www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

A USER-CENTERED TAXONOMY FOR SPECIFYING MIXED REALITY SYSTEMS FOR AEC INDUSTRY

SUBMITTED: August 2008
PUBLISHED: February 2011 at http://www.itcon.org/2011/29
EDITOR: Turk Ž.

Xiangyu Wang, Senior Lecturer,
Construction Management and Property Program
Faculty of the Built Environment
The University of New South Wales, Australia
xiangyu.wang@unsw.edu.au

Phillip S. Dunston, Associate Professor,
School of Civil Engineering
Purdue University, USA
dunston@purdue.edu
https://engineering.purdue.edu/CEM/People/Personal/Dunston/

SUMMARY: *Compared to the attention given to Virtual Reality, the related emerging technology of Mixed Reality (MR) has been little explored, until recently, for its applications in the Architecture, Engineering, and Construction (AEC) arena. Observations of the limited lab-based MR applications for AEC highlight the need for a structured methodology that can address suitability and usability issues in order for the full potential of the technology to be realized. The scientific contribution of this paper is the presentation of a comprehensive multi-dimensional taxonomy for specifying MR technology and characteristics, using Milgram's general taxonomy as a springboard. Characteristics including media representation, input mechanism, output mechanism, and tracking technology are progressively disclosed regarding suitability and usability suggestions and context-driven discussion. Some of the characteristics are presented as continuums. Understanding the relationship between a task purpose and the technology's actual position in the continua may help developers identify usability weaknesses and strengths, with the potential to suggest alternative solutions. The taxonomy provides a thorough classification, enumeration, and discussion of MR technology, which can be used to emphasize the importance of user-centered approaches to MR-based application development. The benefit to developers and researchers is a structured framework for developing prototype MR systems for industrial and experimental use. The benefit for practitioners is an enhanced understanding of how MR-based computer interfaces might be exploited to facilitate their work.*

KEYWORDS: *AEC; Augmented Reality; cognitive load; continuum; Mixed Reality; taxonomy, interactive systems*

REFERENCE: *Xiangyu Wang, Phillip S. Dunston(2011) A user-centered taxonomy for specifying mixed reality systems for aec industry, Journal of Information Technology in Construction (ITcon), Vol. 16, pg. 493-508, http://www.itcon.org/2011/29*

1. INTRODUCTION

A variety of computer technologies and computer science techniques are now used by researchers aiming to improve aspects of architectural design, engineering analysis, construction and maintenance. Mixed Reality (MR), as formally defined by Milgram et al. (1994, 1999), is a special class of Virtual Reality (VR) related technologies for creating environments wherein real world and virtual world objects are presented together in a single display (see Figure 1). The state-of-the-art in MR today is comparable to the early years of VR in that many systems (mainly Augmented Reality systems) have been demonstrated but few have matured beyond lab-based prototypes (Azuma et al. 2001). The ever-increasing power of hardware, rendering applications and tracking technology has primarily motivated the exploration and creation of MR-based systems to benefit manufacturing tasks (Ong and Nee 2004). However, in light of the fact that manufacturing shares many common aspects with the architecture, engineering, and construction (AEC) arena, comparatively little effort has been expended on the MR technology application in AEC. Human-computer interfaces that blend a view of an existing site with relevant design or field information should be an attractive class of technology for the AEC industry. Complex design, construction, and maintenance tasks in the AEC industry present excellent opportunities for exploiting MR-based technologies because there is a significant need for good training tools and the means to access large amounts of technical documentation and information. The promise of MR techniques is in the ability it affords to bring additional digitally managed information as seamlessly as possible into the view of the user while engaged in the work task (s). Mixed Reality has potential to completely transform the current way AEC practitioners perform tasks, but success requires that utilization of the technology does not bring with it an undesirable burden in terms of physical and especially mental requirements.

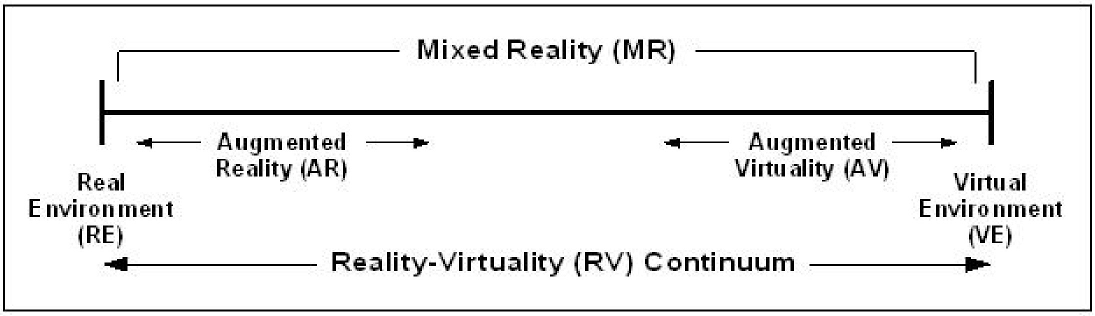

FIG. 1: Reality-Virtuality Continuum (Milgram et al. 1994; 1999)

As shown in Figure 1, MR encompasses the continuum of possible combinations of elements from both virtual and real environments, the continuum between fully real and fully virtual. Augmented Virtuality (AV) and Augmented Reality (AR) are the two major subsets lying within the MR range of the Reality-Virtuality (RV) continuum. AV describes cases where some real object is inserted into a predominantly computer-generated environment, which has been explored in safety programs and manufacturing (Caterpillar Distributed Virtual Reality System 1996). The converse case on the MR continuum is AR, which is a technology or an environment where the additional information generated by a computer is inserted into the user's view of a real world scene. AR allows a user to work in a real world environment while visually receiving additional computer-generated or modeled information to support the task at hand.

A review of the limited MR-based applications in AEC and other domains reveals the lack of a scientifically structured framework for establishing the link between AEC task requirements and the suitability of MR technologies to support such tasks. That is, a taxonomy is needed to facilitate development of MR systems tailored for AEC industry applications. A specification and evaluation methodology is an important first step toward this objective, and it serves as a foundation for evolving methods to develop performance-enhancing MR-based applications by means that are more effective than ad hoc approaches. In order to map MR technologies to an AEC task, MR technological components must be classified, and the task requirements analyzed. Then a match must be established based on user-centered concepts such as perception, cognition, usability and ergonomics, etc., which are not mutually exclusive concepts Using the fundamental concepts of MR as a starting point, the classification approach presented in this article goes further to provide a methodology for devising user-centered, performance enhancing systems and thus can facilitate technology adoption. Real, usable MR systems may then be designed. The more specific focus of this paper is the issue of cognitive load imposed by MR technologies.

2. RELATED WORK

MR system designs may vary considerably due to the variety of media representations, input, output, and tracking (for spatial registration) methods. Categorizing of these options enables the relative assessment of the various MR system configurations. Milgram et al. (1999) formulated a global taxonomy to describe how the "virtual" and "real" aspects of Mixed Reality (MR) environments are merged and experienced based on three continua: Reality-Virtuality, Centricity, and Control/Display (C/D) congruence continua, with the ultimate objective of clarifying conceptual boundaries existing among noted research. Azuma (1997) categorizes AR systems in terms of application domains, sensor technology and display features. Boud et al. (1999) take a slightly different focus looking at the level of interaction with real world objects and how this facilitates learning. Haniff et al. (2000) described AR systems in terms of context (context aware or context free) and granularity. Most AR systems are context-aware and rely on sensors, however, context-free systems are useful for activities not requiring location information (Baber et al. 1999).

Because of its comprehensive nature, Milgram’s taxonomy provides fertile ground for cultivating a specification framework. While Milgram’s taxonomy is useful in discussions of MR systems as they exist, an even more user-based extension of the taxonomy is needed for specifying MR applications and for stimulating valuable human-computer interaction (HCI) research. Our taxonomy is motivated by the need to distinguish among the various technological components necessary for realizing and researching MR-based systems. The goal here is to minimally modify and complement, rather than replace, Milgram’s original MR continuum by considering media representation features and tracking technology, and discussing the input and output features separately.

3. TAXONOMY FOR SPECIFYING MR SYSTEMS

This first component of the proposed specification taxonomy serves as a framework for definition, discussion, research, development, analysis, comparison, and usability evaluation. Four attributes were identified as the dimensions for the taxonomy: media representation, input mechanism, output mechanism, and tracking technology. Since the authors conclude that AR has the greater relevance for supporting the performance of persons functioning in real environments, the taxonomy discussion generally assumes AR applications. There reader may readily draw parallels for AV. Cognitive load is the central and critical concept involved in the investigation and discussion throughout this paper. Cognitive load involves the understanding of how much information can be retained in short-term working memory before those information decay. A common cited example of this principle is the use of 7-digit phone numbers, based on the theory that most people can only retain seven “chunks” of information in their short-term working memory. Cognitive load theory has been devised to provide guidelines that aim to assist in information presentation in an encouraging manner for people to optimize intellectual performance (Sweller et al.1998)

3.1 Media representation

MR systems tend to use more varied media representations than is typical in VR environments. In an AR scene, digital content augments a person’s knowledge or understanding regarding some aspect(s) of the real environment and is typically registered onto the real background. Considering the many different forms of computer-generated information which could be presented via an electronic display, a *virtual augmenting content continuum* consisting of sequenced options is shown in Figure 2. Each option represents a class of information with common characteristics. For instance, “Platform, Tablet, and Screen” is a more complicated format than a simple “indicator” and is a richer media. It could consist of a logical collection of “indicators” that form a schematic and simple artifact like a control panel.

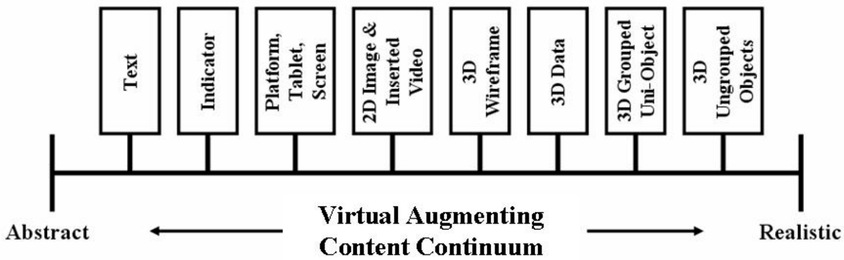

FIG. 2: Virtual Augmenting Content Continuum as Applied to AR

The virtual augmenting content continuum shown in Figure 2 depicts a transition from abstract (schematic symbolism) to concrete (realistic) types of media representation. Moving from left to right across the continuum, the *cognitive load* required to interpret the virtual information is surmised to generally decrease. That is, the form of information requires less transformation to be comprehended. It is apparent that realistic representations convey more detailed and richer information that can speed human comprehension in many situations thus augmenting the cognitive process of the work activity. This continuum arrangement, however, is not meant to imply that concrete media representation is always superior to abstract media representation because each type of representation has its own advantages and appropriate application domains. Concrete representations, for example, may involve too much detail for some cases. Also, two or more types of media representation presented together on a single display—a hybrid representation—may be very useful in some situations and especially when the MR system supports multiple functions.

The real background acts as the registration base for the virtual augmenting content. The term “real” here not only refers to “real” world objects, but encompasses any kind of sampled image data, and includes as the primary example video, but also photographic images (visible or infrared), radar, X-ray and ultrasound, as well as laser scanned 2D and 3D data (both range and light intensity data) (Milgram et al. 1999). The major types of representation of the real background are identified and ordered to form the *augmented background continuum* as shown in Figure 3 according to the extent of realism.

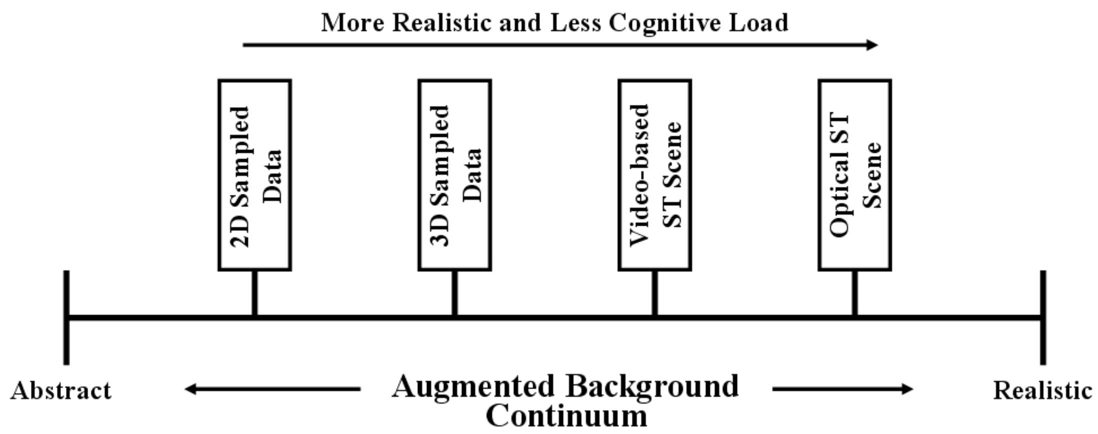

FIG. 3: Augmented Real Background Continuum as Applied to AR

The continuum in Figure 3 goes from the extreme of 2D sampled data with the least fidelity how the real environment is visually perceived by a person to the other extreme of an optical see-through scene with essentially complete fidelity to the real environment. An optical see-through (ST) scene of a real environment is directly observed by human eyes (that are binocular) for accurate depth judgment. A video-based see-through scene is a video stream of a real environment processed by computer. The video could be distorted by the camera lens and does not provide the stereoscopic sense that is normally relied upon for depth judgment. Therefore, the human viewer needs to apply other mental strategies to understand the depth relationships, which induces additional cognitive load. Since it is computationally expensive to render sampled data in a realistic way, simplified representations of different properties (e.g., color, texture, geometry etc.) may be applied. Even though computer software can accurately distinguish these sampled data and accurately render them with corresponding representations, users still have to understand what each representation stands for and also the depth cues are difficult to apprehend. All these necessary interpretations require more cognitive transformations than a video-based ST scene. Also the richer fidelity of 3D sampled data makes it more realistic and less cognitively demanding (i.e., distracting) than 2D sampled data, an important consideration to maintain attention to quality or safe operations. Therefore it is reasonable to infer cognitive load increases as the representation of the real background becomes less realistic as shown in Figure 3.

AR systems have augmenting virtual content and an augmented real background displayed together on a single display. Figure 4 depicts the combined cognitive load distribution for interpreting both the virtual and real entities in a scene if the two continua are considered together in combination. Since the cognitive load trends for each continuum have already been revealed, it is a reasonable assert that the highest cognitive load area is located around the point P1 representing the scenario where text and 2D-sampled data are combined in a display (see Figure 4). Likewise, it is inferred that the lowest cognitive load region should be around the point P2 representing the scenario where 3D ungrouped object and optical see-through scene are combined in a display. The closer the combination is graphically located to P1, the greater will be the cognitive load imposed by the media representation. Likewise, the closer the combination is graphically located to P2, the lesser the cognitive load involved.

However, it is difficult to compare the cognitive load requirements between any combination located along the line from the representing scenario (where a 3D ungrouped object and 2D sampled scene are combined in a display) to the representing scenario (where text and optical see through scene are combined on a display) because there is a tradeoff in cognitive load associated with the virtual and real content. Both points represent a combination of high and low cognitive load requirements. The notion of a quantitative comparison is further complicated by the absence of a common absolute scale on the two axes. The straight band between P1 and P2 could roughly depict the intermediate transition along which the different combinations involve different extents of cognitive load imposed on a user in comprehending the scene. The relative cognitive load of a combination that does not directly locate on the band, might be assessed by its perpendicularly projected intersection with the band. To do so suggests even that some combinations on opposite sides of the band can invoke equal cognitive loads. This supposition is, of course, speculative, but the chart offers an interesting basis for hypothesizing outcomes of tests of relative cognitive load.

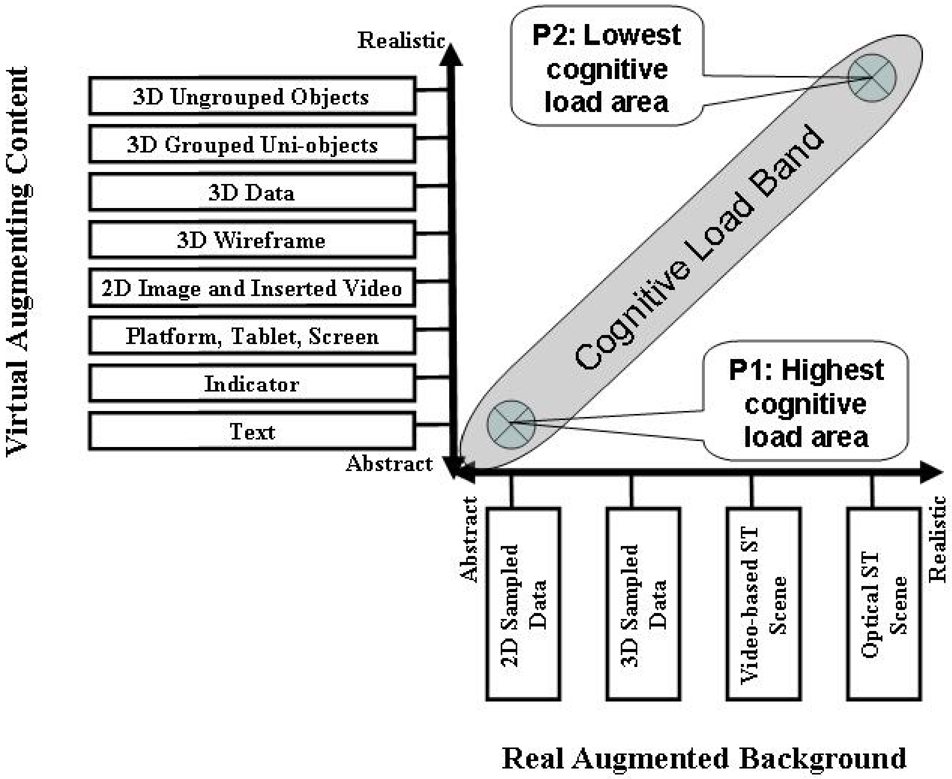

FIG. 4: Hypothetical Cognitive Load for Different Content and Background Combination

3.2 Input Mechanism

The input mechanism—the real world device(s) for directly interacting with the augmenting content—may create a number of problems for the user if it is designed inappropriately. The major aspect considered in this framework is the extent of mismatch, or semantic differences between real and virtual objects, that may leave users confused about functionality. The extent of mismatch will influence the intuitiveness of the input mechanism and play a significant role in HCI performance. That is, the most intuitive input mechanism is one wherein the features of the physical input device most closely correspond to its virtual counterpart as it exists and functions in the virtual world.

An object in 3D space has six degrees of freedom (DOF). To fully control such an object requires an equal number of degrees of freedom in the input signal. There are a number of ways that six dimensional (three translational and three rotational) controlling signals can be generated. The extent of mismatch is based on the lack of consistency between the generation of such 6-DOF signals and the response in the system. Some of the more sophisticated physical input devices (e.g., gestural, tangible, etc.) are selected or designed in such a way that they can be viewed as objects in 3D virtual space, with each of their six dimensions mapping onto fewer dimensions of the input device. In contrast, some of the input devices (e.g., mouse, keyboard, etc.) can generate controlling signals for six degrees of freedom. For more detailed comparison of these input paradigms, readers are referred to the work by Hinckley et al. (1994).

In order to model the extent of mismatch (naturalness) of the input mechanism, an input mechanism continuum is conceived as shown in Figure 5. The continuum is characterized by a trend of decreased extent of mismatch corresponding to increased intuitiveness. The term *mechanism* here attempts to capture both physical tool and underlying technology function. The more intuitive the device for the users, the more the users can engage themselves in effective work with greater attention on the work task and less mental transformation required to resolve the mismatch between the mechanism device and the input result. Thus, in general, cognitive load (mental transformation requirement) can decrease by minimizing the mismatch between the control device and the task at hand.

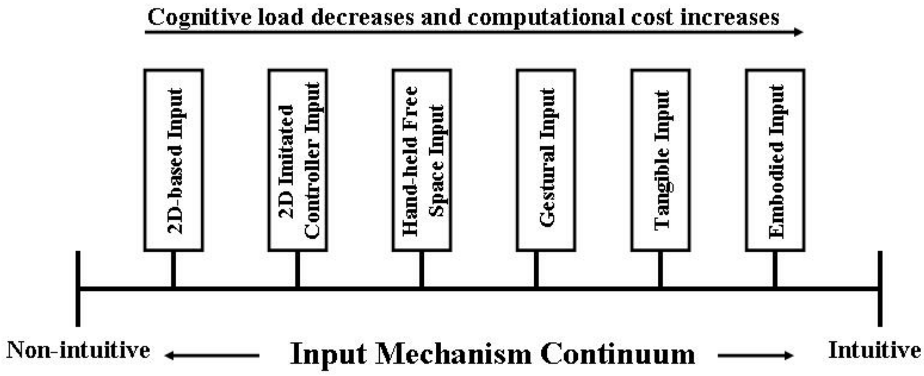

FIG. 5: Input Mechanism Continuum as Applied to AR

Transition toward the intuitive end of this input continuum also reveals a decrease in the significance of the role of a device as an intermediate mechanism between users and the displayed content (Davies 1996). For example, intermediate devices such as joysticks actually place themselves between users and environments, thus requiring more user cognitive mapping to perform tasks. In cases where interaction with virtual models implies some standard physical implement (e.g., drilling implies a drill, hammering implies a hammer), tangible interfaces, often referred to as real-world physical props, may be used as opposed to unique virtual tools, coupled with synthetic force-feedback. Tangible interfaces support the interaction with virtual objects by directly interacting in the physical world with somewhat similar real objects or tools. Real-world props serve as excellent input devices when the corresponding real-world tasks explicitly require the use of such tools. From traditional human factors research, Hinckley et al. (1994) suggest that developers avoid integrating traditional computer input devices such as keyboards and mice in combination with 3D, free-space input devices, because users may have a difficult time switching between multiple devices, especially when immersed in HMD-based environment. However, voice input, a more direct, natural form of interaction, can be achieved as an extra input to increase input capability of the whole system. Finally, the more intuitive input mechanisms involve more sophistication, which leads to greater computational requirements and thus implied cost. From a functional standpoint, high computational intensiveness can also result in system lag, crippling user interaction by disrupting the perceived link between user actions and system responses.

3.3 Output mechanism

The MR system display, discussed here primarily as the *output mechanism*, can be generally classified as, visual, acoustic, tactile, etc. Different classes of output mechanism are discussed in relation to the situational awareness required or preferred for the work task(s) that must be performed by the user. Both efficiency and safety would be chief concerns in field tasks. This framework focuses on display categories that are relevant for feasible use of AR systems in the AEC arena.

Mapping user scenarios and tasks to appropriate display characteristics is essential for the selection of truly useful visual displays. A spatially relevant frame of reference or point of view is of primary importance. Users must have a clear, unobstructed, detailed view of the virtual information as well as real objects or tools. Viewpoints as noted by Milgram et al. (1999) are generally considered to be either egocentric, or exocentric. One way to determine mapping from display to task is through the combined requirements/preferences of *centricity*, *distortion*, and *directness of view* that the task requires and that which the system intends to convey. Depending upon centricity requirements of the task, users may need an egocentric (for object identification, location and distance estimation), exocentric (for navigation information, location of self and other entities), or simultaneous views. In the latter case, a user may be provided with greater awareness than would normally be possible. Distortion refers to the physical image-deformations that are caused by the applied optical elements. Thus, undistorted graphics as perceived by the viewer constitute an orthoscopic mapping (1:1) between size and proportions of displayed images and real environment. Directness of view refers to whether the applied optical element offers a direct or indirect view on the real environment. An optical see-through HMD offers direct view on the real environment, and conversely, a video-based see-through HMD is an example of an indirect view due to the intermediate processing that occurs between video capture and display. Given a particular Mixed Environment (ME) application, representative user tasks may implicitly suggest the mix of centricity, distortion and directness of view required. In turn, these requirements may lend credence to a particular display type. Figure 6 shows how the task requirements regarding centricity, distortion, and directness of view influence the selection of display type. This characterization of displays does not lend itself to a simple continuum that may be referenced as with media representation and input mechanism.

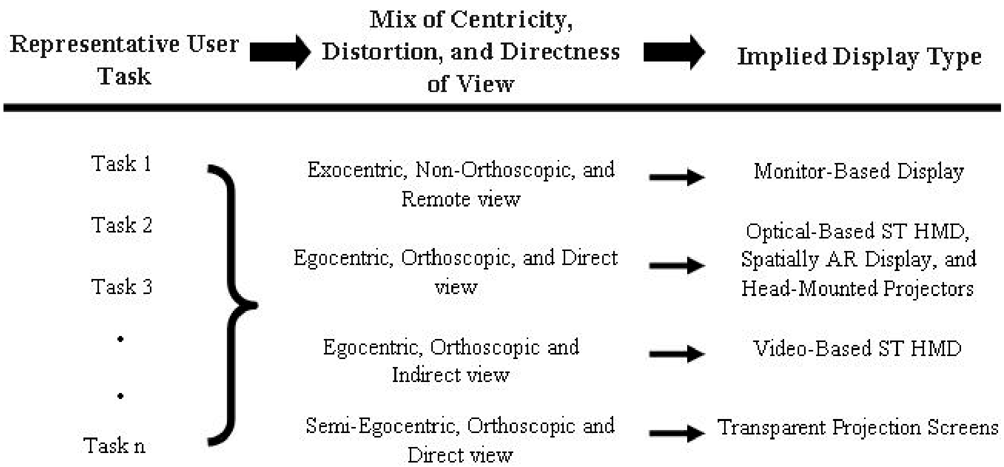

FIG. 6: Task Centricity Requirements' Influence on Display Type (Wang and Dunston 2006)

Another way to conceive the mapping from display to task is through how well the *presence* requirements of the tasks are met in the AR display. Presence is defined as the extent to which the observer is intended to feel connected to the displayed scene from within and goes beyond mere egocentricity (Milgram and Takemura 1994). Presence frequently becomes an issue because it is thought to correlate with improved task performance in virtual environments. A better sense of presence does not show its advantages in all types of tasks because some AR tasks or scenarios do not require a strong inside-out sense of the surroundings. A proper perspective for AR may be that of not only feeling a connection to the displayed scene but also a connection to the virtual content that has been merged into the real scene, i.e., a sense of sharing presence in the scene. Each task should have a corresponding appropriate display type. The output mechanism in AR can be formulated as a continuum from non-immersive to immersive as shown in Figure 7 according to Milgram and Takemura (1994). However, not all of the display classes identified above were considered and occupy a position in the continuum as presented. Transitions in the output continuum are made mainly according to the sense of “real” presence.

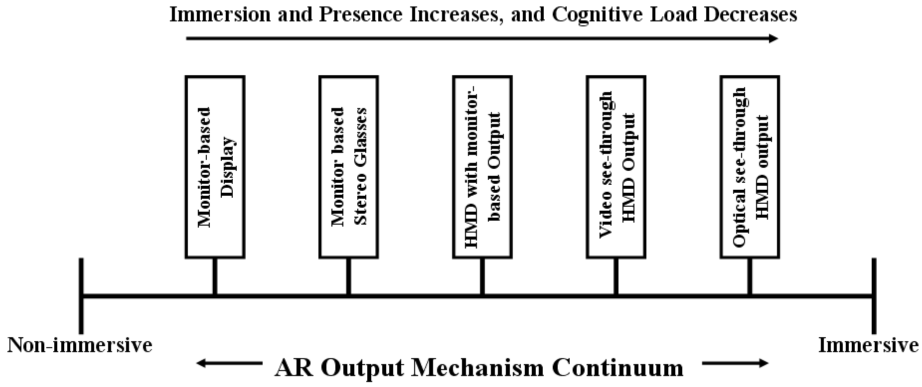

FIG. 7: AR Output Mechanism Continuum Based on Extent of Presence (adapted from Milgram and Takemura 1994)

3.4 AEC Context Considerations and Examples

Tasks and operations in AEC are diverse in nature and complicated. However, they have some features in common. The authors have arrived at a set of AEC context characteristics which contribute to the development of the MR technology taxonomy presented in this paper. Although the focus of this paper is on classifying the combinations of the major technological components of MR systems from the perspective of mental and physical human factors, the authors also have identified four significant factors in AEC application contexts that should be considered in the design of MR systems. The discussion of these AEC context-driven factors are out of the scope of this paper and have been developed in separate papers. However, in order to give some AEC background for this discussion, the following paragraphs give an overview of these context-driven factors. Examples are also presented to illustrate how the AEC context characteristics drove the development of the taxonomy. These four influencing factors are task mental requirements, working environment, physical disposition, and hands occupation. Each of these factors can influence the feasibility and usability of the four MR technological components (media representation, input mechanism, output mechanism, and tracking technology). Implications of these four factors are described in the following sections.

3.4.1 Task Mental Requirements

Task mental requirements have to do with perceptual and cognitive tasks as mentioned above. Perceptual tasks include such actions as “detect”, “receive”, “inspect”, “scan”, and “observe”; cognitive tasks include such actions as “calculate”, “interpolate”, “categorize”, “itemize”, and “compute”; motor tasks include such actions as “activate”, “lower”, “close”, “move”, “connect”, and “press.” MR system design should strive to bolster human performance by compensating for limitations in the user’s mental capabilities (e.g., working memory limits, attention allocation, and bandwidth capacities). For example, text is suitable to be used in descriptive, instructional and procedural tasks, where users need much explanatory information from a database rather than geometrical information. The function description can be attached to each button on the control panel of one heavy equipment in order to help novice operators recall what each button is supposed to do without referring back to the manual or asking other workers. As another example, indicators are more likely used in highlighting the to-be-identified object or feature, and pointing out operation direction or sequence. Wearing an HMD, the project manager can have a virtual route plan for heavy equipment moving direction overlaid onto his real view of the equipment. Such a route can be analyzed for potentials of productivity improvement.

3.4.2 Working Environments

The factors involved in the working environment that need to be considered in regard to MR systems include situational awareness requirements, indoor/outdoor location, noise level, work area hazards, working volume, etc. The working environment may put special limitations on adoption of certain components of MR technology. For example, if the task is to be performed under potentially dangerous conditions, use of solid virtual object, large amounts of text and large size image should be avoided. In such conditions, workers need to keep high situational awareness and update the surroundings in real time. Use of these representations may occupy too much space of the worker's view of the real world.

3.4.3 Physical Disposition

The physical disposition of the work task should be considered in terms of such factors as motion, body position, etc. Many construction tasks require workers to move around, which forces MR systems to be implemented on portable or wearable devices. Construction environments may require both mobile and stationary AR systems. The physical disposition may determine the appropriateness of certain interaction tools or mechanisms. For example, one method of providing a high-fidelity media representation uses sequences of video clips. This method appears to be well-suited for mixed environments (MEs) where people in the worksite perform tasks such as receiving training for procedure guidance and construction methods with minimal interaction. As another example, for MEs involving little or no manipulation, sitting users, or limited facility space, tracking by hand may be more appropriate and comfortable for extended use. Workers in a limited working volume such as an HVAC duct tunnel or special equipment may need a hand-based input mechanism to interact with the virtual information presented in front of the real scene view.

3.4.4 Hands Occupation

Performing a typical construction task augmented by MR technology requires interaction with digital information via a certain input metaphor as well as manual actions for the task at hand. Thus MR may increase both mental and physical occupancy. For example, if hands are required to perform work tasks and to interact with virtual information simultaneously, mental transformation involved in switching between those two tasks should be considered. The less the mismatch between the task object and input mechanism, the less mental transformation is required. Metaphors which have a number of mismatches (semantic differences between real and virtual worlds) may leave users confused about available functionality and mappings. If the virtual object used for tangible input is seamlessly attached to the tool or object in the manual task, then what one sees virtually is what one actually performs. If hands are only required for the input mechanism, the selection of input mechanism should be based on naturalness or intuitiveness, which is discussed in the following section.

4. MR GLOBAL TAXONOMY

Specific MR applications are more thoroughly characterized by Milgram and Colquhoun (1999) in a global (three-axis) continuum: Reality-Virtuality, centricity, and control/display congruence. The authors also propose a global continuum that is geared toward specifying MR systems based on the continua of the three MR technological components (media representation, input mechanism, and output mechanism) presented above and as depicted in Figure 8. Theoretically, any MR system can be characterized/located in this global continuum with clear system configurations identified. For example, the AR CAD system (Dunston et al. 2002) has 3D solid virtual object as the augmenting content, video-based ST scene as the augmented background, tangible device as the input mechanism, and video-based ST HMD as the output mechanism. It can be located to occupy some range in the global continuum as depicted in Figure 8. Three more examples are also given in the figure to demonstrate how the global continuum is used to characterize specific MR applications.

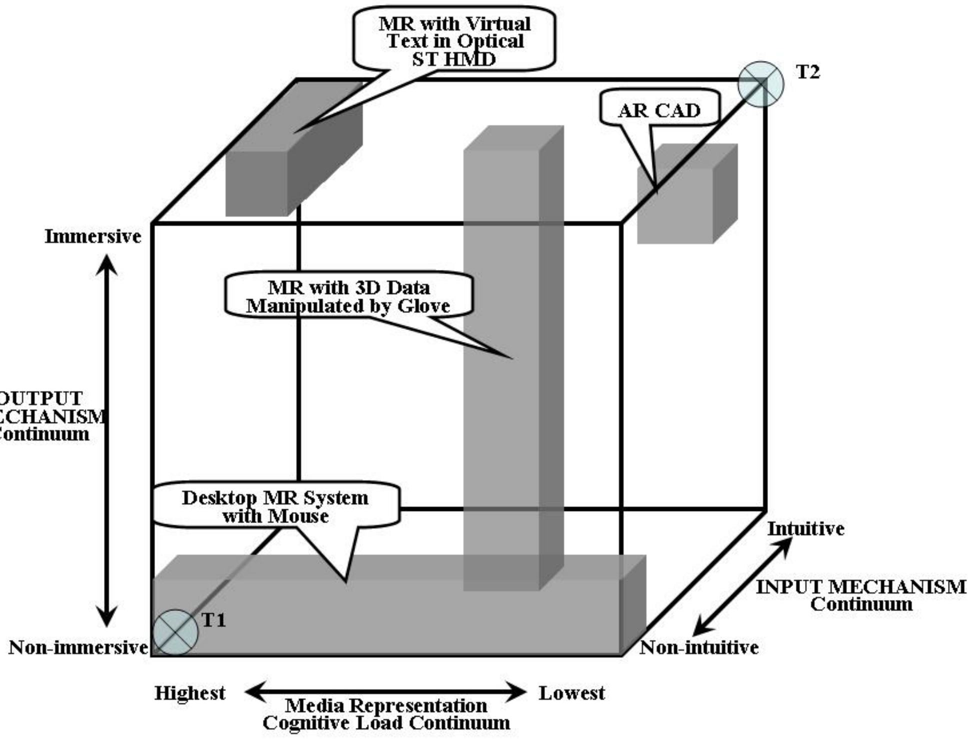

FIG. 8: Approximate Locations of MR Examples within the MR Global Taxonomy

Since the cognitive load trends for each continuum have already been revealed in the continua of each component earlier, the global continuum implies comparison of magnitude of cognition cost for identified MR systems. It is a reasonable assertion that the highest cognitive load area is located around the corner T1 representing an MR system where text, 2D-sampled data scene, 2D mouse, and desktop monitor are used (see Figure 8). Likewise, it is inferred that the lowest cognitive load area should be around the corner T2 representing an MR system where 3D ungrouped objects, body-centered input device, and optical see-through HMD are adopted. The closer an MR system is graphically located to T1, the greater number of mental transformations is imposed on the user. Likewise, the closer the combination is graphically located to T2, the fewer mental transformations are involved. The straight band between T1 and T2 could roughly depict the intermediate transition along which the different combinations involve different extents of cognitive load imposed on users for apprehending the ME. Similar to the earlier suggestion regarding Figure 4, the relative cognitive load associated with any point (combination) in the global continuum might be determined by the projected location along the T1-T2 band.

The global continuum also does imply relative costs of MR systems. The notion of a quantitative comparison is complicated by the absence of a common absolute scale on the three axes. Nevertheless, this graphical scheme does begin to illustrate the relative tradeoffs in cognitive and computational options related to media representations. The lowest overall cognitive load the system is designed to achieve will require the most computational resources, which in turn involves higher development costs. The technological components involved in MR systems located at T1 consist of virtual text that demands the lowest rendering capability, a non-immersive output mechanism, and non-intuitive input mechanism. The low computational resource demand and device selection makes such an MR system cheaper and easier to build. Conversely, the 3D virtual object, immersive output mechanism, and intuitive input mechanism for the MR system located on T2 all demand high computational resources, which makes the MR system more expensive and also more challenging to develop.

5. TRACKING TECHNOLOGY

As indicated at the beginning of Section 2, the fourth technological component of a functional MR system is the method of tracking that is employed. Accurate registration and positioning of virtual objects in the real environment requires accuracy in tracking the user's head (viewing perspective) as well as sensing the locations of real objects in the environment. Typically, the greatest single obstacle to building effective field-ready AR systems is the requirement of accurate, long-range sensing and tracking. For the present discussion, usability issues in selecting a tracking technology are addressed.

AR systems can be classified as context-free and context-aware according to the need of calibration. As a context-free system, video-based calibration-free approaches (Kutukalos et al. 1996; Iu et al. 1996) avoid the need for position tracking and can overlay the video of real entities with digital content without knowledge of camera parameters. However, most of the tracking technologies have been developed for context-aware AR approaches. For mixed environments that require large user-roaming areas, sophisticated ultrasonic tracking systems are one option which may be used to increase user range. Magnetic trackers are typically limited to a range of a few meters, yet do not require a clear line-of-sight. Magnetic tracking is suitable for scenarios with small working volumes and minimum electromagnetic interference. Body-mounted magnetic transmitters are powered through small cables, resulting in some user tethering. Ultrasonic, optical, and infrared tracking systems avoid tethering and thus allow greater freedom of motion. However, a possible tradeoff is the fact that these systems are susceptible to body interference since line-of-sight is required. Optical systems can be well suited for real time applications, especially when compared with magnetic position trackers that have lower data rates (Azuma et al. 2001). Since each of the available technologies has weaknesses, future tracking systems that can meet the high accuracy and mobility demands of AR will probably be hybrid systems, such as a combination of inertial and optical technologies (Azuma 1997). State et al. (1996) combined active-target magnetic and vision sensing. Azuma and Bishop (Azuma and Bishop 1994) developed a hybrid of inertial sensors and active-target vision to create an indoor augmented reality system. A more recent commercial example is the InterSense IS-1200 VisTracker™, a wide-area, wearable, 6-DOF vision-inertial hybrid self tracking system designed specifically for precision Augmented Reality and mobile computing applications. The continued effort of commercial developers to create such systems bodes well for eventual transfer of AR technology to industrial applications.

6. MR COGNITIVE LOAD INDEX

The global continuum discussed in Section 4 provides a qualitative means to roughly compare the total cognitive load among different MR systems. In an effort to merge consideration of the technological components which have scalable cognitive load implications, the authors have conceptualized an MR Cognitive Load Index (MR CLX) scheme to quantitatively measure and compare the cognitive load for different MR systems. The MR CLX is a multi-dimensional rating procedure that provides an overall cognitive load score based on a weighted average of ratings on four factors (subscales) originated from the three technological components of MR systems discussed above: media representation (consist of augmenting content and augmented background), input mechanism, and output mechanism. A definition of each subscale is provided in Table 1. The algorithm behind the development of the MR CLX refers to the NASA Task Load Index (NASA TLX) which is commonly used to rate and compare systems or methods with respect to the work load associated with a given task.

For more details of the statistical algorithm of the NASA TLX, the reader is referred to the paper of Hart et al. (1984). The NASA TLX deals with the total task load involved in interaction with a system (including mental, physical, etc.), however, the MR CLX merely assesses the total cognitive load involved in interaction with an MR system. Also, the contribution made here is the identification of the four major factors that might be the sources of cognitive load, and the way they are considered together to determine the magnitude of total cognitive load.

Although it is clear that definitions of cognitive load do indeed vary among experimenters and among subjects (contributing to confusion in the cognitive load literature and between-rater variability), the specific sources of loading imposed by different tasks are an even more important determinant of cognitive load experiences. Thus, the MR CLX combines subscale ratings that are weighted according to their subjective importance to raters in a specific task, rather than their a priori relevance to raters' definitions of cognitive load in general. The MR CLX is a two-part evaluation procedure consisting of both weights and ratings. The degree to which each of the four factors contributes to the cognitive load of a specific task is determined by raters' responses to pair-wise comparisons among the four factors. Magnitude ratings on each subscale are obtained after performance of a task. Ratings of factors deemed most important in creating the cognitive load of a task are given more weight in computing the comprehensive cognitive load score, thereby enhancing the sensitivity of the scale.

TABLE 1. Definitions of Rating Scales in MR Cognitive Load Index

RATING SCALE DEFINITIONS

Title	Scale Range	Descriptions
AUGMENTING CONTENT (AC)	*Abstract – Realistic*	How much mental and cognitive activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.) in processing information from the augmenting content? Was the augmenting content easy or demanding, simple or complex to interpret and comprehend?
AUGMENTED BACKGROUND (AB)	*Abstract – Realistic*	How much mental and cognitive activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.) in processing information from the augmented background? Was the task easy or demanding, simple or complex to interpret and comprehend?
INPUT MECHANISM (IM)	*Non-intuitive – Intuitive*	How much mental and cognitive activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.) in using the input mechanism? Is the input mechanism intuitive or non-intuitive to compensate for the extent of mismatch, or semantic differences, between real and virtual objects?
OUTPUT MECHANISM (OM)	*Non-immersive – Immersive*	How much mental and cognitive activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.) in using the output mechanism? Is output mechanism immersive or non-immersive to provide an appropriate level of presence in the mixed environment?

The weights and ratings may or may not covary. For example, it is possible for augmenting content to be the primary source of cognitive loading for a task, even though the magnitude of the cognitive load imposed by augmenting content might be low. Conversely, the input mechanism might be the primary source of cognitive load, and the cognitive load imposed by using the input mechanism might be rated as high for some versions of the task and low for others.

The first requirement in utilizing the MR CLX is for each rater to evaluate the weight (relative contribution) of each factor to the cognitive load of a specific task. These weights account for two potential sources of between rater variability: differences in cognitive load definition between raters within a task, and differences in the sources of cognitive load between tasks. The weights themselves provide diagnostic information about the nature of the cognitive load imposed by the task. There are six possible pair-wise comparisons of the four scales. Subjects attending the MR cognitive index measurement need to decide the member of each pair that contributed more to the cognitive load of the MR system. The number of times that each factor is selected is tallied. The tallies can possibly range from 0 (not relevant) to 3 (more important than any other factor). The second requirement is to obtain numerical ratings for each scale that reflect the magnitude of that factor in a given task. The scales are illustrated in Figure 9. Each scale is divided into 20 equal intervals anchored by bipolar descriptors (e.g. concrete/realistic). The comprehensive cognitive load score/index for each subject is computed by multiplying each rating with the weight given to that factor by that subject. The sum of the weighted ratings for each task is divided by 6 (the sum of the weights), resulting in a comprehensive cognitive load score. Such weighted ratings are then used as a dependent measure in whatever type of analyses the experimenter chooses.

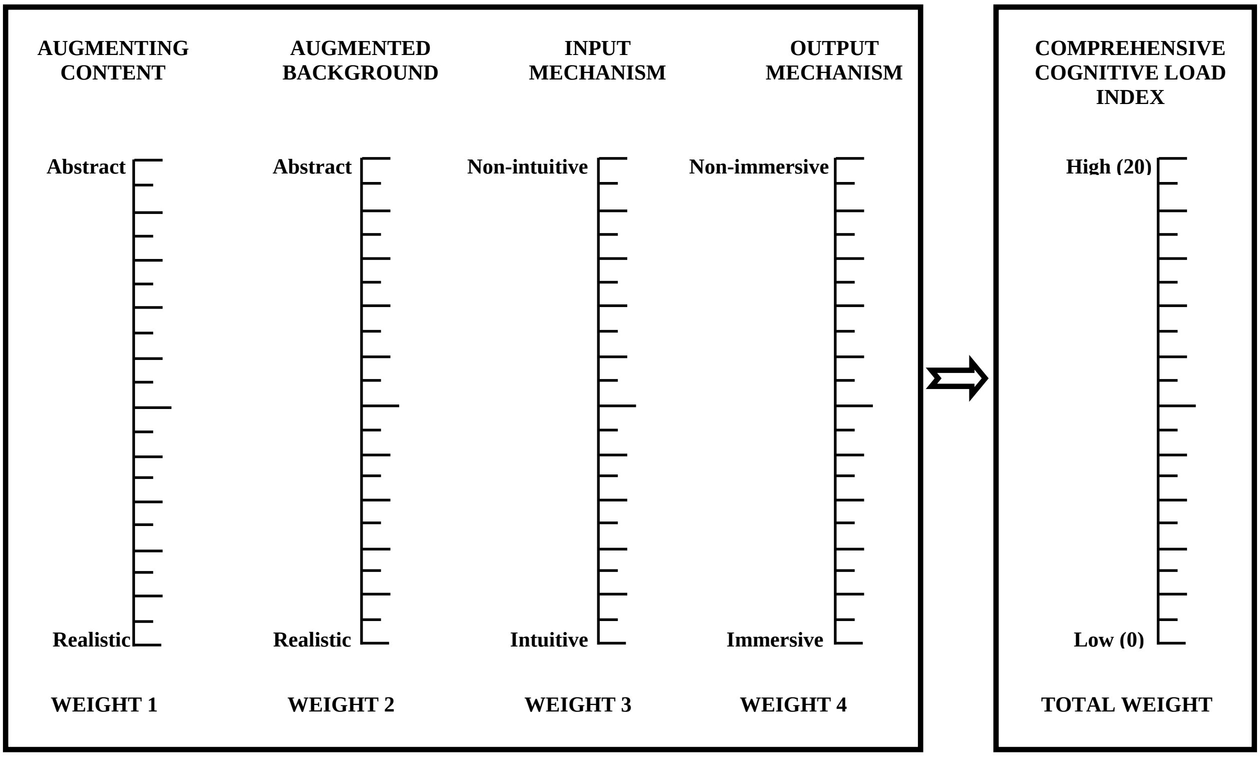

FIG. 9: Rating Scales Composition of Comprehensive Cognitive Load Index

Figure 10 depicts the composition of a comprehensive cognitive load score graphically. The bar graph on the left represents the four factor ratings. The width of the bars reflects the importance of each factor (its weight) and the height represents the magnitude of each factor (its rating) in a particular MR system for a particular task. The weighted cognitive load score (the bar on the right) represents the average area of the subscale bars.

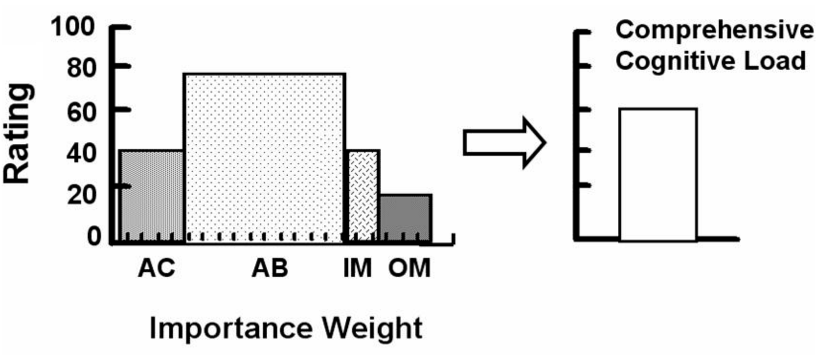

FIG. 10: Graphic example of the composition of a comprehensive cognitive load score

The MR CLX should be tested in a variety of experimental MR prototypes on different tasks. It is conceptualized that the derived comprehensive cognitive load scores should have substantially less between-rater variability than uni-dimensional workload ratings, and the sub-scales provide diagnostic information about the sources of cognitive load.

7. AR COGNITIVE CONTINUUM

A final note regarding location within Milgram's Reality-Virtuality (RV) continuum is worthwhile. Milgram acknowledges the use of only a reality/virtuality ratio as a criterion for classifying MR displays is insufficient for comparing alternative MR systems designs when systems involve varying combinations of real and virtual. The authors assert that typical industrial uses will likely not employ such fluidity due to the narrowly specified functions of AR systems. Therefore, a system might be defined as Augmented Reality based upon the fundamental intent of the system to support the user's effectiveness in a real world environment. Conversely, Augmented Virtuality would describe systems fundamentally designed to support work tasks that result in changes in the virtual world environment.

With AR thus distinguished from AV, an AR continuum (see Figure 11) can be conceived and related to the earlier discussion, for the purpose of identifying actual location along the continuum. This continuum can define the location of an AR system within the AR continuum from the perspective of overall cognitive load, based on the concept and methodology for measuring the MR comprehensive cognitive load index. The structure behind measurement on this continuum can identify the usability issues, which might be addressed, and suggestions can be made to improve specific AEC systems by making changes in technological components.

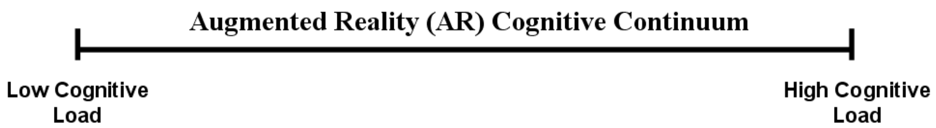

FIG. 11: Augmented Reality Comprehensive Cognitive Load Continuum

8. SUMMARY AND CONCLUSIONS

This paper has presented the formulation of a comprehensive multi-dimensional taxonomy for specifying Mixed Reality (MR) technology and characteristics. The discussion demonstrates the taxonomy as it may be applied to Augmented Reality but is also transferable to Augmented Virtuality. Characteristics regarding media representations, interaction methods, and tracking technology were progressively disclosed and some of them are presented as continua with implications for suitability and usability suggestions. Although there are rare evaluation works on the taxonomy in the world, the main conclusions and implications of the taxonomy for the design of future MR systems for the AEC sector are as follows:

- Two or more types of media representation presented together on a single display—a hybrid representation—may be very useful in some situations in AEC and especially when the MR system supports multiple functions.
- The richer fidelity of 3D sampled data makes it more realistic and less cognitively demanding (i.e., distracting) than 2D sampled data, an important consideration to maintain attention to quality or safe operations in construction field operations.
- As for the input mechanisms, in general, cognitive load (mental transformation requirement) can decrease by minimizing the mismatch between the control device and the AEC task at hand.
- The more intuitive input mechanisms involve more sophistication, which leads to greater computational requirements and thus implied cost to the current AEC computing infrastructure.
- Given a particular AR application, representative AEC tasks may implicitly suggest the mix of centricity, distortion and directness of view required. In turn, these requirements may lend credence to a particular display type.
- A proper perspective for AR may be that of not only feeling a connection to the displayed scene but also a connection to the virtual content that has been merged into the real scene of a specific AEC operation, i.e., a sense of sharing presence in the scene.

A concept and methodology for measuring an MR comprehensive cognitive load index was also developed and presented. Based on the MR comprehensive load index, the concept was formulated to define the location of an AR system within an AR continuum from the perspective of overall cognitive load, with the purpose of presenting an ordered quantitative ranking, according to which alternative AR technological components strategies can be compared.

So far, MR technologies and characteristics have been specified from the perspective of human-machine interface in this paper. In order to successfully design effective MR systems for AEC applications, designers and researchers also have to understand the user's goals and tasks, and the circumstances under which users must work to accomplish a goal. In light of the limited number of attempts of lab-based MR prototypes for AEC operations, theoretical research should be conducted to optimize the attempts to transfer MR technology into AEC industry utilization.

Future work will develop an AEC task taxonomy and task analysis method for understanding the nature of a given task from a user-centered perspective. Future work should also identify a set of influencing factors that should be analyzed and considered while selecting from among the MR technological component alternatives. After this step is achieved, a structured and iterative methodology developed for mapping appropriate Mixed Reality technology to a specific AEC task will be explored. To increase the likelihood of success in technology transfer, the methodology for developing user-centered, performance enhancing MR-based systems should be formulated, where AEC tasks should be generically analyzed and categorized according to common functional features, which could be mapped to a collection of suitable or required MR-related technology strategies. Also, a technology selection process should be identified to choose appropriate technology characteristics including information representations, interaction methods and, tracking technology for a specific task category. Such a thorough mapping methodology can be used to guide new MR-based system design as well as help evaluate existing systems.

9. ACKNOWLEDGMENT

The authors acknowledge support from National Science Foundation Grant No. CMS-0239091. Opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the National Science Foundation.

10. REFERENCES

Azuma, R. T., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B. (2001). “Recent Advances in Augmented Reality.” *IEEE Computer Graphics and Applications*, 21 (6), 34-47.

Azuma, R. T. and Bishop, G. (1994) “Improving Static and Dynamic Registration in an Optical See-Through HMD.” *Proc. Siggraph 95,* ACM Press, 197-204.

Baber, C., Haniff, D. and Woolley, S. (1999). “Contrasting Paradigms for the Development of Wearable Computing”, *IBM Systems Journal*, 38 (4), 551-565.

Boud, A. C., Haniff, D. J., Baber, C. and Steiner, S. J. (1999). “Virtual Reality and Augmented Reality as a Training Tool for Tasks”, *Information Visualization ’99*, 14-16 July, London, England, IEEE Computer Society, 32-36.

Caterpillar Distributed Virtual Reality System (1996) URL: http://viswiz.gmd.de/DVP/Public/deliv/deliv.443/DVR.html, accessed in December 2004.

Davies, C. (1996) “Cognition, Perception, and Experience in the Virtual Environment: Do You See What I See?” *Information from the SIGGRAPH '96 Panel Session*.

Dunston, P. S., Wang, X, Billinghurst, M., and Hampson, B. (2002). “Mixed reality benefits for design perception.” *Proceedings of the 19th International Symposium on Automation and Robotics in Construction (ISARC 2002)*, William Stone, editor, NIST Special Publication 989, 191-196.

Haniff, D. J., Baber, C and Edmondson, W. H. (2000). “Categorizing Augmented Reality Systems.” *Journal of 3D Images*, 14 (4), 105-110.

Hart. S. G., Battiste, V., and Lester, P. T. (1984). “Popcorn: A Supervisory Control Simulation for Workload and Performance Research. *In Twentieth Annual Conference on Manual Control*, Washington, D. C., NASA Conference Publication 2341, 431-454.

Iu, S. and Rogovin, K. W. (1996) “Registering Perspective Contours with 3-D Objects Without Correspondence Using Orthogonal Polynomials.” *Proceedings of VRAIS '96*, Santa Clara, 37-44.

Kutulakos, K. N. and Vallino, J. (1996) “Affine Object Representations for Calibration-Free Augmented Reality.” *Proceedings of VRAIS '96*, Santa Clara, 25-36.

Milgram, P. and H. Colquhoun (1999). A Taxonomy of Real and Virtual World Display Integration. Mixed Reality - Merging Real and Virtual Worlds. Y. O. a. H. Tamura, Ohmsha(Tokyo) & Springer Verlag (Berlin): 1-16.

Milgram, P. and F. Kishino (1994). A Taxonomy of Mixed Reality Visual Displays. IEICE Trans, on Information and Sytems (Special Issue on Networked Reality) E77-D (12): 1321-1329.

Milgram, P., H. Takemura, et al. (1994). Augmented Reality: A class of displays on the reality-virtuality continuum. SPIE: Telemanipulator and Telepresence Technologies, Boston, MA.

NASA Task Load Index (TLX) v. 1.0, Human Performance Research Group, NASA Ames Research Center, Moffett Field, California, URL: http://iac.dtic.mil/hsiac/docs/TLX-UserManual.pdf

Ong, S. K. and Nee, A. Y. C. (2004). *Virtual and Augmented Reality Applications in Manufacturing*, Springer Verlag, ISBN 1-85233-796-6.

State, A., Hirota, G., Chen, D. T., Garrett, B., and Livingston, M. (1996) “Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking.” *Proceedings of SIGGRAPH ’96*, New Orleans, 429-438.

Sweller, J., Van Merriënboer, J., & Paas, F. (1998). "Cognitive architecture and instructional design". Educational Psychology Review 10: 251–296.

Wang, X., Shin, D. H., Dunston, P. S. (2003). “Issues in Mixed Reality-Based Design and Collaboration Environments.” *CD Proceedings of ASCE 2003 Construction Research Congress*, March 19-21, Honolulu, Hawaii, 8 pages.

Wang, X. and Dunston, P. S. (2006). “Compatibility Issues in Augmented Reality Systems for AEC: An Experimental Prototype Study.” *Automation in Construction*, Elsevier, 15 (3), 314-326.