An Information-Theoretic Approach to the Cost-benefit Analysis of Visualization in Virtual Environments

Min Chen, Member, IEEE, Kelly Gaither, Member, IEEE, Nigel W. John, and Brian McCann, Member, IEEE

Abstract—Visualization and virtual environments (VEs) have been two interconnected parallel strands in visual computing for decades. From a broad perspective, the uses of visualization and virtual environments (VEs) have much in common. Both facilitate computer-supported activities involving primarily visual perception and human-computer interaction. Many systems that enable VE research and applications, such as the CAVE (Cave Automatic Virtual Environment) in the 1990s [31] and the RAVE (Reconfigurable Automatic Virtual Environment) in the 2000s [16], are also considered large visualization infrastructures. A variety of visualization applications, ranging from biomedical data visualization to text and document visualization, have been implemented to run in VEs.

1 INTRODUCTION

From a broad perspective, the uses of visualization and virtual environments (VEs) have much in common. Both facilitate computer-supported activities involving primarily visual perception and human-computer interaction. Many systems that enable VE research and applications, such as the CAVE (Cave Automatic Virtual Environment) in the 1990s [31] and the RAVE (Reconfigurable Automatic Virtual Environment) in the 2000s [16], are also considered large visualization infrastructures. A variety of visualization applications, ranging from biomedical data visualization to text and document visualization, have been implemented to run in VEs.

Despite the shared common ground, visualization publications rarely feature virtual reality or augmented reality capabilities, while research in VEs seldom addresses commonly understood challenges in information visualization, scientific visualization, or visual analytics. Concerns regarding the financial return on investment of historical VE hardware, recurring operation and maintenance, and to a lesser degree, software, have in many ways overshadowed the potential values that VEs may have as a viable discovery environment. Additionally, doubts about the value of conducting visual analytics and sense making in a VE have been a topic of consideration with mixed consensus. At first glance, the cost-benefit metric for visualization processes proposed by Chen and Golan [19] indicate that visualization in VEs may suffer from high cost and the lack of abstraction, but a cursory look at the history of VEs and the creativity in this space as a whole suggests there is more to understand.

In this paper, we investigate the cost-benefit of visualization in VEs from three perspectives: information theory, cognitive sciences, and practical applications. We use the term virtual environment (VE) as an encompassing term for immersive and semi-immersive virtual environments, mixed and augmented reality, visual as well as non-visual perception, and device-based as well as natural interaction. This investigation serves as a theoretical assessment about the usability of VEs in visualization as well as the applicability of Chen and Golan’s cost-benefit metric [19].

We frame our discourse based on immersion and presence, the most fundamental properties of VEs (Section 3). In the context of VEs, we first examine the three elementary quantities of the information-theoretic metric for cost-benefit analysis, namely alphabet compression, potential distortion, and cost (Section 4). We then support the theoretical propositions and analysis with theories and discoveries in the literature of cognitive science (Section 5). This is followed by practical evidence reported in the literature of visualization and VEs (Section 6). Our investigation leads to an analysis of the cost-benefit of performing four different levels of visualization tasks in VEs. This analysis enables us to consider the cost and benefit of immersion and presence at each level. It offers theory- and evidence-based explanations of the past implementations, while suggesting new opportunities and challenges.
Our contributions include (i) the application of theory-based cost-benefit analysis to an important but often overlooked area of visualization (Section 4), (ii) an effort of validating the theoretical analysis using evidence from cognitive science (Section 5) and practical phenomena (Section 6), and (iii) a demonstration that the theory can guide us to explore answers to practical questions (Section 7).

This paper is also supported by a number of appendices, including A: the mathematical definitions of several information-theoretic measures for the desirable self-containment, B: more detailed discourse on evidence from cognitive science, C: more detailed discourse on evidence from practical uses of visualization in VEs, D: a collection of answers to the 13 questions identified by the 2017 Workshop on Immersive Analytics, and E: more detailed discourse on the merits and demerits of performing different levels of visualization tasks in VEs, together with several predictions for long-term validation.

As the theoretical proposition in [19] can only be falsified by finding a counter example, we consider this broad application of the cost-benefit analysis to visualization in VEs as an important falsification exercise in theoretical research in visualization.

2 Related Work

In this paper, we use the term Virtual Environments (VEs) as an encompassing term for immersive and semi-immersive environments, large theater- or dome-based infrastructures, gigapixel displays, virtual reality systems, mixed reality systems, augmented reality systems, augmented virtuality systems, and web-based VEs. There are numerous VE systems and applications reported in the literature. Readers who are interested in exploring the broad spectrum of VEs may consult a number of books and literature surveys on the subject [94, 111] as well as in specific areas, including, but not limited to, presence [84, 95], haptics [25], augmented reality [4, 91], usability evaluation [13], medicine and healthcare [2, 26, 112], flight simulation [49, 79], education [9], sports [59], and cultural and natural heritage [1].

Milgram et al. outlined the Reality-Virtuality Continuum [60] that defines a continuous scale ranging between the completely virtual and the completely real. The area between these two extremes is referred to as mixed reality, which encompasses the technology of augmented reality where the virtual augments the real and the technology of augmented virtuality where the real augments the virtual. Schnabel et al. enriched this continuum by relating the connection between action and perception to the extent of interaction with real objects [83]. In this work, we will explore this continuum by examining the cost-benefit of virtuality and reality in visualization processes.

The theoretical research in the field of VEs has been largely focused on the concept of presence. Researchers have engaged in extensive discourse as to what constitutes the sense of presence and what may contribute to such a sense. Sheridan [88] and Heeter [38] were among the first to initiate this discourse. Sheridan [88] outlined three causes of presence: the extent of sensory information, the control of the relation between sensors and an environment, and the ability to modify a physical environment. Heeter [38] drew distinction between three types of presence, namely personal, social, and environmental presence. Schloerb [82] divides the notion into two categories, subjective and objective presence. Slater and Wilbur [96] related these two categories to two distinctive terms, “presence” and “immersion”. Lombard and Ditton [52] defined six aspects of presence: social richness, realism, transportation, immersion, social actor with medium, and medium as social actor. Slater et al. [95] further defined the dimensions of presence and immersion. Schuemie et al. gave a comprehensive review about this line of inquiry [84]. In this paper, we relate the concepts of presence and immersion to the abstract properties of alphabet compression, potential distortion, and cost in visualization processes [19]. We examine when and where presence and immersion may be beneficial to visualization users, and when and where they incur a noticeable amount of cost.

The theoretical research in the field of visualization has resulted in a large number of taxonomies (e.g., [11, 103]), many conceptual models (e.g., [28, 63, 109]), and a few theoretic frameworks (e.g., [21, 44, 116]). A more comprehensive list of references can be found at [22]. Recently Chen et al. [20] suggested that the theoretical foundation of visualization includes four major aspects, namely taxonomies and ontologies, principles and guidelines, conceptual models and theoretic frameworks, and quantitative laws and theoretic systems. This work falls into the category of conceptual models and theoretic frameworks. We aim to use information theory [87] to bring a substantial amount of activities in VEs into the information-theoretic framework for visualization [18]. Once visualization activities in VEs can be considered data intelligence processes, we can categorize these activities based on the four levels of visualization tasks [19], and apply the information-theoretic metric for cost-benefit analysis to these activities in an abstract and objective manner. This work also provides an opportunity to evaluate the theoretic findings in [19] to see if it can explain complex phenomena in visualization and VEs, if its analytical discourse can be supported by evidence in cognitive sciences and real-world applications, and if it can be used to suggest new guidelines, hypotheses, and predictions.

There has always been an interest in VEs in the field of visualization. For example, in 1995, Disz et al. [31] reported visualization experience in a CAVE, and Taylor et al. [100] presented promising results for immersive visualization for scientific applications. van Dam et al. and others reported some earlier VE-based visualization applications [36, 39, 98, 108]. In recent years, Ip et al. [40] reported the use of a large VE system for gigapixel analytics. Reda et al. [75] reported the use of CAVE2 for visualizing large, heterogeneous data. Bock et al. [12] showcased a dome-based VE infrastructure for Open Science events. Papadopoulos et al. [72] presented an immersive gigapixel display, and Papadopoulos and Kaufman [71] presented techniques that enable focus-and-context visualization using such a system. Müller et al. [64] presented an evaluation of biological data visualization using a large VE system. We hope that this work will stimulate new interests in delivering visualization solutions using VEs.

Some of the well-known pieces of wisdom, such as “maximizing data ink ratio” [107], “overview first, zoom, and details on demand” [89], and guidelines on 3D visualization of non-spatial data [32], have cast a negative shadow on VE applications. There have also been guidelines proposed for VEs. For example, Jerald proposed 13 design guidelines [42], including “focus on the user experience rather than the technology,” “design for visceral communication in order to induce presence and inspire awe in users,” “avoid the uncanny valley by not trying to make characters appear too close to the way real humans look,” and so on. Oculus, a VE technology supplier, offers an introduction to best practices [67], which include “safety first,” and “experiment, experiment, experiment.” All these guidelines seem to pose more questions than answers as to what makes some VE systems and applications work and what hinders others from success. While a variety of practical factors (e.g., the size of an avatar) have been, and should be, examined in probing these questions, it is also desirable to identify the most fundamental measures that can frame such inquiries. It is the current lack of theoretical abstraction that motivated this work.

3 Dimensions of Virtual Environments

Research in virtual environments (VEs) differs from that in computer graphics and visualization by placing a significant emphasis on the concepts of immersion and presence. While many have contributed to the formulation of these two concepts, we chose to adopt Slater et al.’s definitions [95] as a basis for our investigation. Here the term “dimension” is a common analogy referring to relatively independent aspects or attributes of VEs [95, 97].

Immersion is an attribute used to describe a technology. It characterizes the extent to which a VE is capable of delivering an inclusive, extensive, surrounding, and vivid illusion of reality to the senses of a human participant. There are six dimensions of immersion [95]:

- **inclusion** – the extent to which physical reality is shut out;
- **extension** – the range of sensory modalities accommodated;
- **surrounding** – the extent of visual coverage (e.g., panoramic, telescopic, microscopic, x-vision, etc.);
- **vividness** – the fidelity of the information conveyed (e.g., display resolution, color resolution, content richness, and variety of energy simulated within a particular modality);
In theory, the steps in Figure 2 can be infinitesimally small in time, but allow the probabilistic behaviors of a gas to be observed as a sequence of events. This sequence is a collective and accumulative attribute primarily about the human-centric processes, where the concept of presence is a sense of believing and social presence [90], e.g., participating in face-to-face interaction vs. remote communication; and a sense of co-presence [66], e.g., being together with other actors vs. unconnected individual actors.

In general, most technical advances in VEs have been driven by higher specifications of immersion and increased requirements for presence. Comparing a VE featuring more immersion or presence with a VE featuring less, the former generally delivers more data to a user through its movement skill of a participant, and other factors. This leads naturally to a question regarding the cost-benefit of different VEs.

**Preserve** is an attribute used to describe a human participant. It characterizes the state of consciousness, i.e., the psychological sense of being in a VE. In contrast to immersion, describing presence is often subjective and qualitative in nature, although some aspects may be measurable objectively and quantitatively. The state of consciousness can be described by, but not limited to, the following senses:

- a sense of believing [88], e.g., being at a place vs. viewing a set of images;
- a sense of naturalism [15], e.g., acting as if in the real-world vs. acting unnaturally;
- a sense of social presence [90], e.g., participating in face-to-face interaction vs. remote communication;
- a sense of co-presence [66], e.g., being together with other actors vs. unconnected individual actors.

We can finely divide the time steps, and the interaction among the three classes of processes can be defined as forward connections as shown in Figure 2. The connections 1, 4, and 7 indicate the state transitions within the same class of processes. A VE system that delivers output at time $t_{i+1}$ is expected to keep the state at time $t_i$. The position of a human participant at time $t_{i+1}$ is expected to be caused by a movement from the corresponding position at time $t_i$.

Fig. 2. A sequence of events in a VE can be considered as a series of processes flowing along a pathway in a complex space of all possible states of the entities involved. The main entities are the system of the VE and the human participant(s). In a mixed reality environment, parts of the reality are also involved as the third entity. While such processes result in changes at each stage, information is passed from the processes at stage $i$ to those at stage $i+1$ along paths marked by ①-⑨. The dotted lines indicate those paths typically available only in mixed reality environments.

4 Theoretic Propositions and Analysis

Processes and States. A sequence of interactive events in a VE can be considered as a processing flow as illustrated in Figure 2. In most VEs, there are two main types of processes: virtual environment (VE) processes and human processes. VE Processes include all machine-centric processes that enable the devices in a VE to change their states, e.g., generating new images, sounds, or force-feedback functions. Human Processes are human-centric and encompass any processes that enable the participants in a VE to change their states, e.g., attention, perception, interpretation, memory, emotion, speech, and body actions. In mixed reality environments, including augmented reality and augmented virtuality, a participant’s reality may also change. We refer to the causes of such changes as Real Environment Processes.

In theory, the steps in Figure 2 can be infinitesimally small in time, the resulting changes can be infinitesimally detailed, while the sequence can be innumerable long and the processes can be immeasurably complex. In practice, one can construct a coarse approximation of a processing flow for a specific set of tasks. We will adopt this approach when we examine practical VE systems.

**Presence** is another series of transformation of alphabets, i.e., from a data alphabet $V$ to an alphabet $Z$. This is not typical in a conventional data analysis and visualization workflow as observed in [19]. If the observation in [19] can be applied to visualize alphabets, the reduction of potential distortion is defined as the difference in terms of Shannon entropy between $V$ and $Z$. Meanwhile, the design emphasis on dispersion visualization alphabets, which may not happen in every session, can still be included as a letter in the VE alphabet. When it is transformed to a visualization workflow as observed in [19]. Potential Distortion happens when the VE alphabet size and complexity of alphabet increase more than an everyday visualization environment, what is the benefit that may be caused by the transformation. Instead of measuring the potential distortion, we consider only visualization applications and their function in virtual environments.

**Alphabets and Letters.** In abstract, let all possible states of VE processes (e.g., combinations of different computer-generated scenes, sounds, force-feedback, etc.) be the letters of a very large alphabet $V$, all possible states of human processes (e.g., combinations of the physical and cognitive states of all human participants in a VE) be the letters of a very large alphabet $H$, and all possible states of the reality observable to the VE system and the human participants in the VE be the letters of a very large alphabet $R$. Therefore, the change from one state to another is the same as the change from one letter to another.

Because the variables for these states remain more or less the same in a processing flow, we can maintain the same set of letters in each alphabet (i.e., $V$, $H$, or $R$) in the processing flow, but allow the probabilities of its letters to vary from one moment to another. For example, a participant may have a state of “fallen on the floor”. Although this state may not happen in every session, it can still be included as a letter in the alphabet $H$. Its probability varies depending on the task a participant is performing, the mobility skill of a participant, and other factors.

One observation that we can make is that the Shannon entropy

\[ H = -\sum_{i} p_i \log p_i \]

1 It is helpful to note that the abundance of these letters and the complexity of these alphabets should not be the reason to shelve a theoretical notion. In the history of thermodynamics, from which information theory is rooted, the kinetic theory, which models a gas based on the probabilistic behaviors of a huge number of particles, was difficult to appreciate before 1900s.
of each of the three alphabets in VEs does not have a general trend of reduction along the processing flow. Moreover, any increase of immersion and presence will most likely result in an increase of the size and complexity of alphabet $\cal V$, hence an increase of the Shannon entropy of $\cal V$. This is not typical in a conventional data analysis and visualization workflow as observed in [19].

However, when considering a visualization application in a VE, there is another series of transformation of alphabets, i.e., from a data alphabet at the beginning to a decision alphabet at the end. Here we refer to these alphabets, which are denoted as $Z_1$, $Z_2$, ..., $Z_n$, collectively as visualization alphabets. Unlike alphabet $\cal V$, these visualization alphabets may differ significantly in terms of data type or data resolution. Some of these alphabets, such as visualization images, will be a constituent part of the VE alphabet $\cal V$. But others, such as human perception about various data patterns, will be part of the human alphabet $\cal H$. In a mixed reality environment, some of these alphabets will be a constituent part of the real environment alphabet $\cal R$. It is not difficult to imagine that in some cases, the availability of the reality, i.e., letters in $\cal R$ are limited or too costly; hence one uses aspects of a VE alphabet $\cal V$ to simulate these letters in $\cal R$. In other cases, the desired immersion and presence cannot be achieved entirely using a VE alphabet $\cal V$ or it is too costly to achieve; hence one mixes some aspects of a reality alphabet $\cal R$ with those of $\cal V$. A fundamental question is: since VEs normally cost more than an everyday visualization environment, what is the benefit that would justify the extra cost?

### Cost-benefit Analysis

If the observation in [19] can be applied to VEs, the visualization alphabets, $Z_1$, $Z_2$, ..., $Z_n$, should also exhibit a general trend of Alphabet Compression, since the decision alphabet is usually much smaller than the original data alphabet in terms of Shannon entropy. Let $H(Z_i)$ be the Shannon entropy of alphabet $Z_i$. When it is transformed to $Z_{i+1}$ (e.g., from data to visual representation or from visualization image to perceived features), alphabet compression is defined as the difference in terms of Shannon entropy between the two alphabets, $H(Z_i) - H(Z_{i+1})$. On the other hand, the reduction of Shannon entropy must be balanced by the Potential Distortion that may be caused by the transformation. Instead of measuring the errors of $Z_{i+1}$ based on a third-party and likely-subjective metric, we can consider a reconstruction of $Z_i$ from $Z_{i+1}$. If a person has some knowledge about the data, the context, or the previous transformations, it is possible for the person to have a better reconstruction than one without such knowledge. In [19], this is presented as one of the main reasons that explain why visualization is useful. The potential distortion is measured by the Kullback-Leibore divergence $D_{KL}(Z_i || Z_{i+1})$, where $Z_i'$ is reconstructed from $Z_{i+1}$. Furthermore the transformation and reconstruction need to be balanced by the Cost involved, which may include the cost of computational and human resources, cognitive load, time required to perform the transformation and reconstruction, the adversary cost due to errors, and so on. Together the trade-off of these three measures are expressed in Eq. (1):

$$\text{Benefit} = \text{Alphabet Compression} - \text{Potential Distortion}$$

$$\text{Cost} = H(Z_i) - H(Z_{i+1}) - D_{KL}(Z_i || Z_{i+1})$$

(1)

The exact mathematical definitions of $H$ and $D_{KL}$ can be found in Appendix A. This metric suggests several principles in data intelligence. For example, Alphabet Compression has a positive impact as long as Potential Distortion or Cost is not increasing. Human knowledge can reduce the Potential Distortion and Cost in reconstructing data from visualization. The Cost reflects economic, cognitive, and other types of resources. Below we examine the dimensions of several typical VE systems, and relate these dimensions to the three abstract components in the metric (alphabet compression, potential distortion, and cost). Table 1 summarizes the above four types of VEs in terms of their design emphases (i.e., dimensions of immersion and presence), and the corresponding measures in the cost-benefit metric. We elaborate on each below, linking circled numbers in the table with those in the text.

#### 4.1 Theater-based Education Systems

Many visualization applications in VEs are designed for educational purposes, and they are a form of disseminative visualization [19]. They typically run in conjunction with a theater-based setup, which can accommodate tens to hundreds of participants. The large number of participants pose challenges in some dimensions of immersion and presence, such as extension, plot-interactivity, social presence, and co-presence as described in Section 3. Their design mostly focuses on the following dimensions:

- **inclusion**, e.g., using a very dark theater to block out reality;
- **surrounding**, e.g., using a large panoramic display featuring many more pixels than a typical commodity display screen;
- **vividness**, e.g., using high quality computer-generated imagery resulting from high resolution modelling and sophisticated rendering techniques such as global illumination; and
- **sense of believing**, e.g., seeing a black hole as a phenomenon as if it is observable to naked eyes.

Consider that the VE alphabet $\cal V$ includes primarily the data being visualized, visual imageries, commentary voice, and accompanying music. The initial visualization alphabet (i.e., the data alphabet) $Z_1$ is a subset of $\cal V$. Here we use the circled number to relate the statement to the mathematical description in Table 1. Hence, ideally, the mutual information $I(\cal V; Z_1)$ between the two alphabets should be maximized, and contain roughly the same amount of the entropy of $Z_1$. The additional visual and audio effects result in additional entropy $H(\cal V) - H(\cal V | Z_1)$. Meanwhile, the design emphasis on inclusion implies the minimization of the entropy of the reality $H(\cal R)$.
The final visualization alphabet (i.e., the decision alphabet) $Z_0$ is vaguely defined in such VEs. The participants are expected to absorb as much information as possible. This can be defined as the minimization of the potential distortion when a participant remembers the VE alphabet $V$ as a reconstructed alphabet $V[\mathcal{L}]$. The potential distortion is defined as the Kullback-Leibler divergence $D_{KL}(V||V)$. When the decision alphabet $Z_0$ is not precisely defined, we can also define the minimization of the potential distortion as the maximization of the mutual information between the final takeaway messages and the VE alphabet, $I(Z_0;V)$. If the decision alphabet $Z_0$ is relatively small and clearly defined, e.g., understanding the results of an election, the mutual information will be small. Thus it is necessary to increase alphabet compression $\mathcal{L}$. The advantage of maintaining a large and complex alphabet $V$ throughout a processing flow will disappear.

Interestingly, this type of VE purposely demands a huge amount of cognitive attention from the participants throughout a processing flow $\mathcal{L}$. This demand is facilitated by several immersion dimensions such as inclusion, surrounding, and vividness. The participants in the VE bear the responsibility to absorb as much information as possible, generally assumed to be an acceptable responsibility. Because such cognitive load is the cost borne by the participants, the provider of the disseminative visualization does not bear this cost.

However, as mentioned previously, there are large facility and operation costs paid by the VE provider. In some cases, such as the London Planetarium, the financial costs are partly or fully covered by the participants as an entrance fee. In other cases, governments and private sponsors are able to fund these types of educational activities as a good cause. The VEs that have an entry charge implicitly assume a significantly higher cost for providing the participants with a novel experience of accessing information. Meanwhile, for the information provider, the quality of immersion and presence is of utmost importance to eliminate potential distractions that would divert participants' cognitive load to other tasks.

4.2 Real-time Mixed Reality Systems

Many mixed reality systems are designed to support the needs for real-time visualization. For example, given an initial dataset (e.g., a computed tomography scan, or a planned route on a map), a mixed reality system may enable the user to visualize the data in conjunction with aspects of the reality (e.g., a patient, or a landscape in the real world). The visualization tasks are usually reasonably well-defined, e.g., verifying the position of an anatomical feature shown in the visualization against the actual geometry of the patient, or determining the geographical features in the landscape that correspond to the planned route. These are typical observational visualization tasks. Performing such a task falls neatly into the visualization workflows discussed in [19]. The initial dataset is a letter in the data alphabet (i.e., $Z_1$), and the visualization tasks are represented by the decision alphabet (i.e., $Z_0$). Most use cases of mixed reality systems have a clearly defined tasks, hence $Z_{eq}$ is expected to have a much smaller entropy. The alphabet compression from $Z_1$ towards $Z_0$ is thus critical $\mathcal{L}$.

In a perfectly idealized situation, one might wish to have the relevant aspects of the reality (e.g., the patient or the landscape) captured as a high-resolution 3D model by a computer system, and the captured reality could then be visualized using high-fidelity rendering in conjunction with the dataset. In other words, the VE alphabet $V$ will include aspects of the reality as well as the data. However, the current technological limitation gives rise to many problems. For example, the relevant aspects of the reality might change dynamically, and any captured 3D model would become unsynchronized with the reality almost immediately after its capture. The computational costs for processing a high-resolution 3D model and rendering high-fidelity visualization could be incompatible with the real-time task requirement and the operational environment. A low-resolution model or low-fidelity visualization of the model would incur more cognitive load of the user in relating the visualization to the reality.

A mixed reality system addresses the aforementioned technological limitation by introducing the reality as part of the visualization solution. We use $V \otimes R$ to denote the combined alphabet for the mixed reality. It allows the visualization to focus on depicting the data rather than the reality, i.e., the mutual information between $V$ and $Z_0$ should be maximized $\mathcal{L}$, while the visualization process at each stage should minimize the potential distortion due to the mixed reality setting $\mathcal{L}$.

In terms of immersion and presence, the real environment alphabet $R$ delivers a substantial amount of the requirements for extension, surrounding, vividness, plot-interactivity, and the senses of believing, naturalism, social presence, and co-presence. Hence, it is usually the more the better $\mathcal{L}$. The main technical challenge is with aspects of match between the true reality and the perceived information through viewing the integrated visual representation of data $V$ (assuming $Z_1 \subseteq V$) and $R$. The cognitive load for match can be rather high $\mathcal{L}$.

In terms of cost-benefit analysis, the alphabet compression from $V \otimes R$ to the decision alphabet $Z_0$ is expected to be very high. The potential distortion depends on the immersion attribute match, which can be influenced by many factors, such as the capability of the mixed reality system and the user’s experience in registering $V$ against $R$. In comparison with the idealized VE system that captures all required aspects of the reality $R$, the mixed reality approach is likely to be more economic in the short to medium term. The potential distortion due to deficiencies in achieving adequate match can be alleviated by having more information in the VE alphabet $V$ about the reality. The more overlapping between the virtuality $V$ and the reality $R$, the more mutual information $I(V;R)$, and the less potential distortion. Hence, if the technologies for capturing some aspects of reality are becoming more usable and less costly, it is possible to increase the amount of data for representing the reality, such as a 3D model of a patient captured using camera or a 3D landscape captured using drones. If such 3D models were captured prior to the real-time visualization, they could potentially be used to enhance the user’s ability to match the virtuality with the reality, while reducing the cognitive load in registering $V$ against $R$.

4.3 “Big Data” Visualization Systems

Many large VE infrastructures are equipped with gigapixel displays. Typically, they have been used for visualizing some very large datasets [75], such as gigapixel images [40], and large biomolecular models for simulating the dynamic behaviors of millions of atoms [64]. The visualization tasks involved usually fall into any of the four levels of visualization [19]. For example, creating an archeological exhibition [8] is a disseminative visualization task. Interactive exploring multi-gigapixel images for object identification [40] is an observational visualization task. Identifying patterns in social networks is an analytical visualization task. Validating or debugging a large biomolecular model is a model-developmental visualization task.

Because the applications in question often do not demand presence dimensions, such as the sense of believing, naturalism, or social presence, these VEs usually place less emphasis on some immersion dimensions such as inclusion, extension, and match. They are sometimes referred to as semi-immersive environments. The requirements for displaying “big data” naturally leads to an emphasis on surrounding and vividness. In most cases, the users are given a substantial amount of control, hence a high-level of plot-interactivity. In many applications, the VEs allow multiple users to perform their visualization tasks collaboratively, though the sense of co-presence usually arrives naturally through the reality rather than through the display media and interaction devices of a VE.

The relative merits and demerits of using a gigapixel display in comparison with using one or a few conventional desktop displays (referred to as megapixel displays) are always a concern in the minds of many technology providers and users. We can consider the potential merits and demerits using the information-theoretic metric for cost-benefit analysis.

For observational, analytical, and model-developmental visualization tasks, the VE alphabet $V$ usually focuses on the data alphabet $Z_1$. The visualization tasks are expected to be the same for a gigapixel display and a few megapixel displays. The decision alphabet $Z_0$ is thus the same. Both types of displays are expected to deliver the same amount of alphabet compression. Hence we reuse the premises in [19]. The visualization process should ideally have high rate alphabet com-
Fig. 3. If a VE-based training system were considered as a workflow for observational visualization (above), there would be a contradiction with the cost-benefit metric [19] for optimizing visualization processes. However, it is more appropriate to consider a VE-based training system as a workflow for model-development visualization (below). Because we have limited knowledge about the structure of the model, the variables that affect the model, and the evolution of the model, the VE tends to maximize the amount of reality that can be simulated.

The primary reason for using VE-based training is the lack of access to the required reality \( \mathbb{R} \). For example, it would be inappropriate to train certain medical procedures on real patients; it would be foolish to set fires on many arbitrary buildings in order to train firefighters; and it would be costly to create many different scenarios in sports using real players. Hence, one creates a VE alphabet \( \mathbb{V} \) to approximate \( \mathbb{R} \). The desired variations in \( \mathbb{R} \) are stimulated by different letters in \( \mathbb{V} \).

Let us focus on visual stimuli, and consider the VE-alphabet \( \mathbb{V} \) as the data alphabet \( \mathbb{Z}_1 \) and all the possible actions in response to the visual stimuli as a decision alphabet \( \mathbb{Z}_n \). Here we use \( \mathbb{Z} \) to indicate that this is a very rough approximation, and the actual \( \mathbb{Z} \) is more complex as discussed below. There is a trend in alphabet compression from \( \mathbb{Z}_1 \) to \( \mathbb{Z}_n \), and in many training applications, the transformations may take a split second.

This seems to suggest an inconsistency with the theory proposed in [19]. For a typical visualization application, any embellishment of \( \mathbb{Z}_1 \) would incur more cost for additional processing. In other words, as shown in Figure 3, users should be able to react more quickly or even more accurately if \( \mathbb{Z}_1 \) is pre-processed in the direction towards \( \mathbb{Z}_n \). However, the practical experience suggests that the demand is to embellish \( \mathbb{Z}_1 \) with more realism in many dimensions of immersion and presence.

The reason for this apparent contradiction is that VE-based training is a form of model-developmental visualization and the actual alphabets \( \mathbb{Z}_1, \mathbb{Z}_2, \ldots, \mathbb{Z}_n \) along a visualization process have to include the model being developed. When one is developing a machine-centric model, such as a decision tree in [99], the initial alphabet \( \mathbb{Z}_1 \) encompasses all possible variations of the decision tree model in the context (\( \mathbb{M} \)), all possible variations of the inputs to the model (\( \mathbb{I} \)), and all possible variations of the outputs to the model (\( \mathbb{O} \)). At the beginning of the workflow, we do not know how these three components are related to each other. So \( \mathbb{Z}_1 \) has the highest level of uncertainty as it encompasses all combinations of three types of variations \( \mathbb{Z}_n = \mathbb{M} \times \mathbb{I} \times \mathbb{O} \). Through a visualization-assisted learning workflow, we gradually narrow down a specific model and establish the functional relationships among the three variations, which is represented by \( \mathbb{Z}_n \) at the end of the process. Typically \( \mathbb{Z}_n \) encompasses only one model, or a few models. For simplicity, let us make \( m_{\text{best}} \in \mathbb{M} \) as the chosen optimal model. It can be written as \( \mathbb{Z}_n = \{ m_{\text{best}} \in \mathbb{M}, i \in \mathbb{I}, o \in \mathbb{O} \} \). Since the number of letters in \( \mathbb{Z}_1 \) is at the scale of \( |\mathbb{M}| \times |\mathbb{I}| \times |\mathbb{O}| \), and that of \( \mathbb{Z}_n \) is at the scale of \( |\mathbb{M}| \), \( \mathbb{Z}_n \) has much lower entropy than \( \mathbb{Z}_1 \). By juxtaposing all possible models at the beginning of the training with a data space and the model(s) converged at the end of the training with a decision space, the trend of alphabet compression is consistent with the theory proposed in [19].

The mathematical formulation of a VE-based training system is fundamentally the same as that of the aforementioned visualization-assisted learning workflow. The main difference is that we cannot currently directly visualize a human-centric model (e.g., the brain function for controlling a type of motion), but we can normally do so for a machine-centric centric model (e.g., a decision tree). Because in a real-world environment, a human-centric model \( m_{\text{best}} \in \mathbb{M} \) may require more complex and detailed inputs (e.g., type of building, location of the fires, etc.) than some abstract information (e.g., 30% of a building is on fire), making \( \mathbb{Z}_1 = \mathbb{V} \) closer to the reality \( \mathbb{R} \) partly reflects our attempt to gain the knowledge as to the inputs that the model may depend on. Without the definite knowledge about these inputs, the best one can do is to provide as much immersion and presence as possible [18] within the constraints of financial cost [18]. Normally we would like users of a VR-based training system to have the most faithful perception of the visual stimuli [19]. In fact, if possible, we would like to use the visualization to stimulate the perception of the reality being simulated, which challenges the visualization techniques further [19]. Meanwhile, it is likely that the more immersion and presence, the less cognitive load [18] and the better training outcome.

**Summary of Theoretical Findings.** The merits and demerits of performing visualization tasks in VEs may have some correlation with
the levels of visualization tasks, which correspond to the complexity of
the search space concerned [19]. In particular,

- The increase of presence leads to the increase of attention and
in some cases enjoyment, which is desirable to the presenter in
Disseminative Visualization (Level 1). However, failing to meet
such cost of attention often leads to inattentional blindness in
visualization.
- The increase of presence leads to the reduction of potential distor-
tion by making use of humans’ memory and a priori knowledge, which is desirable in some Observational Visualization (Level 2) where perceived information must be associated with reality efficiently and effectively.
- The increase of presence leads to the increase of alphabet compres-
sion and the reduction of potential distortion and learning cost in some Model-Developmental Visualization (Level 4) where human participants’ behavioral models can be studied, and hu-
mans’ learning capabilities can be utilized.
- The increase of presence usually leads to the decrease of alphabet compression and increase of cost, which is often undesirable in Observational and Analytical Visualization (Level 2 and Level 3), especially when non-intuitive mapping (not easy to learn and remember) from data to virtual objects is deployed.
- Analytical visualization tasks and (algorithmic) model-
developmental tasks typically present a large and complex search
space for the target patterns or optimized solutions. The increase
of immersion and presence has potential to provide a means to
explore a large and complex search space. We also touch
briefly on an open question: How can we introduce intuitive and
effective presence to support humans’ intelligence in discovering
target patterns or optimized solutions?

5 Evidence from Cognitive Science

In this section, we draw evidence from cognitive science to support
the theoretical discussions in Section 4. In particular, we examine
the aspects of attention, visual search, working memory, and motor coordination. The most relevant findings in cognitive science are
detailed in Appendix B.

Attention. The evidence in cognitive science shows that attention or
selective attention is essential for humans to make efficient and effective
use of the limited cognitive resources available to each individual
[3]. The fine coordination of eye, hand and body movements provide objective details about the organization of attention, working memory and sensorimotor control [37].

For a large display in a VE, participants have to adjust their gaze as well as move their heads. When participants are at a relatively
closer proximity to the display, walking around also becomes neces-
sary. These additional movements also incur additional requirements for
information retention. Hence, there is a high cognitive load for
maintaining a certain level of awareness across the external information available. For disseminative visualization, a VE system attracts and
attracts demands more attention from participants, and can potentially facilitate
the delivery of more information for educational purposes 1 3 4. For observational and analytical visualization, on the other hand, such
a demand has to be carefully managed 5 6. The more cognitive re-
sources are devoted to the attention for retrieving external
information, the less cognitive resources are available for the attention to internal
events (e.g., analytical reasoning and decision making) 7.

Visual Search and Working Memory. Humans are efficient visual
searchers. Cognitive studies have confirmed humans’ ability to under-
stand a visual scene at a glance [68]. Retention, on the other hand, is
not our strength. Humans’ short-term (verbal) memory is famously
limited to around seven items [61].

Most visualization techniques provide an effective means for exter-
nal memorization, and utilize our ability in visual search to compensate
for limited working memory resources. In “big data” visualization
applications, a high-resolution display can provide more display band-
width for external memorization and enable visual search tasks with
less interactions than a low-resolution display. On the other hand,
any humans’ soft knowledge about the “big data”, including the previ-
ous visualization experience of the data, is retained through long-term
memory, which does not have the same limitation as working memory.
When such knowledge is utilized for visual search, selective atten-
tion becomes more effective 8. If the high resolution of a display is
achieved by a very large display surface, the demand for more cognitive
load related to attention may undermine the benefit of visual search
with less interaction 9.

In real-time mixed reality applications, the challenge of the match
dimension is often related to visual search and memorization. The inte-
grated presentation of two types of visual stimuli (i.e., virtual and real
objects) is not what one encounters in everyday life. Hence, this
unfa-
miliarity may reduce humans’ aforementioned visual search capability 10. There can be mismatch between the integrated visualization
and the user’s mental models gained from real-life experience. Any
mismatch between the two types of visual stimuli (e.g., due to poor
registration) can create further difficulties. Hence, the solutions to
these issues include (i) an improvement of the match between the two
stipulations in order to reduce the user’s cognitive load for “men-
tal registration” during visual search, and (ii) introducing training in
order to improve the relevant mental models of the user retained in the
long-term memory 11.

Motor Coordination. One lesson from the past 50 years or so of
literature is that moving our bodies is one of the most demanding tasks
we perform as humans. The number of variables in human movement
control is estimated to be about 2^600, with considerably simplified assumptions about motor activations [115]. The evidence in cognitive science confirms that the “models” of
humans’ motor coordination are highly complex. In order for users
to develop the “lost” motor coordination skills (e.g., due to medical
conditions) or some “new” skills (e.g., to perform tasks beyond one’s
natural ability), there is a need for model-developmental visualization.
The use of VEs with a high level of immersion and presence provides
more stimulus information to a variety of the variables of a model under
training. This also provides opportunities for researchers to develop the
understanding of such a model and its main variables.

6 Evidence from Practical Applications

Visualization has been a ubiquitous tool for supporting scientific and
scholarly activities in almost all disciplines. Many visualization appli-
cations have been developed to run in VEs. These include applications
in education and e-learning (e.g., [9]), design and testing (e.g., [69]),
sports training (e.g., [59]), volume visualization (e.g., [45, 46]),
information visualization (e.g., [64, 75]), medicine and healthcare (e.g.,
[2, 26, 112]), environmental planning (e.g., [71, 72]), information dis-
semination and public engagement (e.g., [12]), and culture and heritage
(e.g., [1]). In this section, we examine three visualization applications
in VEs, and discuss the cost-benefit of such applications based on
the evidence reported in the literature. More detailed description and
analysis of these case studies can be found in Appendix C.

Data Visualization on Large Displays. Many empirical studies
were carried out to evaluate the utility of large displays for visualization
[7, 14, 41, 51, 64, 76, 80, 118], resulting in mixed conclusions about the
relative merits of such VE systems. Moorland [62] summarized a set
of challenges in delivering effective visualization on large displays.
Müller et al. [64] reported an empirical study on using large high-
resolution displays for comparative visualization. It is an unbiased
piece of investigation into the effectiveness of using large displays (or
everywhere). They compared a large display (6m × 2.2m, 10,800 ×
4,096 pixels) with a 24-inch desktop monitor. They examined visual-
ization tasks for judging the geometric differences among 40 biological
structures. The results of the study showed that accuracy and response
times did not differ significantly between different devices. Participants
did not have clear preference towards the large VE display or the des-
top monitor. In such a case, the desktop monitor was seen as a more
 economical choice.
From the perspective of information-theoretic cost-benefit analysis, we can observe that the visualization task was to examine the relationship amongst 40 data objects, and is at the level of analytical visualization. Because the total number of possible relationships is relative low (780), the task was carried out with brute-force observation, in other words, more similar to typical observational visualization. The task has a well-defined decision alphabet, and hence the alphabet compression is substantial. The dependent variables (e.g., accuracy and response time) of the study relate directly to the potential distortion $\mathcal{D}$ and cognitive cost $\mathcal{C}$ in the cost-benefit metric. From the perspective of cognitive science, the visualization task is a relatively complex visual search task, and demands working memory retain some interim comparative judgements. Hence any additional head and body movement may incur more cognitive load. In their results, there is a small trend of high response time for the large display, which might indicate such extra load. Meanwhile, the benefit of the large resolution display is unclear as participants viewed two types of displays at different distances. The requirement for display resolution is also complex for geometrical comparison, as the judgement is likely made at multiple levels of overview and details.

The study indicates that achieving sufficient cost-benefit of using VEIs in observational and analytical visualization for “big data” is not trivial. Nevertheless, once we understood the three abstract measures of alphabet compression, potential distortion, and cost, we can explore this avenue further by considering visualization tasks that may demand more alphabet compression. For example, consider the cost-benefit ratio in the 40-structure study by Muller et al. [64] as the benchmark, would examining relationships among 400 or 4,000 structures change the benchmark ratio? Could the cost be reduced if some analytics algorithms were used to prioritize the comparative activities?

**Surgical Training**

Domain experts in medicine are early adopters of VEIs, particularly in the context of training surgical procedures. Traditionally surgical training is an apprenticeship model whereby trainees observe the procedure being performed, before attempting it for themselves (under guidance) on real patients. However, this apprenticeship model is being challenged because of the quality and safety standards in surgical training, reduction in training hours, and constant technological advances. As a result, pressure on training outside the operating room has significantly increased. A variety of training aids are available, such as mannequins, but are often unrealistic compared with the real patient. VE-based training has been widely accepted as a complementary training methodology for well over two decades (e.g., [48,50,86,119]). Typically a VE helps to develop hand eye coordination and other psychomotor skills, while catering for different patient types and enabling the exploration of what-if scenarios when something goes wrong.

The application of surgical training is a form of model-development visualization. It places a particular emphasis on **vividness** and the sense of believing that the virtual patient is real. The VE alphabet $\mathcal{V}$ encodes the variations of the rendered virtual scenes (e.g., endoscopic views), animation of the virtual patient (e.g., from respiration), and haptic effects calculated on the virtual surgical tools. The human alphabet $\mathcal{H}$ encodes the variations such as the visual attention of the surgeon, the sensation felt on the surgeon’s hands, and the decision on how to proceed from the current state. The real environment alphabet $\mathcal{R}$ encodes the variations such as the parameter settings on the input interface and the state of the haptic actuator. The mental models to be trained are for not only the eye-hand coordination but also the decision mechanism in response to different scenarios. The cost-benefit of using such VEIs has already been confirmed by many practitioners.

Minimally invasive surgical (MIS) procedures currently provide the most opportunities for surgical training using VEIs (see Appendix C for details). As MIS can also be deployed in conjunction with real-time mixed reality systems, the visualization tasks involved also fall into the level of observational visualization, as the surgeon needs to observe a variety of data from both the virtual and real environments frequently and at a quick glance, and to make rapid decisions. It is a research ambition to evolve such systems further to surgical guidance systems to be deployed in real operation rooms. In other words, there are continuing research efforts to increase the space of the real environment alphabet $\mathcal{R}$. The visualization tasks performed in such surgical guidance systems will be mission-critical, and the necessity for achieving high rate alphabet compression (i.e., from data to decision) $\mathcal{D}$ with minimal potential distortion will be paramount.

After surveying a large collection of reports on developing AR and VR applications in radiotherapy, Cosentino et al. identified that the accuracy of the registration and cost of hardware are the two important factors affecting the deployment of such VE systems in practice [26]. The accuracy of registration is a form of match, directly corresponding the potential distortion in reconstructing the reality $\mathcal{R}$ and the data to be visualized $\mathcal{Z}$, and indirectly related to the cost (e.g., cognitive load in matching and the damaging consequences due to errors). Cosentino et al. pointed out that between 2002 and 2013 there was a reduction of the registration error from 0.08mm to 0.02mm. They observed that the majority of the reviewed studies used costly hardware not widely available commercially, but widely-available commodity devices (e.g., Wii remote, iPad, iPhone, iPod Touch) started to appear in recent studies. They also noticed that there was little mention of the problems of user discomfort, requirements of special training, or equipment cost in these recent studies based on commodity devices. However, the computation power available on such commodity devices is still not quite adequate for real-time registration. In other words, there is a trade-off between the potential distortion and the cost, which has shaped the current state of deployment focusing on teaching and training applications.

**Sports Training**

Sporting activities can lend themselves very well to being replicated within a VE. In the context of visualization, domain experts in sports are interested in using VEIs to provide alternative ways of training a skill, and analysing performance. Miles et al. [57] provide a comprehensive review of the use of VEIs for training in ball sports. They identified several key research challenges, including: what technologies achieve the best results; should stereoscopy be used and is a high fidelity VE always better; what types of skills appear to be best suited to training in VEIs; and do sports skills reliably transfer from VE training conditions to real-world scenarios?

Many challenges highlighted in [57] relate to different dimensions of immersion and presence. For example, the necessity of “closer approximation of the target skill and the environmental conditions of the target context” reflects the need to simulate as much reality as possible. From the perspective of cognitive science, such requirements reflect the complexity of the human model for motor coordination. The emphasis on “specific motor control skills” (e.g., ball passing in rugby [58]) enables the reduction of the complexity of the variable space through domain experts’ understanding of what may affect such skills. In other words, this facilitates the reduction of the complexity of the VE alphabet, and thereby the reduction of the cost of using such visualization in a VE $\mathcal{R}$. In addition, the discussions in [57] on the relative merits of stereoscopic displays and the necessity of high fidelity imagery also reflect the need to understand variable space of individual models under training. While stereoscopic displays introduce depth perception as a variable in the training of a model $\mathcal{R}$, it may also introduce new variables (e.g., fatigue and discomfort, view distortion) that are undesirable to be part of the model $\mathcal{R}$. During a training session, a player processes visual stimuli at a very high speed, achieving extremely high rate of alphabet compression $\mathcal{D}$. Hence the challenge about image fidelity is about how much compression is done by the computer (in the case of low fidelity) and how much is done by humans (in the case of high fidelity).

Miles et al. [58] reported a VE-system for training ball passing skills in rugby as shown in Fig. 1(d). The system simulates a number of variables, such as the flight trajectory of the virtual ball, and wind direction and strength. They noted that the use of stereoscopy made no significant difference to the accuracy of depth perception in this simulation. This is a typical visualization task in model development. Similar to visualization-assisted machine learning [99], it is necessary to monitor the variable space of a model, and to relate the performance of the model with various initial conditions. For VE-based training, the visualization capability is readily available on site. It is highly desirable to utilize such capability for supporting the model development.
7 Answering Practical Questions

So far we have shown that the cost-benefit analysis based on information theory can explain why different types of VEs have different impacts on each of the levels of visualization tasks, and such explanations can be supported by evidence from cognitive science and practical applications. If the above theoretical discourse is correct, we should also expect the cost-benefit analysis can be applied to practical problems that have not yet been solved. While there will be a journey from any theory to a corresponding practical solution, the theory should at least offer an effective pathway to a solution.

As part of IEEE VIS 2017, the attendees of the Workshop on Immersive Analytics: Exploring Future Interaction and Visualization Technologies for Data Analytics (http://immersivanalytics.net), posed a number of questions for discussions during the Workshop. Since the discussions on many questions were largely from a practical perspective and often inconclusive, they offer an opportunity to test the usefulness of the cost-benefit analysis based on information theory. There are a total of 16 questions. As detailed in Appendix D, we have attempted the answers to 13 of these questions. Here we use our answers to Q11 as an example to demonstrate that the cost-benefit analysis can offer an effective pathway to help advance the discourse.

Q11. Do we really need 3D visualization for 3D data?

We assume that the term “3D visualization” implies the use of a 3D volumetric display or a 2D stereo display. This question is indeed at the heart of the cost-benefit analysis. Let us compare the process for generating a visualization alphabet on a 3D visualization environment with the process involving a plain 2D environment. For the same 3D data alphabet, the former is likely to result in less Alphabet compression, less Potential Distortion, less cognitive Cost, but more economic Cost. The Potential Distortion and cognitive Cost in the reverse mapping from the visualization alphabet to the data alphabet depends partly on the viewer’s knowledge about the data being visualized. If a viewer is familiar with the variations in the data alphabet, such as different chairs, the Potential Distortion and cognitive Cost can be very similar between the two types of visualization environments. Hence, the higher Alphabet Compression and lower economic Cost in the plain 2D environment can bring more cost-benefit. On the other hand, if the variations in the data alphabet are unfamiliar to the viewer, such as the swimming shapes of a large school of fish, the plain 2D environment will likely result in more Potential Distortion and cognitive Cost. Here we use the term “alphabet” throughout the discussion to emphasize that we are not considering only a single dataset rather all possible datasets that a viewer can encounter in a particular context.

Hence, the question does not have a yes or no answer, but an optimization solution based on the cost-benefit metric. In addition, we also need to look forward to the decision alphabet following the visualization process. Some types of potential distortion (e.g., the shape of individual fish) may have less impact on the decision about the collective shape of schooling fish. In such a scenario, one may ask if using a gigapixel display would bring much more benefit than an original desktop display. Similarly, one can also apply the analysis to compare 3D geometric models displayed as outlines, wireframe, shaded, and photorealistic objects using a 2D display.

Recently, a web-based forum, VxsGuides (visguides.dbvis.de), was established specifically for discussing guidelines in visualization [30]. Before long, several VE-related discussion threads emerged, including “(Don’t / Do) replicate the real world in VR?” “What are the main disadvantages of 3D visualizations in general?” “Visual Variables for visualizations in VR;” and “Facilitate depth perception for 3D visualizations.” Although the cost-benefit analysis was only mentioned briefly in one of the threads partly because of the conversational nature of the forum, it can be used to frame the scientific questions if the discourse inspires further research effort.

For example, when one considers a visual channel (or visual variable) in visualization, it is not difficult to examine two basic processes, i.e., $P_1$: from a data variable to a visual channel, and $P_2$: from visual channel to perceived data variable. Information-theoretically, these are three alphabets and two transformations among them. Both transforma-

8 Conclusions

In this paper, we have applied information theory in general, and the recently proposed cost-benefit model [19] in particular, to an array of visualization tasks in VEs. The cost-benefit analysis allows us to examine different aspects of VEs and visualization in abstraction, and to make generalized observations. The evidence from cognitive science supports our analysis of various cognitive costs in VEs, and the evidence from practical applications substantiates the benefits of using VEs for visualization in conditions suggested by the theoretical analysis. We believe that this theoretical study has resulted in several contributions. It provides a theory-based approach to analyzing the cost-benefit of visualization in VEs, and offers a set of findings as summarized in Section 1. It extends the original definition of four levels of visualization, and provides further evidence to validate the cost-benefit metric through its application to a large research area intersecting visualization and VEs. The mutual corroboration between the theoretical discourse and practical observation is encouraging. We hope that many researchers, including ourselves, will explore various challenges presented in Appendix E, while seizing the opportunity of continuing reduction of the cost of some VE devices.

References

[1] A. C. Addison. Emerging trends in virtual heritage. IEEE MultiMedia, 7(2), 2000.
[2] A. Alaraj, M. G. Lemole, J. H. Finkle, R. Yudowsky, A. Wallace, C. Luciano, P. P. Banerjee, S. H. Rizzi, and F. T. Charbel. Virtual reality training in neurosurgery: Review of current status and future applications. Surgical Neurology International, 2(52), 2011.
[3] J. R. Anderson. Cognitive Psychology and its Implications. Worth Publishers, 6th edition, 2004.
[4] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. IEEE Computer Graphics & Applications, 21(6):34–47, 2001.
[5] A. Baddeley. Working memory and conscious awareness. Theories of Memory, pages 11–20, 1992.
[6] D. Baddeley and G. Hitch. Working memory. Psychology of Learning and Motivation, 8:87–99, 1974.
[7] R. Ball, C. North, and D. Bowman. Move to improve: Promoting physical navigation to increase user performance with large displays. In Proc. ACM CHI, pages 191–200, 2007.
[8] J. A. Barceló, M. Forte, and D. H. Sanders, editors. Virtual Reality in Archaeology, volume 843 of British Archaeological Reports. Archaeopress, 2000.
[9] L. Bell, P. Franks, and R. B. Trueman, editors. Teaching and Learning in Virtual Environments: Archives, Museums, and Libraries. Libraries Unlimited, 2016.
reprocessing tool for quantitative data analysis in a virtual environment. In Proc. ACM VRST, pages 212–215, 2006.

[37] M. Hayhoe and D. Ballard. Eye movements in natural behavior. Trends in Cognitive Sciences, 9(4):188–194, 2005.

[38] C. Heeter. Being there: The subjective experience of presence. Presence, 1(2):227–219, 1992.

[39] B. Hentschel, I. Tedjo, M. Probst, M. Wolter, M. Behr, C. Bischof, and T. Kuhlen. Interactive blood damage analysis for ventricular assist devices. IEEE Trans. Visualization and Computer Graphics, 14(6):1515–1522, 2008.

[40] C. Y Ip and A. Varshney. Salience-assisted navigation of very large landscape images. IEEE Trans. Visualization and Computer Graphics, 17(12):1737–1746, 2011.

[41] M. R. Jakobsen and K. Hornbæk. Sizing up visualizations: Effects of display size in focus+context, overview+detail, and zooming interfaces. In Proc. ACM CHI, pages 1451–1460, 2011.

[42] J. Jerald. The VR Book: Human-Centered Design for Virtual Reality. ACM and Morgan & Claypool, 2015.

[43] N. Kijmongsokilchai, A. Abdul-Rahman, and M. Chen. Empirically measuring soft knowledge in visualization. Computer Graphics Forum, 36(3):73–85, 2017.

[44] G. Kindlmann and C. Scheidegger. An algebraic process for visualization design. IEEE Trans. Visualization and Computer Graphics, 20(12):2181–2190, 2014.

[45] B. Laha, D. A. Bowman, and J. J. Socha. Effects of vr system fidelity on analyzing isosurface visualization of volume datasets. IEEE Trans. Visualization and Computer Graphics, 20(4):513–522, 2014.

[46] B. Laha, K. Sensharma, J. D. Schiffbauer, and D. A. Bowman. Effects of immersion on visual analysis of volume data. IEEE Trans. Visualization and Computer Graphics, 18(4):597–606, 2012.

[47] M. F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities? Vision Research, 41(25):3595–3565, 2001.

[48] C. R. Larsen, J. Oestergaard, B. S. Ottesen, and J. L. Soerensen. The efficacy of virtual reality simulation training in laparoscopy: a systematic review of randomized trials. Acta obstetricia et gynecologica Scandinavica, 91(9):1015–1028, 2012.

[49] A. T. Lee. Flight Simulation: Virtual Environments in Aviation. Routledge, 2005.

[50] G. S. Letterie. How virtual reality may enhance training in obstetrics and gynecology. American journal of obstetrics and gynecology, 187(3):S37–S40, 2002.

[51] C. Liu, O. Chapuis, M. Beaudouin-Lafon, E. Lecolinet, and W. Mackay. Effects of display size and navigation type on a classification task. In Proc. ACM CHI, pages 4147–4156, 2014.

[52] M. Lombard and T. Ditton. Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. In Proc. Workshop on Presence, 2000.

[53] S. J. Luck and E. K. Vogel. The capacity of visual working memory for features and conjunctions. Nature, 390(6657):279–281, 1997.

[54] J. Mackinlay. Automating the design of graphical presentations of relational information. ACM Trans. Graphics, 5(2):110–141, 1986.

[55] J. S. Matthys and B. R. Fagen. Visual control of foot placement when walking over complex terrain. Journal of Experimental Psychology: Human Perception and Performance, 40(1):106, 2014.

[56] N. Mennie, M. Hayhoe, and B. Sullivan. Look-ahead fixations: anticipatory eye movements in natural tasks. Experimental Brain Research, 179(3):427–442, 2007.

[57] H. C. Miles, S. R. Pop, S. J. Watt, G. P. Lawrence, and N. W. John. A review of virtual environments for training in ball sports. Computers & Graphics, 36(6):714–726, 2012.

[58] H. C. Miles, S. R. Pop, S. J. Watt, G. P. Lawrence, N. W. John, V. Perrot, P. Mallet, D. R. Mestre, and K. Morgan. Efficacy of a virtual environment for training ball passing skills in rugby. In Transactions on Computational Science XXIII, pages 98–117. Springer, 2014.

[59] H. C. Miles, S. R. Pop, S. J. Watt, G. P.Lawrence, and N. W. John. A review of virtual environments for trainingin ball sports. Computers & Graphics, 36(4):714–726, 2012.

[60] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: Understanding the roles of animation, interaction, and static presentation. Computer Graphics Forum, 31(3):1215–1224, 2012.

[61] T. R. Coles, D. Meglan, and N. W. John. The role of haptics in medical training simulators: A survey of the state of the art. IEEE Trans. Haptics, 4(1), 2011.

[62] F. Cosentino, N. W. John, and J. Vaarkamp. An overview of augmented and virtual reality applications in radiotherapy and future developments enabled by modern tablet devices. Journal of Radiotherapy in Practice, pages 1–15, 2013.

[63] T. M. Cover and J. A. Thomas. Elements of Information Theory. John Wiley & Sons, 2nd edition, 2006.

[64] C. Demiralp, C. E. Scheidegger, G. L. Kindlmann, D. H. Laidlaw, and J. Heer. Visual embedding: A model for visualization. IEEE Computer Graphics and Applications, 34(1):10–15, 2014.

[65] R. Desimone and J. Duncan. Neural mechanisms of selective visual attention. Annual Review of Neuroscience, 18(1):193–222, 1995.

[66] A. Diehl, A. Abdul-Rahman, M. El-Assaday, B. Bach, D. Keim, and M. Chen. VisGuides: A forum for discussing visualization guidelines. In Proc. EuroVis Short Papers, 2018.

[67] T. L. Diz, M. E. Papka, M. Pellegrino, and R. Stevens. Sharing visualization experience among remote virtual environments. In M. Chen, P. Townsend, and J. A. Vince, editors, High Performance Computing for Computer Graphics and Visualisation. Springer, 1996.

[68] N. Elmqvist. 3d visualization for nonspatial data: Guidelines and challenges, accessed in June 2018.

[69] D. C. V. Essen, W. T. Newsome, and J. H. Maunsell. The visual field representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability. Vision Research, 24(5):429–448, 1984.

[70] M. S. Graziano and C. G. Gross. Spatial maps for the control of movement. Current Opinion in Neurobiology, 8(2):195–201, 1998.

[71] M. S. Graziano, C. S. Taylor, T. Moore, and D. F. Cooke. The cortical control of movement revisited. Neuron, 36(3):349–362, 2002.

[72] E. J. Griffith, M. Koutek, F. H. Poss., , T. Heus, and H. J. J. Jonker. A
