A FRAMEWORK FOR RELATING HEAD-MOUNTED DISPLAYS TO MIXED REALITY DISPLAYS

Paul Milgram & Herman W. Colquhoun Jr.
Department of Mechanical and Industrial Engineering
University of Toronto
http://etclab.mie.utoronto.ca/

The Head Mounted Display (HMD) is discussed as a subset of Mixed Reality (MR) displays. A definition of MR is given, in terms of image mixtures along a Reality-Virtuality (RV) Continuum, including the subclasses of augmented reality (AR) and augmented virtuality (AV). In relation to actual task execution, the relative need for local guidance information versus more global planning and navigational information is discussed. A taxonomic framework for classifying MR systems is presented, in terms of not only the RV continuum, but also the degree of centricity of the observer relative to a nominal viewpoint and the extent of control-display congruence. Several practical examples of MR systems are presented, all from the domain of surgery, and for each a volume within the MR design space is proposed.

INTRODUCTION

Our objective in this paper is to discuss the relationship between Head-Mounted Displays (HMD's), as a general display concept, and another class of displays which are with increasing frequency referred to under the heading *"Mixed Reality" (MR)*. We first briefly discuss the concept of MR, and propose a practical classification scheme for different MR applications. In general, we contend that, due to the unique constraint of HMD's that the user must actually *wear* the device, essentially all HMD's can be considered to be contained within the set of MR displays. We also show that it is necessary to take into account the same design factors that constrain MR systems: extent of reality/virtuality, centricity and control-display congruence, when classifying HMD systems.

DEFINITIONS

For a strict working definition of a head-mounted display, we limit ourselves to any display which is mounted on the observer's head and through which information is displayed in a heads-up fashion – i.e. within the normal visual field. In order to extend our discussion, however, we shall relax the first constraint and consider also any display in which data are presented in a heads-up fashion, but not necessarily head-mounted. In addition, although we are interested primarily in visual data, it is important to keep in mind the significance of auditory HMD's, for which information is displayed within the normal heads-up auditory field. Furthermore, although haptic inputs and vestibular information are currently much less likely to be synthesised for display purposes, the presence of natural haptic and vestibular inputs clearly has a major influence in determining the efficacy of practical head mounted display systems.

For an extensive definition of Mixed Reality (MR), the reader is referred to (Milgram & Colquhoun, 1999), within a recent volume (Ohta & Tamura, 1999) which serves as an up-to-date overview of the state of the art in MR research. In brief, as summarised in Figure 1, Mixed Reality refers to any display in which both real and virtual images are combined in some way and in some proportion. Perhaps the best known subclass of MR is the set of *Augmented Reality (AR)* displays, which encompass essentially any case of computer enhancement of real-world images. In the interest of symmetry, a converse class of *Augmented Virtuality (AV)* displays has been proposed, which encompasses all cases for which the background image, or substratum, is virtual and upon which real-world image data are superimposed. One well-known example of AV is the computer graphics technique of texture mapping.

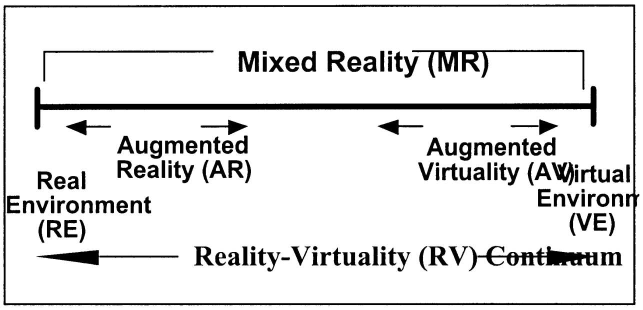

Figure 1: 'Definition' of Mixed Reality, in terms of the Reality -Virtuality Continuum

Although the concepts of AR and AV, and thus MR, appear superficially to be quite simple, further consideration reveals the need to clarify terms which might otherwise seem intuitively obvious, including *what is 'reality'?* and *what is 'virtuality'?*. Our working definition of these concepts, for the purpose of classifying MR displays, relies on the extent to which quantitative spatial information is available (within a computer) about image content. As discussed in (Milgram & Colquhoun, 1999), we take as a necessary condition for a *completely virtual* environment that it be *completely modelled*. This condition derives from the notion that it is not possible to create and simulate all aspects of a virtual world unless we have complete knowledge, i.e. a complete model, about everything that we wish to simulate in that world. In applying to *completely real* environments the converse case of where no quantitative spatial information at all is available, i.e. when an environment is *completely unmodelled*, we are obviously not claiming that a completely real environment can not be modelled. Rather, we are contending that, if no quantitative spatial information about the environment depicted in an image is known (i.e. about what objects are within it, the dynamic properties of those objects, and/or where they are located), yet the environment exists, it can not be virtual, so it must be real.

Of course, neither the case of completely real nor completely virtual environments should strictly speaking be considered as examples of MR, since neither involves any mixing of image types. Instead, as shown in Figure 1, the two extreme cases are useful for establishing the poles of a continuous spectrum of cases, that is, a *Reality-Virtuality (RV) Continuum*, between which lie all cases of Mixed Reality.

MR DESIGN SPACE

Centricity

In the preceding discussion, we considered only the reality/virtuality ratio as a criterion for classifying MR displays. Within that context, it is clear that most HMD's can be classified somewhere along the RV Continuum. At one extreme are those cases in which the wearer is immersed in a totally virtual world, as presented through the HMD, while at the other are those in which the wearer is immersed in a real world.[1] When we go beyond RV considerations and take into account further what kinds of tasks the user must actually carry out, however, it is important to distinguish between operations which require information in support of local task execution and those which require support for global awareness. A well recognised summary of how some of those information needs contrast with each other is presented in Table 1.

Generally speaking, in order to retain adequate global situational awareness (SA), a world referenced view of the observed environment is desirable. The principal problem associated with such a viewpoint, however, is that it can be difficult to acquire the equivalent of an "out-the-window" perspective. Conversely, an ego-referenced view is usually more effective for maintaining good local SA, which is necessary for local guidance and control. The problem, however, is that it is difficult to maintain a good global perspective, necessary for planning and navigation. A thorough treatment of these ideas can be found in the various reports of Wickens and his colleagues, e.g. (Wickens, 1998), (Faye & Wickens, 1995), (Wickens, 1995), who also treat the concept of a continuum spanning the space between purely egocentric and purely exocentric views. In addition, dynamic changes in viewpoint along this *centricity continuum* can be modelled by an effective *tether* joining the location of the (virtual) camera to the observer's nominal viewpoint, or "own ship", within the image. (Wickens, 1998), (Milgram & Colquhoun, 1999).

[1]The question for which cases of total immersion is possible arises here. Clearly total immersion in a totally real world is quite feasible, by virtue, for example, simply of putting on an optical see-through head-set – essentially equivalent to a simple pair of spectacles. Total immersion in a totally wired world is more challenging, however, in light of the near impossibility of extinguishing other non-visual sensory cues, especially vestibular inputs. Most instances of the latter case, therefore, should in fact be considered examples of augmented virtuality (AV).

***Global* Situational Awareness**
• Navigation information
• Location of self
• Location/movement of other system entities
• Commands / directions / instructions from above
***Local* Situational Awareness**
• Target identification
• Target location
• Terrain / object distance estimation
• Cueing presence of relevant objects

Table 1: Situational needs of HMD user.
(Adapted from (Blackwood et al., 1997))

One import consequence of the distinction between different situational needs is that local and global situational awareness are not uniformly achievable along the RV continuum. On the one hand, the closer one gets to the virtual side of the continuum, the more freedom one has to vary one's viewpoint, zoom in and out of the image, etc. On the real side of the RV continuum, conversely, images are typically generated by either a real-time sensing system, such as a remote camera, or some kind of database of stored images. Unless the system is capable, therefore, of providing real-world images which are effectively slaved to the head and body movements of the observer, one does not ordinarily have the same viewpoint flexibility of a virtual world display.

Control-Display Congruence

Thus far we have dealt with the frame of reference within which an observer views her environment, which may comprise any combination of real or virtual images. We have also seen that only with an image whose basis is virtual do we have the flexibility of maintaining both local and global SA. As one moves farther away from the right hand (virtual) side of the RV continuum, however, the constraints of reality impose themselves, primarily due to the fact that unmodelled data can not easily be transformed spatially.

A third factor which must be taken into account in MR design is also related to the observer's viewpoint relative to the world within which she is situated, but pertains more to the means available to the operator to influence that environment, primarily either by manipulating objects within it or by travelling through it. Although several components contribute to this factor, for practical purposes we lump them together here into a single dimension, which we call *Control-Display (C/D) Congruence.* A summary of some of those factors is depicted in Figure 2. (Milgram & Colquhoun, 1999), (Zhai & Milgram, 1998)

Briefly, the three factors shown in Figure 2, from top to bottom, relate to the following respectively:

- the degree of directness with which the user is able to manipulate something, varying from essentially hands-on to the use of some kind of a tool, either real or metaphorical;
- the orientation of the user relative to the display surface, as well as the orientation of the (virtual) camera feeding the display device relative to the task space;
- the number of integration steps (i.e. transfer function) between the operator's input and the system's output.

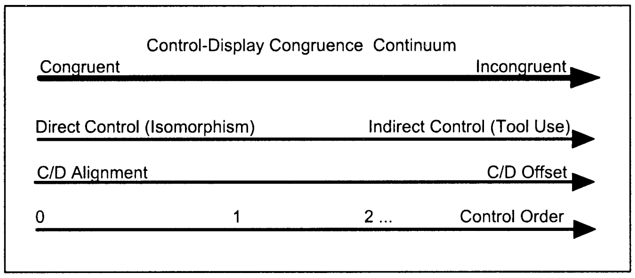

Figure 2: Control-Display (C/D) Congruence Continuum

MR Taxonomy Cube

A graphical depiction of the three factors discussed above is presented in Figure 3. In the remainder of this paper, we endeavour to illustrate the usefulness of this taxonomy for distinguishing a variety of MR applications, with an eye towards its potential usefulness for HMD research.

MR TAXONOMY EXAMPLES

In this section we examine a number of examples of Mixed Reality, and model the location of each respectively within the MR design cube with respect to the three axes discussed above. Due to the wide variety of procedures and technologies associated with it, we consider here only examples from the field of surgery. (Waterworth, 1998)

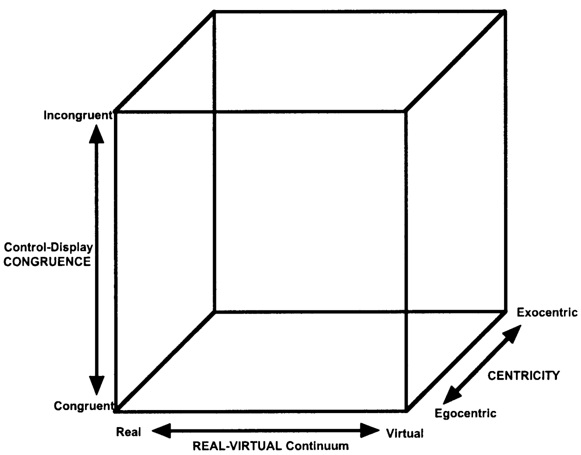

Figure 3: Global Taxonomy of MR Display Integration

a) Open Surgery

As a baseline case, we consider open surgery, which refers to the conventional case in which the surgeon opens the body and uses hands and instruments to operate. The MR taxonomy model is shown in Figure 4a. Since the only inputs to the surgeon, be they visual, haptic or olfactory, are real, this case clearly belongs at the real (left) side of the RV continuum. Furthermore, because the surgeon's actual viewpoint corresponds exactly with the nominal one, this case is at the extreme egocentric end of the centricity continuum. Finally, we have located open surgery very close to the highly congruent end of the third axis, although it has been extended slightly vertically, in light of the fact that the surgeon clearly does not have an unlimited ability to manipulate her tools or to approach all structures from any arbitrary angle.

b) Minimally Invasive Microscopic Surgery

Microscopic surgery is similar to open surgery, in the sense that the surgeon operates directly with her instruments within a space which she has opened up while viewing the world through a magnifying optical link. Because that space is relatively very small and because the degree of dexterity is reduced, the C/D congruence has been extended upwards in Fig. 4b. Otherwise, this case is very similar to that shown in Fig. 4a. One important difference between the two, however, is that with microscopic surgery a viewing device has been placed between the human and the operative field. This permits modification of the real-world image, as shown in the following section. It also fulfils the criteria mentioned earlier for this to be considered a HMD.

c) Minimally Invasive Microscopic Surgery with AR

Whenever some means of modifying an image exists, due to the presence of either an optical or an electronic link between the human and the real world, it becomes possible, in theory, to add computer generated enhancements, i.e. augmented reality (AR). This idea is shown in Fig. 4c, where we have taken the case of Fig. 4b and stretched the design space rightwards. One example of this case is the augmented reality virtual tape measure for making intraoperative 3D measurements. (Kim, Milgram, & Drake, 1997)

d) Minimally Invasive Endoscopic Surgery

Endoscopic surgery is similar to the preceding two cases, in the sense that both are *minimally invasive*. In other words, surgery is performed through either natural body openings or small artificial incisions. The principal difference relative to microscopic surgery, however, is that the surgeon is now viewing the world through a remotely operated camera which has been inserted into the operative field. Fig. 4d, which illustrates this case, differs from Fig. 4b in two principal ways. First of all, because the surgeon is no longer viewing the world along the same line of sight as for microscopic surgery, but instead must view the world through a camera (which may be oriented in a variety of ways relative to the operative site, and which is manipulated not by the surgeon but typically by an associate, while the actual display surface is situated elsewhere), the range of possibilities along the centricity axis is quite broad, as shown. Secondly, because of both the reduced degrees of manipulative freedom and the displacement of the viewing device from the operative site, the range of C/D congruence possibilities is also shown as being quite broad. Note that in both cases, however, we are still operating in a primarily real environment. Another important distinction is that, in the typical case, as noted, the surgeon must look up at the screen while operating elsewhere manually. It is therefore arguable whether this can justifiably be considered a heads-up display, according to the definition given above.

e) Minimally Invasive Endoscopic Surgery with AR

The extension of Fig. 4d to 4e is similar to the modification of Fig. 4b to 4c. By virtue of the fact that in ordinary endoscopic surgery the world is viewed through some kind of video link, it is possible to modify the image by adding computer generated graphics. In theory both global navigation and local guidance information should be useful. However, in practice, both kinds are difficult to obtain, since it is first necessary to be able to track the position and orientation of the camera sensor within the body during an operation. Although the problem is relatively tractable with rigid endoscopes, flexible endoscopes are much more difficult to track. Nevertheless, assuming for now the technical feasibility of acquiring the proper information, Fig. 4e illustrates the effect of augmenting the display with appropriate information. Note that, not only has the MR cube been stretched rightwards, but, more significantly, we should be able to reduce the degree of incongruity through the provision of useful navigation and/or guidance information using AR.

f) AR Surgical Planning and Diagnosis

Fig. 4f illustrates the case in which the surgeon, wearing a see-through HMD[2], is standing next to the patient and essentially looking *through* him, at a preoperatively scanned image (be it ultrasound or MRI or CT) of his internal organs. (e.g. Bajura, Fuchs, & Ohbuchi, 1992) Because in this case the surgeon is standing next to the patient and looking directly at him, the C/D congruence must be very high and the point of view highly egocentric. Since the 3D graphic images must be registered to the real world for such a system to work, the surgeon should be able to move around the patient relatively freely while still viewing the correct image internally. Because in principle any amount of graphic information can be added, we have stretched the cube most of the way across the RV continuum. However, because the basic substratum, the patient, remains real, note that the cube does not go all the way to the right.

CONCLUSION

Our primary aim in this paper has been to define the concept of Mixed Reality displays, as summarised in Figure 1, and to illustrate how the taxonomy presented in Figure 3 can be used to classify a broad variety of practical Mixed Reality display systems. In relation to Head-Mounted Displays, although these can in principle fall essentially anywhere along the RV continuum, the farther to the real end of the continuum they lie, the more conducive they should be for supporting local guidance rather than global awareness. In general, designers should endeavour to avoid the low congruence end of the vertical axis and to maintain an effective viewpoint centricity which is compatible with the local and/or global requirements of the task. Unfortunately, due to operational constraints and/or technological limitations, it is not always easy to satisfy all criteria simultaneously.

[2]Note that the HMD in this case could be either an *optical see-through* device or a *video see-through* device. For practical reasons, it is usually more common for video see-through to be used, since this is more conducive to creating the impression of images lying *inside* the patient, rather than floating on top of and outside of him.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge support from the Institute for Robotics and Intelligent Systems (IRIS) and MacDonald Dettwiler Space and Advanced Robotics, under the “Interactive Intelligent Remote Operations (IIRO) project.

REFERENCES

Bajura, M., Fuchs, H., & Ohbuchi, R. (1992). Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *Computer Graphics (Proc. of SIGGRAPH’92), 26*(2), 203-210.

Blackwood, W. O., Anderson, T. R., Bennett, C. T., Corson, J. R., Endsley, M. R., Hancock, P. A., Hochberg, J., Hoffman, J. E., & Kruk, R. V. (1997). *Tactical Displays for Soldiers - Human Factors Considerations.* Washington, DC: National Academy Press.

Faye, E. L., & Wickens, C. D. (1995). *Strategies for Display Integration in Navigational Guidance and Situation Awareness* (ARL-95-4/NASA-95-1). Savoy, IL: Aviation Research Laboroaty, Institute of Aviation, University of Illinois at Urbana-Champaign.

Kim, M., Milgram, P., & Drake, J. (1997). *Computer assisted 3D measurements for micro-surgery.* Paper presented at the 41st Annual Meeting of Human Factors and Ergonomics Society, Albuquerque, NM.

Milgram, P., & Colquhoun, H. (1999). A taxonomy of real and virtual world display integration. In Y. O. H. Tamura (Ed.), *Mixed Reality: Merging Real and Virtual Worlds.* (pp. 5-30). Tokyo: Ohmsha / Springer-Verlag.

Ohta, Y. & Tamura, H. (Eds). (1999). *Mixed Reality: Merging Real and Virtual Worlds.* Tokyo: Ohmsha / Springer-Verlag.

Waterworth, J. A. (1998). *Virtual Reality in Medicine: A survey of the state of the art.* (WWW Report). Umea: Dept. of Informatics, Umea University, Umea, Sweden.

Wickens, C. D. (1995). *Integration of navigational information for flight* (ARL-95-11/NASA-95-5). Savoy, IL: Aviation Research Lab, Institute of Aviation, University of Illinois at Urbana-Champaign.

Wickens, C. D. (1998). Frame of Reference for Navigation. In D. G. A. Koriat (Ed.), *Attention and Performance, Vol. 16* (Vol. 16,). Orlando, FL: Academic Press.

Zhai, S., & Milgram, P. (1998). *Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices.* Paper presented at the CHI’98, ACM Conference on Human Factors in Computing Systems, Los Angeles, Ca.

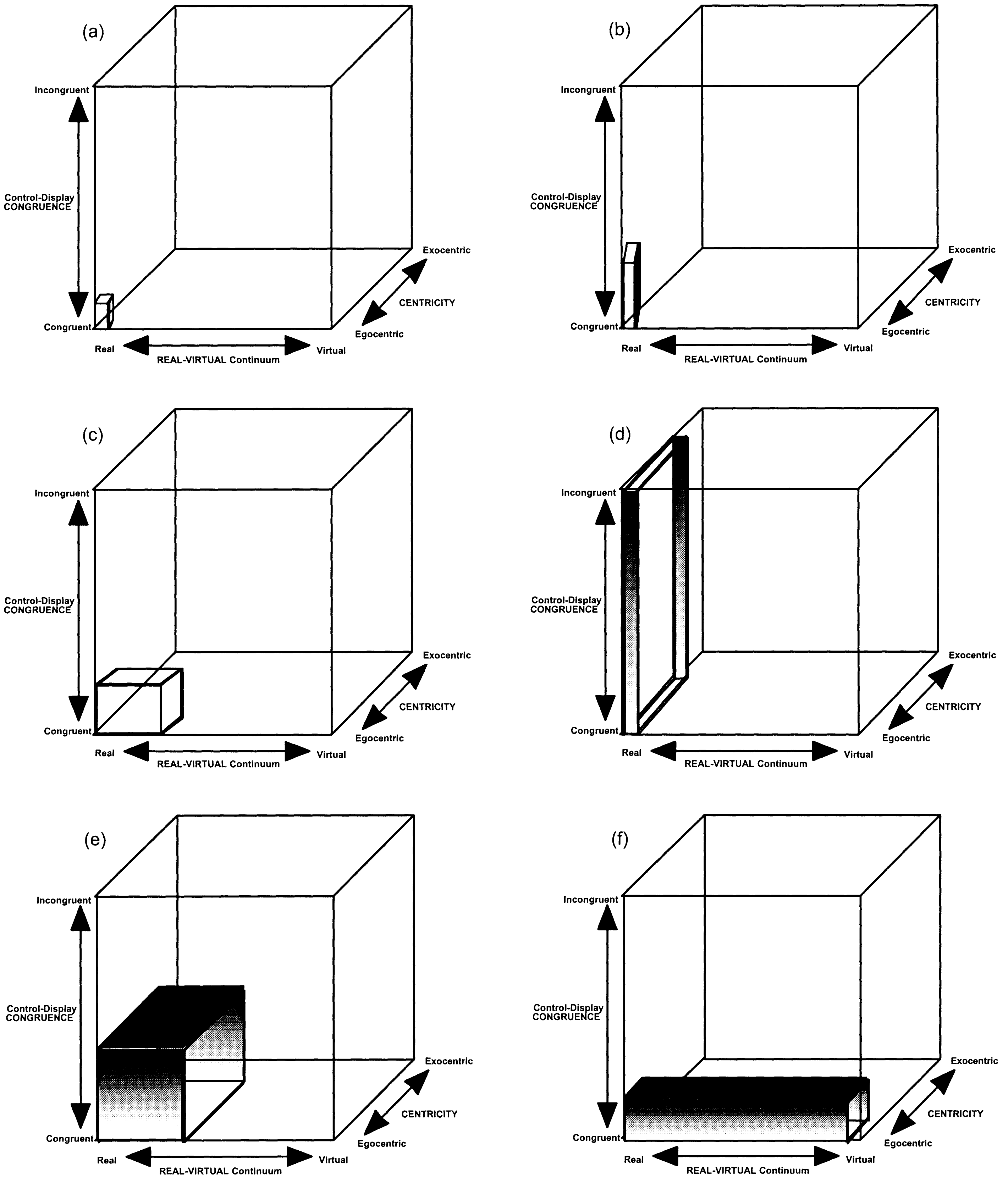

Figure 4: Surgical Examples of Mixed Reality: a) Open Surgery; b) Minimally Invasive Microscopic Surgery; c) Minimally Invasive Microscopic Surgery with AR; d) Minimally Invasive Endoscopic Surgery; e) Minimally Invasive Endoscopic Surgery with AR; f) AR surgical planning and diagnosis.