known, several classes of existing hybrid display environments can be found, which could reasonably be considered to constitute MR interfaces according to our definition:

1. Monitor based (non-immersive) video displays—i.e. "window-on-the-world" (WoW) displays—upon which computer generated images are electronically or digitally overlaid [1]-[4]. Although the technology for accomplishing such combinations has been around for some time, most notably by means of chroma-keying, practical considerations compel us to be interested particularly in systems in which this is done *stereoscopically* [5], [6].
2. Video displays as in Class 1, but using immersive head-mounted displays (HMD's), rather than WoW monitors.
3. HMD's equipped with a see-through capability, with which computer generated graphics can be optically superimposed, using half-silvered mirrors, onto directly viewed real-world scenes [7]-[12].
4. Same as 3, but using video, rather than optical, viewing of the "outside" world. The difference between Classes 2 and 4 is that with 4 the displayed world should correspond orthoscopically with the immediate outside real world, thereby creating a "video see-through" system [13], [14], analogous with the optical see-through of option 3.
5. Completely graphic display environments, either completely immersive, partially immersive or otherwise, to which video "reality" is added [1].
6. Completely graphic but partially immersive environments (e.g. large screen displays) in which real physical objects in the user's environment play a role in (or interfere with) the computer generated scene, such as in reaching in and "grabbing" something with one's own hand [15], [16].

In addition, other more inclusive computer augmented environments have been developed in which real data are sensed and used to modify users' interactions with computer mediated worlds beyond conventional dedicated visual displays [17]-[20].

As far as terminology goes, even though the term "Mixed Reality" is not in common use, the related term "Augmented Reality" (AR) has in fact started to appear in the literature with increasing regularity. As an operational definition of Augmented Reality, we take the term to refer to any case in which an otherwise real environment is "augmented" by means of virtual (computer graphic) objects, as illustrated in Fig. 1. The most prominent use of the term AR in the literature appears to be limited, however, to the Class 3 types of displays outlined above [8], [10]-[12]. In the authors' own laboratories, on the other hand, we have adopted this same term in reference to Class 1 displays as well [5], [21], not for lack of a better name, but simply out of conviction that the term Augmented Reality is quite appropriate for describing the essence of computer graphic enhancement of video images of real scenes. This same logic extends to Classes 2 and 4 displays also, of course.

Class 5 displays pose a small terminology problem, since that which is being augmented is not some direct representation of a real scene, but rather a *virtual* world, one that is generated primarily by computer. In keeping with the logic used above in support of the term Augmented Reality, we therefore proffer the straightforward suggestion that such displays be termed "*Augmented Virtuality*" (AV), as depicted in Fig. 1†. Of course, as technology progresses, it may eventually become less straightforward to perceive whether the primary world being experienced is in fact predominantly "real" or predominantly "virtual," which may ultimately weaken the case for use of both *AR* and *AV* terms, but should not affect the validity of the more general *MR* term to cover the "grey area" in the centre of the virtuality continuum.

We note in addition that Class 6 displays go beyond Classes 1, 2, 4 and 5, in including directly viewed real-world objects also. As discussed below, the experience of viewing one's own *real* hand directly in front of one's self, for example, is quite distinct from viewing an image of the same real hand on a monitor, and the associated perceptual issues (not discussed in this paper) are also rather different. Finally, an interesting alternative solution to the terminology problem posed by Class 6 as well as composite Class 5 AR/AV displays might be the term "*Hybrid Reality*" (HR)††, as a way of encompassing the concept of blending many types of distinct display media.

2. The Need for a Taxonomy

The preceding discussion was intended to introduce the concept of Mixed Reality and some of its various manifestations. All of the classes of displays listed

† Cohen (1993) has considered the same issue and proposed the term "Augmented Virtual Reality." As a means of maintaining a distinction between this class of displays and Augmented Reality, however, we find Cohen's terminology inadequate

†† One potential piece of derivative jargon which immediately springs to mind as on extension of the proposed term "Hybrid Reality" is the possibility that (using a liberal dose of poetic license) we might refer to such displays as "Hyberspace"!

above clearly share the common feature of juxtaposing "real" entities together with "virtual" ones; however, a quick review of the sample classes cited above reveals, among other things, the following important distinctions:

- Some systems {1, 2, 4} are primarily video based and enhanced by computer graphics whereas others {5, 6} are primarily computer graphic based and enhanced by video.
- In some systems {3, 6} the real world is viewed directly (through air or glass), whereas in others {1, 2, 4, 5} real-world objects are scanned and then resynthesized on a display device (e.g. analogue or digital video).
- From the standpoint of the viewer relative to the world being viewed, some of the displays {1} are exocentric (WoW monitor based), whereas others {2, 3, 4, 6} are egocentric (immersive).
- In some systems {3, 4, 6} it is imperative to maintain an accurate 1 : 1 orthoscopic mapping between the size and proportions of displayed images and the surrounding real-world environment, whereas for others {1, 2} scaling is less critical, or not important at all.

Our point therefore is that, although the six classes of MR displays listed appear at first glance to be reasonably mutually delineated, the distinctions quickly become clouded when concepts such as real, virtual, direct view, egocentric, exocentric, orthoscopic, etc. are considered, especially in relation to implementation and perceptual issues. The result is that the different classes of displays can be grouped differently depending on the particular issue of interest. Our purpose in this paper is to present a taxonomy of those principal aspects of MR displays which subtend these practical issues.

The purpose of a taxonomy is to present an ordered classification, according to which theoretical discussions can be focused, developments evaluated, research conducted, and data meaningfully compared. Four noteworthy taxonomies in the literature which are relevant to the one presented here are summarised in the following.

- Sheridan [22] proposed an operational measure of *presence* for remotely performed tasks, based on three determinants: extent of sensory information, control of relation of sensors to the environment, and ability to modify the physical environment. He further proposed that such tasks be assessed according to task difficulty and degree of automation.
- Zeltzer [23] proposed a three dimensional taxonomy of *graphic simulation systems*, based on the components autonomy, interaction and presence. His "AIP cube" is frequently cited as a framework for categorising virtual environments.
- Naimark [24], [25] proposed a taxonomy for categorising different approaches to recording and reproducing visual experience, leading to *real-space imaging*. These include: monoscopic imaging, stereoscopic imaging, multiscopic imaging, panoramics, surrogate travel and real-time imaging.
- Robinett [26] proposed an extensive taxonomy for classifying different types of technologically mediated interactions, or *synthetic experience*, associated exclusively with HMD based systems. His taxonomy is essentially nine dimensional, encompassing causality, model source, time, space, superposition, display type, sensor type, action measurement type and actuator type. In his paper a variety of well known VR-related systems are classified relative to the proposed taxonomy.

Although the present paper makes extensive use of ideas from Naimark and the others cited, it is in many ways a response to Robinett's suggestion (Ref. [26], p. 230) that his taxonomy serve as "a starting point for discussion." It is important to point out the differences, however. Whereas technologically mediated experience is indeed an important component of our taxonomy, we are not focussing on the same question of how to classify different varieties of such interactions, as does Robinett's classification scheme. Our taxonomy is motivated instead, perhaps more narrowly, by the need to distinguish among the various technological requirements necessary for realising, and researching, mixed reality displays, with no restrictions on whether the environment is supposedly immersive (HMD based) or not.

It is important to point out that, although we focus in this paper exclusively on mixed reality *visual* displays, many of the concepts proposed here pertain as well to analogous issues associated with other display modalities. For example, for auditory displays, rather than isolating the participant from all sounds in the immediate environment, by means of a helmet and/or headset, computer generated signals can instead be mixed with natural sounds from the immediate real environment. However, in order to "calibrate" an *auditory augmented reality* display accurately, it is necessary carefully to align binaural auditory signals with synthetically spatialised sound sources. Such a capability is being developed by Cohen and his colleagues, for example (Ref. [27]), by convolving monaural signals with left/right pairs of directional transfer functions. Haptic displays (that is, information pertaining to sensations such as touch, pressure, etc.) are typically presented by means of some type of hand held master manipulator (e.g. Ref. [28]) or more distributed glove type devices [29]. Since synthetically produced haptic information must in any case necessarily be superimposed on any existing haptic sensations otherwise produced by an actual physical manipulator or glove, *haptic AR* can almost be considered the

natural mode of operation in this sense. *Vestibular AR* can similarly be considered a natural mode of operation, since any attempt to synthesize information about acceleration of the participant's body in an otherwise virtual environment, as is commonly performed in commercial and military flight simulators for example, must necessarily have to contend with existing ambient gravitational forces.

3. Distinguishing Virtual from Real: Definitions

Based on the examples cited above, it is obvious that as a first step in our taxonomy it is necessary to make a useful distinction between the concept of *real* and the concept of *virtual*. Our need to take this as a starting point derives from the simple fact that these two terms comprise the foundation of the now ubiquitous term "Virtual Reality." Intuitively this might lead us simply to define the two concepts as being orthogonal, since at first glance, as implied by Fig. 1, the question of whether an object or a scene is real or virtual would not seem to be difficult to answer. Indeed, according to the conventional sense of VR (i.e. for completely virtual immersive environments), subtle differences in interpreting the two terms is not as critical, since the basic intention there is that a "virtual" world be synthesized, by computer, to give the participant the impression that that world is not actually artificial but is "real," and that the participant is "really" present within that world.

In many MR environments, on the other hand, such simple clarifications are not always sufficient. It has been our experience that discussions of Mixed Reality among researchers working on different classes of problems very often require dealing with questions such as whether particular objects or scenes being displayed are real or virtual, whether images of scanned data should be considered real or virtual, whether a real object must look 'realistic' whereas a virtual one need not, etc. For example, with Class 1 AR systems there is little difficulty in labelling the remotely viewed video scene as "real" and the computer generated images as "virtual." If we compare this instance, furthermore, to a Class 6 MR system in which one must reach into a computer generated scene with one's own hand and "grab" an object, there is also no doubt, in this case, that the object being grabbed is "virtual" and the hand is "real." Nevertheless, in comparing these two examples, it is clear that the reality of one's own hand and the reality of a video image are quite different, suggesting that a decision must be made about whether using the identical term "real" for both cases is indeed appropriate.

Our distinction between real and virtual is in fact treated here according to *three* different aspects, all illustrated in Fig. 2. The first distinction is between *real objects* and *virtual objects*, both shown at the left of the figure. The operational definitions† that we adopt here are:

- Real objects are any objects that have an actual objective existence.
- Virtual objects are objects that exist in essence or effect, but not formally or actually.

In order for a real object to be viewed, it can either be observed directly or it can be sampled and then resynthesized via some display device. In order for a virtual object to be viewed, it must be *simulated*, since in essence it does not exist. This entails use of some sort of a description, or *model*††, of the object, as shown in Fig. 2.

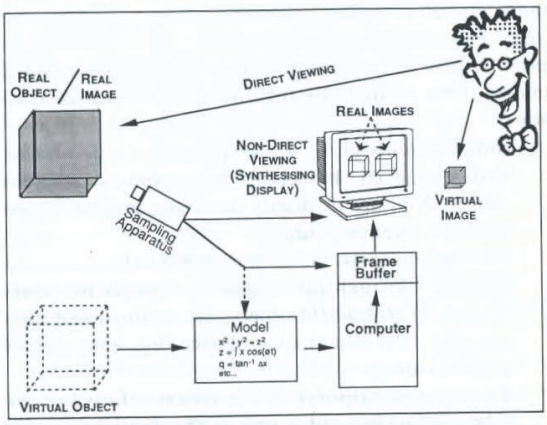

Fig. 2 Different aspects of distinguishing *reality* from *virtuality*: i) Real vs Virtual Object; ii) Direct vs Non-direct viewing; iii) Real vs Virtual Image.

The second distinction concerns the issue of *image quality* as an aspect of reflecting reality. Large amounts of money and effort are being invested in developing technologies which will enable the production of images which look "real," where the standard of comparison for realism is taken as *direct viewing* (through air or glass) of a real object, or "*unmediated reality*" [24]. *Non-direct viewing* of a real object relies on the use of some imaging system first to sample data about the object, for example using a video camera, laser or ultrasound scanner, etc., and then to resynthesize or reconstruct these data via some display medium, such as a (analogue) video or (digital) computer monitor. Virtual objects, on the other hand,

† All definitions are consistent with the Oxford English Dictionary [30].

†† Note that virtual objects can be designed around models of either non-existent objects or existing real objects, as indicated by the dashed arrow to the model block in Fig. 2. A model of a virtual object can also be a real object itself of course, which is the case for sculptures, paintings, mockups, etc., however, we limit ourselves here to computer generated syntheses only.

by definition can not be sampled directly and thus can only be synthesized. Non-direct viewing of either real or virtual objects is depicted in Fig. 2 as presentation via a Synthesizing Display. (Examples of non-synthesizing displays would be include binoculars, optical telescopes, etc., as well as ordinary glass windows.) In distinguishing here between direct and non-direct viewing, therefore, we are not in fact distinguishing real objects from virtual ones at all, since even synthesized images of formally non-existent virtual (i.e. non-real) objects can now be made to look extremely realistic. Our point is that just because an image "looks real" does not mean that the object being represented *is* real, and therefore the terminology we employ must be able carefully to reflect this difference.

Finally, in order to clarify our terms further, the third distinction we make is between real and virtual images. For this purpose we turn to the field of optics, and operationally define a *real image* as any image which has some luminosity at the location at which it appears to be located. This definition therefore includes direct viewing of a real object, as well as the image on the display screen of a non-directly viewed object. A *virtual image* can therefore be defined conversely as an image which has no luminosity at the location at which it appears, and includes such examples as holograms and mirror images. It also includes the interesting case of a stereoscopic display, as illustrated in Fig. 2, for which each of the left and right eye images on the display screen is a real image, but the consequent fused percept in 3D space is virtual. With respect to MR environments, therefore, we consider any virtual image of an object as one which appears *transparent*, that is, which does not occlude other objects located behind it.

4. A Taxonomy for Merging Real and Virtual Worlds

In Sect. 2 we presented a set of distinctions which were evident from the different Classes of MR displays listed earlier. The distinctions made there were based on whether the primary world comprises real or virtual objects, whether real objects are viewed directly or non-directly, whether the viewing is exocentric or egocentric, and whether or not there is an orthoscopic mapping between the real and virtual worlds. In the present section we extend those ideas further by transforming them into a more formalised taxonomy, which attempts to address the following questions:

- How much do we know about the world being displayed?
- How realistically are we able to display it?
- What is the extent of the illusion that the observer is present within that world?

As discussed in the following, the dimensions proposed for addressing these questions include respectively *Extent of World Knowledge*, *Reproduction Fidelity*, and *Extent of Presence Metaphor*.

4.1 Extent of World Knowledge

To understand the importance of the Extent of World Knowledge (EWK) dimension, we contrast this to the discussion of the Virtuality Continuum presented in Sect. 1, where various *implementations* of Mixed Reality were described, each one comprising a different proportion of real objects and virtual objects within the composite picture. The point that we wish to make in the present section is that simply counting the relative number of objects, or proportion of pixels in a display image, is not a sufficiently insightful means for making design decisions about different MR display technologies. In other words, it is important to be able to distinguish between design options by highlighting the differences between underlying basic prerequisites, one of which relates to how much we know about the world being displayed.

To illustrate this point, in Ref. [2] a variety of capabilities are described about the authors' display system for superimposing computer generated stereographic images onto stereovideo images (subsequently dubbed ARGOS™, for Augmented Reality through Graphic Overlays on Stereovideo [5], [21]. Two of the capabilities described there are:

- a virtual stereographic pointer, plus tape measure, for interactively indicating the locations of real objects and making quantitative measurements of distances between points within a remotely viewed stereovideo scene;
- a means of superimposing a wireframe outline onto a remotely viewed real object, for enhancing the edges of that object, encoding task information onto the object, and so forth.

Superficially, in terms of simple classification along a Virtuality Continuum, there is no difference between these two cases; both comprise virtual graphic objects superimposed onto an otherwise completely video (real) background. Further reflection reveals an important fundamental difference, however. In that particular implementation of the virtual pointer / tape measure, the "loop" is closed by the human operator, whose job is to determine where the virtual object (the pointer) must be placed in the image, while the computer which draws the pointer has no *knowledge* at all about what is being pointed at. In the case of the wireframe object outline, on the other hand, two possible approaches to achieving this can be contemplated. By one method, the operator would interactively manipulate the wireframe (with 6 degrees of freedom) until it coincides with the location and attitude of the object, as she perceives it—which is fundamentally no different from the pointer example. By the other method, however, the computer would

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.