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*



(a) an OpenSpace event [12] (b) surgical simulation [110] (c) gigaimage analytics [40] (d) sports training [58]

Fig. 1. Four examples of typical virtual environments (VEs) used for visualization applications. Designers, users, and other stakeholders of VEs often wonder what makes some systems and applications work and what hinders others from success. While a variety of practical factors have been, and should be, examined in probing this question, it is also desirable to identify the most fundamental measures that can frame such inquiries.

Abstract—Visualization and virtual environments (VEs) have been two interconnected parallel strands in visual computing for decades. Some VEs have been purposely developed for visualization applications, while many visualization applications are exemplary showcases in general-purpose VEs. Because of the development and operation costs of VEs, the majority of visualization applications in practice have yet to benefit from the capacity of VEs. In this paper, we examine this status quo from an information-theoretic perspective. Our objectives are to conduct cost-benefit analysis on typical VE systems (including augmented and mixed reality, theater-based systems, and large powerwalls), to explain why some visualization applications benefit more from VEs than others, and to sketch out pathways for the future development of visualization applications in VEs. We support our theoretical propositions and analysis using theories and discoveries in the literature of cognitive sciences and the practical evidence reported in the literatures of visualization and VEs.

Index Terms—Theory of visualization, virtual environments, virtual reality, augmented reality, mixed reality, cost-benefit analysis, information theory, cognitive sciences, visualization applications, immersive analytics, four levels of visualization.

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.

- *Min Chen is with the University of Oxford, UK. E-mail:min.chen@oerc.ox.ac.uk.*
- *Kelly Gaither and Brian McCann with the University of Texas at Austin, USA. E-mail:kelly@tacc.utexas.edu,bmccann@tacc.utexas.edu.*
- *Nigel W. John is with University of Chester, UK. E-mail:nigel.john@chester.ac.uk.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

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, gigaimage 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, transportion, 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 performance models for interactive and 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);

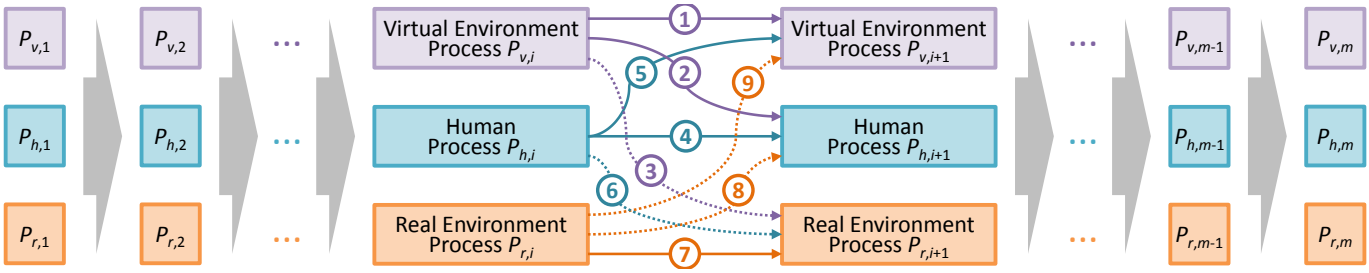


Fig. 2. A sequence of events in a VE can be considered as a series of processes flowing along a pathway in a complex space of all possible states of the entities involved. The main entities are the system of the VE and the human participant(s). In a mixed reality environment, parts of the reality are 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.

- *match* – the degree of correlation between the information conveyed by the VE and a participant's proprioceptive feedback about body movements; and
- *plot-interactivity* – the extent to which a participant can influence the storyline or the sequence of events in a VE.

These dimensions of immersion are considered to be measurable objectively and quantitatively. There have been a number of experiments designed to obtain these measurements for specific VEs.

Presence 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.

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 available information channels (visual, audio, etc.). while often incurring more costs than basic systems. This leads naturally to a question regarding the cost-benefit of different VEs. Because the dimensions of immersion and presence may not be easy to remember, we continue to use italic fonts for these terms in the remaining text.

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 innumerably 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.

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 know 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.

Meanwhile, a human participant can receive a variety of information conveyed by the VE system as indicated by connection ②, and machine-sensors can pick up aspects of a human state as indicated by connection ⑤. In a mixed reality environment, information is also passed from the real environment processes to the human participant and machine sensors as indicated by ⑧ and ⑨. When an object in the real environment is manipulated by a human participant or a device in the VE (e.g., a robotic arm), we conceptualize this phenomenon as information communication from a human process or a VE process to a real environment process.

We note that all of these processes receive information from a previous state, process the information, and deliver changes as a transformation to a new state. This processing flow bears a remarkable resemblance to a data analysis and visualization workflow [19]. Furthermore, the concept of immersion is a collective and accumulative attribute primarily about the machine-centric VE processes, while the concept of presence is a collective and accumulative attribute primarily about the human-centric processes. Hence, it is not surprising that VEs have been used for visualization applications. While VEs have emerged in the consumer gaming market, for the purposes of this paper, 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-feedbacks, etc.) be the *letters* of a very large *alphabet* $\mathbb{V}$, 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* $\mathbb{H}$, 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* $\mathbb{R}$.[1] 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., $\mathbb{V}$, $\mathbb{H}$, or $\mathbb{R}$) 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 $\mathbb{H}$. 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

[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.

Table 1. The design emphases of some typical VE systems, and their abstraction in terms of the cost-benefit measures. The numbers in black circled are used as references in the text whenever a characterization of a VR system relates to one of the three measures in the cost-benefit metric. The exact mathematical definitions for the formulae in columns of Alphabet Compression and Potential Distortion are given in Appendix A. The red text indicates the requirements that could lead to less optimization of the measure concerned (e.g., less alphabet compression and more cost).

Typical VE Uses	Design Emphasis	Alphabet Compression	Potential Distortion	Cost
Theater-based education (e.g., a large dome theater)	inclusion, surrounding, vividness, sense of believing, large audience	❶ $\mathcal{H}(\mathbb{V}) > \mathcal{H}(\mathbb{Z}_1)$ ❷ maximize $\mathcal{I}(\mathbb{V};\mathbb{Z}_1)$ ❸ minimize $\mathcal{H}(\mathbb{R})$ ❹ maximize $\mathcal{H}(\mathbb{Z}_i) - \mathcal{H}(\mathbb{Z}_{i+1})$	❺ minimize $\mathcal{D}_{KL}(\mathbb{V}'\|\|\mathbb{V})$	❻ maximize attention
Real-time mixed reality (e.g., an image-guided surgery system)	extension, *surrounding*, *vividness*, match, *naturalism*, *sense of co-presence*	❼ maximize $\mathcal{H}(\mathbb{Z}_i) - \mathcal{H}(\mathbb{Z}_{i+1})$ ❽ maximize $\mathcal{I}(\mathbb{V};\mathbb{Z}_1)$ ❾ maximize $\mathcal{H}(\mathbb{R})$	❿ minimize $\mathcal{D}_{KL}((\mathbb{V}\otimes\mathbb{R})'\|\|(\mathbb{V}\otimes\mathbb{R}))$ ⓫ minimize $\mathcal{D}_{KL}(\mathbb{Z}_1'\|\|\mathbb{Z}_i)$	⓬ minimize cognitive load
Large dataset visualization (e.g., corpus visualization using a large power-wall)	surrounding, plot-interactivity	⓭ maximize $\mathcal{H}(\mathbb{Z}_i) - \mathcal{H}(\mathbb{Z}_{i+1})$ ⓮ maximize $\mathcal{I}(\mathbb{V};\mathbb{Z}_1)$ ⓯ minimize $\mathcal{H}(\mathbb{V})$	⓰ minimize $\mathcal{D}_{KL}(\mathbb{Z}_1'\|\|\mathbb{Z}_i)$	⓱ minimize costs, e.g., cognitive load, time, error-related cost, ...
VR-based training (e.g., multi-player skill training in sports)	surrounding, match, plot-interactivity, all senses of presence	⓲ maximize $\mathcal{H}(\mathbb{Z}_i) - \mathcal{H}(\mathbb{Z}_{i+1})$ ⓳ maximize $\mathcal{I}(\mathbb{V};\mathbb{R})$	⓴ minimize $\mathcal{D}_{KL}(\mathbb{V}'\|\|\mathbb{V})$ ㉑ minimize $\mathcal{D}_{KL}(\mathbb{V}'\|\|\mathbb{R})$	㉒ minimize financial cost ㉓ minimize cognitive load

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 $\mathbb{V}$, hence an increase of the Shannon entropy of $\mathbb{V}$. 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 $\mathbb{Z}_1, \mathbb{Z}_2, \ldots, \mathbb{Z}_n$, collectively as *visualization alphabets*. Unlike alphabet $\mathbb{V}$, 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 $\mathbb{V}$. But others, such as human perception about various data patterns, will be part of the human alphabet $\mathbb{H}$. In a mixed reality environment, some of these alphabets will be a constituent part of the real environment alphabet $\mathbb{R}$. It is not difficult to imagine that in some cases, the availability of the reality, i.e., letters in $\mathbb{R}$ are limited or too costly; hence one uses aspects of a VE alphabet $\mathbb{V}$ to simulate these letters in $\mathbb{R}$. In other cases, the desired immersion and presence cannot be achieved entirely using a VE alphabet $\mathbb{V}$ or it is too costly to achieve; hence one mixes some aspects of a reality alphabet $\mathbb{R}$ with those of $\mathbb{V}$. 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, $\mathbb{Z}_1, \mathbb{Z}_2, \ldots, \mathbb{Z}_n$, 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 $\mathcal{H}(\mathbb{Z}_i)$ be the Shannon entropy of alphabet $\mathbb{Z}_i$. When it is transformed to $\mathbb{Z}_{i+1}$ (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, $\mathcal{H}(\mathbb{Z}_i) - \mathcal{H}(\mathbb{Z}_{i+1})$. 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 $\mathbb{Z}_{i+1}$ based on a third-party and likely-subjective metric, we can consider a reconstruction of $\mathbb{Z}_i$ from $\mathbb{Z}_{i+1}$. 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 $\mathcal{D}_{KL}(\mathbb{Z}_i'||\mathbb{Z}_i)$, where $\mathbb{Z}_i'$ is reconstructed from $\mathbb{Z}_{i+1}$. 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):

$$\begin{aligned}\frac{\text{Benefit}}{\text{Cost}} &= \frac{\text{Alphabet Compression} - \text{Potential Distortion}}{\text{Cost}} \\ &= \frac{\mathcal{H}(\mathbb{Z}_i) - \mathcal{H}(\mathbb{Z}_{i+1}) - \mathcal{D}_{KL}(\mathbb{Z}_i'||\mathbb{Z}_i)}{\text{Cost}}\end{aligned} \tag{1}$$

The exact mathematical definitions of $\mathcal{H}$ and $\mathcal{D}_{KL}$ 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 $\mathbb{V}$ includes primarily the data being visualized, visual imageries, commentary voice, and accompanying music. The initial visualization alphabet (i.e., the data alphabet) $\mathbb{Z}_1$ is a subset of $\mathbb{V}$ ❶. Here we use the circled number to relate the statement to the mathematical description in Table 1. Hence, ideally, the mutual information $\mathcal{I}(\mathbb{V};\mathbb{Z}_1)$ between the two alphabets should be maximized, and contain roughly the same amount of the entropy of $\mathbb{Z}_1$ ❷. The additional visual and audio effects result in additional entropy $\mathcal{H}(\mathbb{V}) - \mathcal{H}(\mathbb{V}|\mathbb{Z}_1)$. Meanwhile, the design emphasis on *inclusion* implies the minimization of the entropy of the reality $\mathcal{H}(\mathbb{R})$ ❸.

The final visualization alphabet (i.e., the decision alphabet) $\mathbb{Z}_n$ 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 $\mathbb{V}$ as a reconstructed alphabet $\mathbb{V}'$ [5]. The potential distortion is defined as the Kullback-Leibler divergence $\mathcal{D}_{KL}(\mathbb{V}'||\mathbb{V})$. When the decision alphabet $\mathbb{Z}_n$ 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, $\mathcal{I}(\mathbb{Z}_n;\mathbb{V})$. If the decision alphabet $\mathbb{Z}_n$ 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 [4]. The advantage of maintaining a large and complex alphabet $\mathbb{V}$ 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 [6]. 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., $\mathbb{Z}_1$), and the visualization tasks are represented by the decision alphabet (i.e., $\mathbb{Z}_n$). Most use cases of mixed reality systems have a clearly defined tasks, hence $\mathbb{Z}_n$ is expected to have a much smaller entropy. The alphabet compression from $\mathbb{Z}_1$ towards $\mathbb{Z}_n$ is thus critical [7].

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 $\mathbb{V}$ 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 $\mathbb{V} \otimes \mathbb{R}$ 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 $\mathbb{V}$ and $\mathbb{Z}_1$ should be maximized [8], while the visualization process at each stage should minimize the potential distortion due to the mixed reality setting [11]. In terms of immersion and presence, the real environment alphabet $\mathbb{R}$ 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 [9]. 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 $\mathbb{V}$ (assuming $\mathbb{Z}_1 \subseteq \mathbb{V}$) and $\mathbb{R}$. The cognitive load for *match* can be rather high [12].

In terms of cost-benefit analysis, the alphabet compression from $\mathbb{V} \otimes \mathbb{R}$ to the decision alphabet $\mathbb{Z}_n$ 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 $\mathbb{V}$ against $\mathbb{R}$ [10]. In comparison with the idealized VE system that captures all required aspects of the reality $\mathbb{R}$, 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 $\mathbb{V}$ about the reality. The more overlapping between the virtuality $\mathbb{V}$ and the reality $\mathbb{R}$, the more mutual information $\mathcal{I}(\mathbb{V};\mathbb{R})$, 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 $\mathbb{V}$ against $\mathbb{R}$.

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 $\mathbb{V}$ usually focuses on the data alphabet $\mathbb{Z}_1$. The visualization tasks are expected to be the same for a gigapixel display and a few megapixel displays. The decision alphabet $\mathbb{Z}_n$ 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-

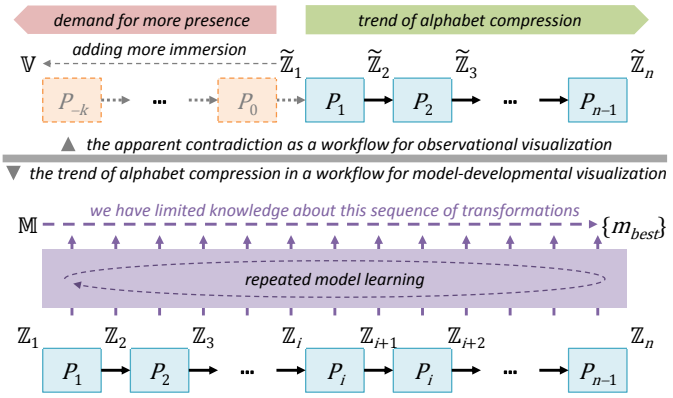


Fig. 3. If a VE-based training system were considered as a workflow for observational visualization (above), there would be a contradiction with the cost-benefit metric [19] for optimizing visualization processes. However, it is more appropriate to consider a VE-based training system as a workflow for 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.

pression, so one can reach the decision using fewer steps [13], while trying to minimize the potential distortion between any intermediate decision alphabet and the original data alphabet [16]. Meanwhile, for observational, analytical, and model-developmental visualization, visual embellishment should be minimized [14][15]. The process should also minimize a variety of costs, such as cognitive load, time needed for performing the tasks, adverse cost of errors, and so on [17].

For a "big data" application, the entropy of $\mathbb{Z}_1$ will be very high. The entropy of a dataset is not necessarily defined by the size of the dataset. It represents the uncertainty about the amount of potential variations in an alphabet. In a specific application context (e.g., landscape images in [40]), the larger the dataset, the more potential variations, and therefore the greater the entropy. Every time, a user observes a portion of the dataset, some uncertainty disappears. Comparing a gigapixel display against a megapixel display, the costs for a user to observe the visualization of a very large dataset include the number of interactions, the amount of body movement, and the imposition for using the equipment (e.g., finance, inconvenience, etc.). In general, we expect a megapixel display requires more interaction, a gigapixel display requires more body movement, and encounters more imposition.

One major cognitive difficulty in observing a very large visualization is the need to remember what was observed a moment ago. With a gigapixel display, this means revisiting a portion of the display with a quick glance at a distance or by walking back to have another close look at the portion again. With a megapixel display, this typically means relocating the portion concerned through a sequence of interactions, which may not always be straightforward. This may inevitably result in poor external memorization, cause some potential distortion, and incur additional cognitive load.

Hence, the type of applications that can benefit from a gigapixel display features datasets unfamiliar to a user, with uncertainty (i.e., potential variations) across different parts of a very large visualization, at different resolutions (i.e., zoom factors). The user's prior knowledge about the dataset and its visualization usually reduces the cost-benefit of using a gigapixel display.

4.4 VE-based Training Systems

One major application of VEs is training, for example, in medicine and sports. The basic workflow for VE-based training involves repeated exercises where a user receives various stimuli and responds to the stimuli with appropriate actions. In most cases, the stimuli are visual imagery, and the actions are the user's motions or interactions. While different applications may place emphasis on different dimensions of immersion and presence, all these dimensions normally have a positive role to play if they can be made available.

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 $\widetilde{\mathbb{Z}}_1$ and all the possible actions in response to the visual stimuli as a decision alphabet $\widetilde{\mathbb{Z}}_n$. Here we use $\widetilde{\mathbb{Z}}$ 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 $\widetilde{\mathbb{Z}}_1$ to $\widetilde{\mathbb{Z}}_n$, 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 $\widetilde{\mathbb{Z}}_1$ 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 $\widetilde{\mathbb{Z}}_1$ is pre-processed in the direction towards $\widetilde{\mathbb{Z}}_n$. However, the practical experience suggests that the demand is to embellish $\widetilde{\mathbb{Z}}_1$ 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 $\mathbb{Z}_1 = \mathbb{M} \times \mathbb{I} \times \mathbb{O}$. 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 $m_{best} \in \mathbb{M}$ as the chosen optimal model. It can be written as $\mathbb{Z}_n = \{[m_{best}, i, o] \,|\, o = m_{best}(i), m_{best} \in \mathbb{M}, i \in \mathbb{I}, o \in \mathbb{O}\})$. Since the number of letters in $\mathbb{Z}_1$ is at the scale of $||\mathbb{M}|| \cdot ||\mathbb{I}|| \cdot ||\mathbb{O}||$, and that of $\mathbb{Z}_n$ is at the scale of $||\mathbb{I}||$, $\mathbb{Z}_n$ has a 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] [18].

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 $m_{best} \in \mathbb{M}$ 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 [19] within the constraints of financial cost [22]. Normally we would like users of a VR-based training system to have the most faithful perception of the visual stimuli [20]. 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 [21]. Meanwhile, it is likely that the more immersion and presence, the less cognitive load [23] 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 distortion* 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 compression* 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 humans' 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 necessary. 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 demands more attention from participants, and can potentially facilitate the delivery of more information for educational purposes [1][3][6]. For observational and analytical visualization, on the other hand, such a demand has to be carefully managed [14][15]. The more cognitive resources 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) [17].

Visual Search and **Working Memory**. Humans are efficient visual searchers. Cognitive studies have confirmed humans' ability to understand 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 external 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 bandwidth 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 previous 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 attention becomes more effective [16]. 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 [17].

In real-time mixed reality applications, the challenge of the *match* dimension is often related to visual search and memorization. The integrated presentation of two types of visual stimuli (i.e., virtual and real objects) is not what one encounters in everyday life. Hence, this unfamiliarity may reduce humans' aforementioned visual search capability [10][11]. 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 types of stimuli in order to reduce the user's cognitive load for "mental 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 [12].

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 applications 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 dissemination 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 experience 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 powerwalls). They compared a large display (6m $\times$ 2.2m, 10,800 $\times$ 4,096 pixels) with a 24-inch desktop monitor. They examined visualization 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 desktop 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 [13]. The dependent variables (e.g., accuracy and response time) of the study relate directly to the potential distortion [16] and cognitive cost [17] 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 higher 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 VEs 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 Müller 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 [13][17]?

Surgical Training Domain experts in medicine are early adopters of VEs, 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 $\mathbb{V}$ 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 $\mathbb{H}$ 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 $\mathbb{R}$ 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 VEs has already been confirmed by many practitioners.

Minimally invasive surgical (MIS) procedures currently provide the most opportunities for surgical training using VEs (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 effort to increase the space of the real environment alphabet $\mathbb{R}$ [9]. 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) [7] with minimal potential distortion will be paramount [10][11].

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 $\mathbb{R}$ and the data to be visualized $\mathbb{Z}_1$, 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 510mm to 12mm. 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 VEs to provide alternative ways of training a skill, and analysing performance. Miles et al. [57] provide a comprehensive review of the use of VEs 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 VEs; 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 about 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 [22]. 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 [20], it may also introduce new variables (e.g., fatigue and discomfort, view distortion) that are undesirable to be part of the model [23]. During a training session, a player processes visual stimuli at a very high speed, achieving extremely high rate of alphabet compression [18]. 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:immersiveanalytics.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 swarming 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, *VisGuides* (visguides.dbvis.de), was established specifically for discussing guidelines in visualization [30]. Before long, several VE-related discussion threads emerged, including "*(Dont / 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 transformations feature significant alphabet compression and potential distortion, since the resolution of the data variable (e.g., [0.000, 1000.000] is usually higher than that of visual channel (e.g., brightness [0, 255]), which is higher than the perceived data variables (e.g., 12 levels of brightness). The comparison among different visual channels in visualization may use different criteria, such as four binary criteria by Bertin [11], accuracy by Mackinlay [54], and pre-attentiveness by Williams [113], Treisman [104], and many others in psychology. In terms of cost-benefit analysis, accuracy corresponds to the potential distortion in combined P_1 and P_2, while pre-attentiveness corresponds to the cost of P_2. For nominal data variables, Bertin's association criterion corresponds to the potential distortion in combined P_1 and P_2. For ordinal variables (or interval or ratio variables) , Bertin's orderedness (or quantifiability) corresponds to the cost and potential distortion in reconstructing from the perceived data variable to the original data variable. Bertin's selectivity corresponds to the alphabet compression of P_2 for all data types.

Hence, it is highly desirable to consider all visual channels in a multifaceted manner, e.g., using the cost-benefit metric in Eq. 1. As there are several dozens of visual channels [23], and likely many more for visualization in VEs, many empirical studies will be needed to establish the cost-benefit measures of these visual channels. For the time being, we may use the approach of cost-benefit analysis to consider and compare visual channels theoretically.

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. Yudkowsky, A. Wallace, C. Luciano, P. P. Banerjee, S. H. Rizzi, and F. T. Charbel. Virtual reality training in neurosurgy: 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] A. D. Baddeley and G. Hitch. Working memory. *Psychology of Learning and Motivation*, 8:47–89, 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.

[10] N. Bernstein. *The coordination and regulation of movement.* London, 1967.

[11] J. Bertin. *Semiology of Graphics*. University of Wisconsin Press, 1983.

[12] A. Bock, E. Axelsson, C. Emmart, M. Kuznetsova, C. Hansen, and A. Ynnerman. Openspace: Changing the narrative of public dissemination in astronomical visualization from what to how. *IEEE Computer Graphics and Applications*, 38(3):44–57, 2018.

[13] D. A. Bowman, J. L. Gabbard, and D. Hix. A survey of usability evaluation in virtual environments: Classification and comparison of methods. *Prsence*, 11(4):404–424, 2002.

[14] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proc. I3D*, pages 35–38, 1997.

[15] D. A. Bowman, R. P. McMahan, and E. D. Ragan. Questioning naturalism in 3D user interfaces. *Communications of the ACM*, 55(9):78–88, 2012.

[16] K. Brodlie, J. Brooke, M. Chen, D. Chisnall, A. Fewings, C. Hughes, N. W. John, M. W. Jones, M. Riding, and N. Roard. Visual supercomputing – technologies, applications and challenges. *Computer Graphics Forum*, 24(2):217245, 2005.

[17] A.-M. Brouwer and D. C. Knill. The role of memory in visually guided reaching. *Journal of Vision*, 7(5):6–6, 2007.

[18] M. Chen, M. Feixas, I. Viola, A. Bardera, H.-W. Shen, and M. Sbert. *Information Theory Tools for Visualization*. A K Peters, 2016.

[19] M. Chen and A. Golan. What may visualization processes optimize? *IEEE Trans. Visualization and Computer Graphics*, 22(12):2619–2632, 2016.

[20] M. Chen, G. Grinstein, C. R. Johnson, J. Kennedy, and M. Tory. Pathways for theoretical advances in visualization. *IEEE Computer Graphics & Applications*, to appear, 2017.

[21] M. Chen and H. Jänicke. An information-theoretic framework for visualization. *IEEE Trans. Visualization and Computer Graphics*, 16(6):1206–1215, 2010.

[22] M. Chen and J. Kennedy. References for theoretic researches in visualization, 2017. https://sites.google.com/site/drminchen/themes/theory-refs.

[23] M. Chen, K. Mueller, and A. Ynnerman. Fusion of visual channels. In C. Hansen, M. Chen, C. R. Johnson, A. Kaufman, and H. Hagen, editors, *Scientific Visualization*, pages 119–127. Springer, 2014.

[24] D. Coffey, F. Korsakov, M. Ewert, H. HaghShenas, L. Thorson, A. Ellingson, D. Nuckley, and D. F. Keefe. Visualizing motion data in virtual reality: Understanding the roles of animation, interaction, and static presentation. *Computer Graphics Forum*, 31(3pt3):1215–1224, 2012.

[25] 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.

[26] 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.

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

[28] 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.

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

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

[31] T. L. Disz, 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.

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

[33] 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.

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

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

[36] E. J. Griffith, M. Koutek, F. H. Post, , T. Heus, and H. J. J. Jonker. A 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*, 1:262271, 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. Saliency-assisted navigation of very large landscape images. *IEEE Trans. Visualization and Computer Graphics*, 17(12):1737–1746, 2011.

[41] M. R. Jakobsen and K. Hornbaek. 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. Kijmongkolchai, 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):3559–3565, 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. Matthis and B. R. Fajen. 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*, 36:714–726, 2012.

[60] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the realityvirtuality continuum. In *Photonics for industrial applications*, pages 282–292, 1995.

[61] G. A. Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2):81, 1956.

[62] K. Moreland. Redirecting research in large-format displays for visualization. In *Proc. LDAV*, pages 91–95, 2012.

[63] K. Moreland. A survey of visualization pipeline. *IEEE Trans. Visualization and Computer Graphics*, 19(3):367–378, 2013.

[64] C. Müller, M. Krone, K. Scharnowski, G. Reina, and T. Ertl. An evaluation of the utility of large high-resolution displays for comparative scientific visualisation. *International Journal of Software and Informatics*, 9(3), 2016.

[65] J. Najemnik and W. S. Geisler. Optimal eye movement strategies in visual search. *Nature*, 434(7031):387–391, 2005.

[66] K. Nowak. Defining and differentiating copresence, social presence and presence as transportation. In *Proc. 4th International Workshop on Presence*, 2001.

[67] Oculus. Introduction to best practices, accessed in June 2018.

[68] A. Oliva and A. Torralba. Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, 155:23–36, 2006.

[69] S. K. Ong and A. Y. C. Nee, editors. *Virtual and Augmented Reality Applications in Manufacturing*. Springer, 2004.

[70] J. K. O'regan, R. A. Rensink, and J. J. Clark. Change-blindness as a result of 'mudsplashes'. *Nature*, 398(6722):34–34, 1999.

[71] C. Papadopoulos and A. E. Kaufman. Acuity-driven gigapixel visualization. *IEEE Trans. Visualization and Computer Graphics*, 19(12):2886–2895, 2013.

[72] C. Papadopoulos, K. Petkov, A. E. Kaufman, and K. Mueller. The reality deck – an immersive gigapixel display. *IEEE Computer Graphics & Applications*, 35(1):33–45, 2015.

[73] J. Pelz, M. Hayhoe, and R. Loeber. The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, 139(3):266–277, 2001.

[74] M. I. Posner, C. R. Snyder, and B. J. Davidson. Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109(2):160, 1980.

[75] K. Reda, A. Febretti, A. Knoll, J. Aurisano, J. Leigh, A. Johnson, M. E. Papka, and M. Hereld. Visualizing large, heterogeneous data in hybrid-reality environments. *IEEE Computer Graphics & Applications*, 33(4):38–48, 2013.

[76] K. Reda, A. E. Johnson, and M. E. Papka. Effects of display size and resolution on user behavior and insight acquisition in visual exploration. In *Proc. ACM CHI*, pages 2759–2768, 2015.

[77] R. A. Rensink. Change detection. *Annual Review of psychology*, 53(1):245–277, 2002.

[78] D. Robinson. The mechanics of human saccadic eye movement. *The Journal of Physiology*, 174(2):245, 1964.

[79] J. M. Rolfe and K. J. Staples, editors. *Flight Simulation*. Cambridge University Press, 2008.

[80] R. A. Ruddle, R. G. Thomas, R. S. Randell, P. Quirke, , and D. Treanor. Performance and interaction behaviour during visual search on large high-resolution displays. *Information Visualization*, 14(2), 2013.

[81] J. A. Saunders and D. C. Knill. Visual feedback control of hand movements. *Journal of Neuroscience*, 24(13):3223–3234, 2004.

[82] D. W. Schloerb. A quantitative measure of telepresence. *Presence*, 4:64–80, 1995.

[83] M. A. Schnabel, X. Wang, H. Seichter, and T. Kvan. From virtuality to reality and back. volume 1, page 15, 2007.

[84] M. J. Schuemie, P. van der Straaten, M. Krijn, and C. A. P. G. van der Mast. Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, 4(2):183–201, 2001.

[85] A. Seydell, B. C. McCann, J. Trommershäuser, and D. C. Knill. Learning stochastic reward distributions in a speeded pointing task. *Journal of Neuroscience*, 28(17):4356–4367, 2008.

[86] N. E. Seymour. Vr to or: a review of the evidence that virtual reality simulation improves operating room performance. *World journal of surgery*, 32(2):182–188, 2008.

[87] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[88] T. B. Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, 1:120–126, 1992.

[89] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. IEEE Symposium on Visual Languages*, pages 336–343, 1996.

[90] J. Short, E. Williams, and B. Christie. *The social psychology of telecommunications*. Wiley, London, 1976.

[91] T. Sielhorst, M. Feuerstein, and N. Navab. Advanced medical displays: A literature review of augmented reality. *Journal of Display Technology*, 4(4):451–467, 2008.

[92] D. J. Simons and D. T. Levin. Change blindness. *Trends in Cognitive Sciences*, 1(7):261–267, 1997.

[93] C. R. Sims, R. A. Jacobs, and D. C. Knill. An ideal observer analysis of visual working memory. *Psychological Review*, 119(4):807, 2012.

[94] M. Slater, Y. Chrysanthou, and A. Steed. *Computer Graphics and Virtual Environments: From Realism to Real-Time*. Addison Wesley, 2001.

[95] M. Slater, A. Steed, and M. Usoh. Being there together: Experiments on presence in virtual environments (1990s). Technical report, Department of Computer Science, University College London, UK, 2013.

[96] M. Slater and S. Wilbur. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence*, 6, 1997.

[97] J. Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42(4):72–93, 1992.

[98] S. Su, A. Chaudhary, P. OLeary, B. Geveci, W. Sherman, H. Nieto, and L. Francisco-Revilla. Enabling scientific workflows in virtual reality. In *Proc. ACM VRCIA*, pages 155–162, 2006.

[99] G. K. L. Tam, V. Kothari, and M. Chen. An analysis of machine- and human-analytics in classification. *IEEE Trans. Visualization & Computer Graphics*, 23(1), 2017.

[100] V. E. Taylor, R. Stevens, and T. Canfield. Performance models of interactive and immersive visualization for scientific visualization. In M. Chen, P. Townsend, and J. A. Vince, editors, *High Performance Computing for Computer Graphics and Visualisation*. Springer, 1996.

[101] E. Todorov and M. I. Jordan. Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11):1226–1235, 2002.

[102] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, 113(4):766, 2006.

[103] M. Tory and T. Moller. Rethinking visualization: A high-level taxonomy. In *Proc. IEEE Information Visualization*, pages 151–158, 2004.

[104] A. Treisman. Focused attention in the perception and retrieval of multidimensional stimuli. *Attention, Perception, & Psychophysics*, 22(1), 1977.

[105] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980.

[106] J. Trommershäuser, L. T. Maloney, and M. S. Landy. Statistical decision theory and the selection of rapid, goal-directed movements. *Journal of the Optical Society of America A*, 20(7):1419–1433, 2003.

[107] E. Tufte. *The Visual Display of Quantitative Information*. 1983.

[108] A. van Dam, A. S. Forsberg, D. H. Laidlaw, J. J. LaViola, Jr., and R. M. Simpson. Immersive vr for scientific visualization: a progress report. *IEEE Computer Graphics & Applications*, 20(6):26–52, 2010.

[109] J. J. van Wijk. The value of visualization. In *Proc. IEEE Visualization*, pages 79–86, 2005.

[110] P.-F. Villard, D. P. Vidal, L. Ap Cenydd, R. Holbrey, S. Pisharody, S. Johnson, A. Bulpitt, N. W. John, F. Bello, and D. Gould. Interventional radiology virtual simulator for liver biopsy. *International Journal of Computer Assisted Radiology and Surgery*, 9(2):255–267, 2014.

[111] J. A. Vince. *Essential Virtual Reality Fast*. Springer, 2013.

[112] P. L. Weiss, E. A. Keshner, and M. F. Levin, editors. *Virtual Reality for Physical and Motor Rehabilitation*. Springer, 2016.

[113] L. Williams. The effects of target specification on objects fixated during visual search. *Acta Psychologica*, 27:355–415, 1967.

[114] J. M. Wolfe. Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, 1(2):202–238, 1994.

[115] D. M. Wolpert and Z. Ghahramani. Computational principles of movement neuroscience. *Nature Neuroscience*, 3:1212–1217, 2000.

[116] L. Xu, T. Y. Lee, and H. W. Shen. An information-theoretic framework for flow visualization. *IEEE Trans. Visualization and Computer Graphics*, 16(6):1216–1224, 2010.

[117] A. L. Yarbus. *Eye Movements During Perception of Complex Objects*, pages 171–211. Boston, MA, 1967.

[118] B. Yost and C. North. The perceptual scalability of visualization. *IEEE Trans. Visualization and Computer Graphics*, 12(5):837–844, 2006.

[119] B. Zendejas, R. Brydges, S. J. Hamstra, and D. A. Cook. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. *Annals of surgery*, 257(4):586–593, 2013.

APPENDICES

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*

APPENDIX A MATHEMATICAL DEFINITIONS OF INFORMATION-THEORETIC MEASURES

In this appendix, we give the mathematical definitions of several information-theoretic measures mentioned in Section 4 including Table 1 in order to maintain the desirable self-containment. For further discussions on these measures, please consult textbooks such as [27].

An alphabet $\mathbb{X}$ is a variable associated with a probability distribution. For each letter (i.e., valid value) of the alphabet, i.e., $x \in \mathbb{X}$, $p(x)$, $p(x)$ is the probability of x that may appear such that $\sum_{\forall x \in \mathbb{X}} p(x) = 1$. *Shannon entropy* $\mathcal{H}$ measures the average amount of uncertainty of an alphabet $\mathbb{X}$ as:

$$\mathcal{H}(\mathbb{X}) = -\sum_{\forall x \in \mathbb{X}} p(x) \log_2 p(x)$$

Here we use base-2 logarithms so that the measurement will have *bit* as it unit. Shannon entropy is always non-negative. When $\mathcal{H}(\mathbb{X}) = 0$, it implies that there is only one probable letter in $\mathbb{X}$ with absolute certainty.

Given an alphabet $\mathbb{X}$ that is associated with two probability distributions $p(x)$ and $q(x)$. We may consider a scenario that $q(x)$ is the original probability distribution of $\mathbb{X}$, and $p(x)$ is the current probability distribution after some events. We denote $\mathbb{X}'$ as an alphabet with the same set of letters as $\mathbb{X}$ but a different probability distribution $p(x)$. *Kullback-Leibler divergence* measures the change from $q(x)$ to $p(x)$ as:

$$\mathcal{D}_{KL}(\mathbb{X}'||\mathbb{X}) = \sum_{\forall x \in \mathbb{X}} p(x) \log_2 \frac{p(x)}{q(x)}$$

Here the conventions for conditions when $p(x) = 0$ or $q(x) = 0$ are $0 \log_2 \frac{0}{0} = 0$, $0 \log_2 \frac{0}{q} = 0$, and $p \log_2 \frac{p}{0} = \infty$. When $p(x) = q(x), \forall x \in \mathbb{X}$, $\mathcal{D}_{KL}(\mathbb{X}'||\mathbb{X}) = 0$, meaning that there is no difference between $\mathbb{X}$ and $\mathbb{X}'$. Kullback-Leibler divergence is sometimes referred to as Kullback-Leibler distance, but it is not a true distance metric since it is not symmetric, i.e., it is not assured that $\mathcal{D}_{KL}(\mathbb{X}'||\mathbb{X}) \equiv \mathcal{D}_{KL}(\mathbb{X}||\mathbb{X}')$.

Given two alphabets $\mathbb{X}$ with probability distribution $p(x)$ and $\mathbb{Y}$ with probability distribution $q(y)$, *Mutual Information* measures the amount of information shared between the two alphabets as:

$$\mathcal{I}(\mathbb{X};\mathbb{Y}) = \sum_{\forall x \in \mathbb{X}} \sum_{\forall y \in \mathbb{Y}} r(x,y) \frac{r(x,y)}{p(x)q(y)}$$

where $r(x,y)$ is the probability of two letters, $x \in \mathbb{X}$ and $y \in \mathbb{Y}$, occurring together. It is not difficult to derive $\mathcal{I}(\mathbb{X};\mathbb{X}) = \mathcal{H}(\mathbb{X})$, meaning that the information shared by an alphabet with itself its the Shannon entropy of the alphabet.

APPENDIX B MORE DETAILED DISCUSSIONS ON EVIDENCE FROM COGNITIVE SCIENCE

In this appendix, 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*. We use ▶ at the beginning of a paragraph to indicate our observations and remarks.

Attention. Attention is a complex cognitive function that selects an aspect of external information (e.g., visual, audio, smell, etc.) or internal events (e.g., thoughts) and maintains a certain level of awareness. 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 anatomy of the human eye reflects the compromises necessary in applying limited attentional resources to varying task demands. The foveated structure of the eye imposes a substantial constraint on the human visual system. We can only physically direct our gaze, and consequently a huge proportion of the *neural resources* in our visual system [33], towards one small area of visual space at a time. To compensate for this, we have evolved sophisticated selective visual attention circuitry allowing us to rapidly redeploy these neural resources as necessary [29].

Eye movements are the direct consequence of shifts in *overt attention*. The fine coordination of eye, hand and body movements provide objective details about the organization of attention, working memory and sensorimotor control [37]. Eye movements reflect information retrieval relevant to the current visual task [117]. The mechanical costs of eye movements are inconsequential [78]. We make many gaze shifts during our day to day activities with varying magnitude, timing, and apparent purpose [47].

▶ 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 necessary. 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 demands more attention from participants, and can potentially facilitate the delivery of more information for educational purposes. For observational and analytical visualization, on the other hand, such a demand has to be carefully managed. The more cognitive resources 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).

Visual Search. Humans are efficient visual searchers. Cognitive studies have confirmed humans' ability to understand a visual scene at a glance [68], to search for known signals embedded in visual noise [65], to identify outlier targets rapidly [105, 114], to take advantage of spatial cuing [74], and to predict probable locations for targets [102].

Working Memory. Retention, on the other hand, is not our strength. Humans' short-term (verbal) memory is famously limited to around seven items [61]. Modern theory emphasizes the importance of *working memory* on cognitive tasks [6], [5]. Working memory includes both visual and phonological (verbal) components, mirroring perceptual modalities. The capacity of visual working memory is difficult to measure precisely. It has been estimated that we can store a conjunction of features representing about four discrete objects [53]. More recently, information theoretic models implying a flexibly allocated capacity account for behavior better than models with a fixed number of slots [93]. Regardless, there is general agreement that working memory is a highly constrained resource.

The limits of visual working memory are more apparent in what we miss than in what we retain. In the phenomenon of *change blindness* [92], [77] large objects in a scene can be introduced, changed, or completely removed without an observer being aware. Visual awareness of the change is masked with a short visual interruption such as a flash, cut, or eye movement [70]. Change blindness is the consequence of selective attention and allocation of limited working memory resources.

▶ Most visualization techniques provide an effective means for external 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 bandwidth 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 previous 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 attention becomes more effective. 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.

► In real-time mixed reality applications, the challenge of the *match* dimension is often related to visual search and memorization. The integrated presentation of two types of visual stimuli (i.e., virtual and real objects) is not what one encounters in everyday life. Hence, this unfamiliarity may reduce humans' aforementioned visual search capability. 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 types of stimuli in order to reduce the user's cognitive load for "mental 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.

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. We typically control only very specific task relevant dimensions. Despite considerable variability across a huge number of kinematic degrees of freedom, the error in a blacksmith's strike point is measured in mm [10]. Optimal feedback control seems to provide a compelling mathematical account of this sort of minimization of error along task-relevant dimensions potentially at the expense of increased variability along task-irrelevant dimensions [101]. The number of variables in human movement control is estimated to be about 2^{600}, with considerably simplified assumptions about motor activations [115]. This is more than the number of atoms in the universe.

Cognitive studies have confirmed many fascinating properties of humans' motor coordination. These include eye-head-hand coordination with precise timing [73], peripheral monitoring of the position of the finger and making corrections to match the intended trajectory [81], taking into consideration our own intrinsic motor variability [106] and environmental variability [85], and making look-ahead fixations to improve the accuracy of grasping [17, 56]. The spatiotemporal complexity of humans' motor coordination challenges any attempt to define a model accurately. For example, visual references to stepping locations are at least two steps ahead in order for humans to maintain an efficient walking gait [55]. The spatiotemporal coordination of eye and body movements is tightly controlled. Pointing gaze towards a location in space commits a huge proportion of neural resources to that area. There is a blurry line between high-level motor planning areas and high-level sensory, attention and association areas. There is a continuous remapping of sensory and remembered information into a manual, or at least motor-centric mapping of space [34, 35]. Organizing sensorimotor control is one of, if not the most important function of cerebral cortex. Despite the mathematical complexity, our brains have been optimized by evolution to solve this particular problem very efficiently.

► 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.

Appendix C More Detailed Discussions on Evidence from Practical Applications

Visualization has been a ubiquitous tool for supporting scientific and scholarly activities in almost all disciplines. Many visualization applications 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 dissemination and public engagement (e.g., [12]), and culture and heritage (e.g., [1]).

Many have provided evidence of the benefit of stereoscopy displays in performing tasks involving complex and unfamiliar 3D objects [45, 46] and very large 3D datasets [71, 72]. In the former case, the complexity and unfamiliarity create a difficulty for viewers to use their knowledge to reconstruct the data from a flat visualization (cf. viewing a 3D model of everyday furniture vs. viewing a medical volume dataset) or have to use more exploration interactions to self-correct errors in perception of the objects [21]. Hence stereoscopy displays can help reduce the potential distortion in perception. In the latter case, the benefits may be obtained in conjunction with large display systems. When the large 3D environmental models are unfamiliar to viewers and are displayed using smaller flat displays, they would need a fair amount of memory capacity, thus cognitive load, to build a mental overview of such models. The need for memorizing something about different parts would also restrict the viewers' capability of visual search and selective attention. Hence, large stereoscopy displays can provide means for reduce potential distortion and cognitive load. As some studies showed that the benefit of stereoscopy displays may not be obvious for viewing familiar visualization (e.g., [57]), further research will be necessary to separate the benefits due to the size of the displays (i.e., *surrounding*) and the stereoscopy functionality (i.e., *vividness*).

In this appendix, we examine three visualization applications in VEs, and discuss the cost-benefit of such applications based on the experience reported in the literature. Similarly, we use ► at the beginning of a paragraph to indicate our observations and remarks.

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 a mixed set of conclusions about the relative merits of such VE systems. Moorland [62] summarized a number of observations about the 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 powerwalls). They compared a large display (6m $\times$ 2.2m, 10,800 $\times$ 4,096 pixels) with a 24-inch desktop monitor. They examined visualization 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 desktop 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 and cognitive cost 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 higher 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 VEs 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 (e.g., relationships among 400 or 4,000

structures), and the need for cost reduction by using some analytics algorithms to prioritize the comparative activities.

Surgical Training Domain experts in medicine are early adopters of VEs, 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.

Minimally invasive surgical (MIS) procedures currently provide the most opportunities for surgical training using VEs, and several commercial systems are available from companies such as 3D Systems Healthcare (CO, USA) and Mentice (Gothenburg, Sweden). MIS procedures may be within the abdominal or pelvic cavities (laparoscopy) or the thoracic or chest cavity (thoracoscopy). They are typically performed far from the target location through small incisions elsewhere in the body. The surgeon's view of the patient is limited to the endoscopic camera view displayed on a monitor. Mixed reality MIS systems are currently being developed for operating theaters, whereby the endoscope camera view is augmented with other information that may not be visible. Haptic feedback on the laparoscopic tools, e.g., the endoscope, may provide the surgeon with additional cues. A processing flow for a typical MIS trainer using the forwarding connections defined in Fig. 2 is:

① Endoscope virtual camera position has changed; virtual endoscopic view on computer monitor is updated and re-rendered.

② Endoscope virtual camera position has changed; Surgeon interprets current view and decides on next step (e.g., insertion, retraction, perform biopsy).

④ Surgeon decides to manipulate endoscope interface moving the endoscope within the virtual patient; Surgeon interprets new view from the endoscopic camera (perhaps in conjunction with medical scan images).

⑤ Surgeon decides to manipulate endoscope interface, to move the endoscope within the virtual patient; virtual endoscopic view on computer monitor is updated and re-rendered.

⑦ A setting changes on the input interface hardware (which is typically fabricated to look and feel like a real endoscopic device); Output interfaces (computer monitor, but could be a head mounted display (HMD); actuator inside input interface hardware provides tactile or force cue) are updated.

► 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 $\mathbb{V}$ encodes the variations of the rendering of the endoscopic view, animation of the virtual patient (e.g., from respiration), and any haptic effect calculated on the virtual endoscope. The human alphabet $\mathbb{H}$ encodes the variations such as the visual attention of the surgeon, any sensation felt on the surgeon's hands, and the decision on how to proceed from an interpretation of the current state. The real environment alphabet $\mathbb{R}$ 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 in such a VE are not only for the surgeon's eye-hand coordination but also for the surgeon's decision mechanism in response to different scenarios. The cost-benefit of using such VEs has already been confirmed by many practitioners.

► 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 effort to increase the space of the real environment alphabet $\mathbb{R}$. 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) with minimal potential distortion will be paramount.

Sports Training. Sporting activities can lend themselves very well to being replicated within a VE. This could be purely for entertainment purposes such as golf and basketball simulators found in arcades, or non-immersive computer games on popular games consoles. In the context of visualization, domain experts in sports are interested in using VEs to provide alternative ways of training a skill, and analysing performance. Miles et al. [57] provide a comprehensive review of the use of VEs for training in ball sports. They identify the key research challenges that are currently being explored, 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 VEs; and do sports skills reliably transfer from VE training conditions to real-world scenarios? The broad coverage of this review and its objective assessment the current successes and challenges can provide our theoretical analysis with necessary evidence in practical applications.

Closely related to the topic of this review, 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 also noted that the use of stereoscopy made no significant difference to the accuracy of depth perception in this simulation.

► 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 about 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. 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, it may also introduce new variables (e.g., fatigue and discomfort, view distortion) that are undesirable to be part of the model. During a training session, a player processes visual stimuli at a very high speed, achieving extremely high rate of alphabet compression. 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. [57] pointed out the importance of performance measure and analysis in VE-based training. 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.

APPENDIX D CAN THE THEORY ANSWER PRACTICAL QUESTIONS?

As part of IEEE VIS 2017, the attendees of the *Workshop on Immersive Analytics: Exploring Future Interaction and Visualization Technologies for Data Analytics* ((http:immersiveanalytics.net). posed a number of questions for discussions during the Workshop. As the discussions on many questions were largely from a practical perspective and often

inconclusive, here we attempt the answers to thirteen of these questions primarily using the information-theoretic metric for measuring the cost-benefit of visualization in VEs. Although the theory cannot fully answer all questions, as demonstrated below, it can help advance the discourse significantly.

Note that the Q8 was missing in the original table in the Google document https:goo.gld5pbRG. We omitted Q12 and Q15 (about designing empirical studies) and Q13 (about existing design methodologies), because they are beyond the scope of this paper. To accommodate different lengths of questions and answers, we reformatted the 16 questions slightly by changing from a table to a list. We also removed the names of those who proposed the questions.

Q1. *Immersion: How immersive is too immersive?*

To formulate an answer to this question, one needs to consider what immersion is and how its quantity is estimated. This paper answers the first question by making use of the existing definitions of the dimensions of VEs (Section 3), and answers the second question by introducing information-theoretic measures to visualization processes in different types of virtual environments (Section 4). The amount of immersion is reflected by the amount of Shannon entropy of the virtual environment and that of the real environment experienced by participants. In addition, we can also measure the amount of Shannon entropy of the data space $\mathbb{Z}_1$ to be visualized and the complexity of visualization tasks $\mathbb{Z}_n$. The paper examines how positive and negative impact of immersion and presence in four categories of VE systems (Section 4) and different levels of visualization (Section 7). One way to consider this question is to rephrase the question as how to optimize the cost-benefit of immersion. The theoretical answer is summarized in Table 1 and Section 7, while in practice we can use the similar discourse in Sections 4 and 6 to analyze a practical application.

Q2. *Walking: During immersive visual exploration, do we walk or do we sit? Do we walk around the data or through the data?*

These two questions can only be answered properly after considering the specific type of data, their possible explicit or metaphoric representations in VEs, and the likely availability of the users' *a priori* knowledge about the data. The dimension *match* is particularly important to the first question. The second question relates to the theoretic discussion about the visual information-seeking mantra in [18, 21]. Information-theoretically, if the user does not have a holistic mental model about the data and such a model is useful for performing the visualization tasks, an initial well-designed walk-around can have a similar effect as an overview, first which is shown to be cost-beneficial [21]. If the user already has a good mental model about the data or such a mental model does not benefit the visualization tasks to be performed, a walk-through the data is likely to have more cost-benefit [18].

Q3. *Abstract Data: Why do we need immersive visualization for non-spatial data? How can we immerse into non-spatial data?*

This question relates to the discussions in Sections 4.1 and 4.3, the first case study in Section 6, and the discussions on analytical visualization and model-developmental visualization for machine-centric models in Section 7.

Q4. *Experiential Analytics: How do we understand and design for this experience? When is it essential and for whom? Are there counter examples where it is unnecessary and slows down the analytical process?*

The first two questions correspond to the discussion on analytical visualization in Section 7. The empirical study by Müller et al. [64] discussed in Section 6 relates to the third question. Clearly much more research effort will be required to answer these three questions.

Q5. *Engagement and Attraction: Immersiveness for engagement (only)? What makes us feel immersed, what do we connect to?*

We believe that this paper has provided detailed answers to the first two questions. Here we assume that the third question means "the connection between a VE and our mind". Section 5 and Appendix A provide a summary answer to this question.

Q6. *Immersive vs 3D: How does immersive analytics differ from 3D data visualization? Non-3D immersive visualization? Most of papers show virtual environments (data visualization) but no data analytics. How we can actually analyze data within immersive environments as we can do in a 2D desktop interface?*

Spatially-3D data visualization can be carried out using immersive and semi-immersive VEs as well as using non-immersive display environments. For the questions about data analytics, see **Q3** and **Q4**.

Q7. *Mapping 3D geospatial datasets into real-world VR environments: how does the quality of the environment impact the understanding of results?*

As discussed in this paper, the impact depends partly on the visualization task, and we can start to examine the impact by first determining which level of visualization the task resides at. For example, the discussions in Sections 4.1 and 7 are particularly relevant to disseminative visualization, while the discussions in Sections 4.2, 4.3, and 7 are relevant to observational visualization.

Q9. *Interaction: Which interaction modalities would you pick? What about mixing modalities of interaction? What are sensible combinations? What can be used to build passive/proactive context and how can that context be used in more explicit/reactive interactions? Which visualization tasks are applicable to immersive analytics?*

These questions are both interesting and challenging. It would require one or a few theoretical papers with profound deliberation as well as many studies in other forms. There has been information-theoretical discourse on interaction in visualization [18], and a few empirical studies have been conducted to measure and estimate the amount of knowledge that humans may convey to visualization processes through interaction [43,99]. This paper provides limited coverage on the interaction modalities mainly because the authors' deliberation on interaction is not quite ready for publication. We hope that future research effort into building a theoretical foundation of visualization will bring comprehensive answers to these questions. Here we introduce some elementary information-theoretic notion to demonstrate the potential applicability of information theory to study interaction in VEs.

Consider all commands that a user may use to interact with a VE system as an alphabet. Each command is thus a letter of the alphabet. With a conventional desktop computer, a user may use a keyboard to type in a command or use a mouse to choose a command from a menu. In these cases, the computer understands the alphabet of commands very well. When a text string or a mouse click does not correspond to any letter of the alphabet, the computer either ignores the input or issues an error message.

In a VE environment, the use of a keyboard or a mouse is typically not as easy as with a desktop. It is not difficult to relate such inconvenience or cumbersomeness with the cost of physical effort and cognitive load. Partly motivated by the need to address such problems and partly by the desire to increase immersion and presence (see Section 3), VE researchers have been developing and experimenting with many interaction modalities. One important technical question is how such interaction modalities may change the alphabet of commands. For example, with gesture recognition, does the alphabet have to include similar gestures for the same commands (e.g., multiple letters corresponding to to the same commands), can any gesture be used for multiple valid actions in a VE (e.g., waving at a friend and issuing a particular command), and will gesture reconstruction incur more potential distortion than recognizing a command issued through a keyboard or mouse?

Q10. *What does it mean to create a visualization in Immersive Analytics?*

In our information-theoretic model of visualization, a visualization process is a series transformation of visualization alphabets, which is part of the other three alphabets (see the paragraphs under the heading of Alphabets and Letters in Section 4). For example, a digital dataset may be represented by a graphical representation in a VE alphabet and

a contextual "dataset" may be a part of a reality alphabet in a mixed-reality environment. The events observed and the decisions made by a user are likely to be part of the human alphabet.

Q11. *Do we really need 3D visualization for 3D data? (related to Q3) What can we perceive/do in 3D immersion that cannot be perceived/done with 2D representations? (related to Q6)*

Here we assume that the term "3D visualization" implies the use of a 3D volumetric display device or a 2D stereo display device. 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 swarming 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 same analysis to compare 3D geometric models displayed as outlines, wireframe, shaded, and photorealistic objects using a plain 2D display.

See also the answers to **Q3**, **Q4**, and **Q6**.

Q14. *Are "classical" definitions (Milgram and Kishino's, and Azuma's) of MR and AR too graphics-centric for data vis? Should we look into more "experience" flavors of MR/AR interpretations?*

We think that the questioner is rightly to suggest the need to accommodate "experience" in formulating concepts in VEs. Because it is difficult to measure experience and knowledge, the cost-benefit metric proposed in [19] avoided direct modeling of experience and knowledge. Instead, a user's observations and decisions are explicitly in the alphabets in a data intelligence workflow, while a user's knowledge is implicitly modeled in the reverse mapping function. We believe that more research effort will be necessary for studying the questions in **Q14**.

Q16. *Does immersion in data differ from immersion in 3D models? If so should we change how we measure it?* Normally 3D models, such as volumetric objects in volume rendering and mesh models in surface rendering, are also considered to be datasets. We suspect that the questioner used the term "data" to imply datasets with fewer than three spatial dimensions. As the questioner must have already observed, for the datasets with fewer than three spatial dimensions, one would often map some non-spatial variables to unused spatial dimensions (e.g., population to height or time to depth). Such a visual mapping is not uncommon in 2D visualization. For example, the y-dimension of a bar chart is commonly used to depict a non-spatial variable. With the aid of other visual variations, more than one non-spatial variable can use the y-dimension, e.g., an error bar on top of a height bar. Regardless of whether using spatial or non-spatial models, 2D or 3D visual representations, and VEs or conventional displays, the user has to perform the reverse mapping from a visual channel to a data variable. This reverse mapping always requires some cognitive load and may cause potential distortion. Hence, the information-theoretic metric for the cost-benefit analysis accommodates both forward and backward mappings in visualization processes, and is ideal for comparing the two types of datasets in VEs. On the one hand, based on the notion of *match* discussed in this paper, some well-designed visual mappings from non-spatial data to spatial dimensions may have a good *match* and demand little cognitive load. On the other hand, some real-world 3D models can be unfamiliar to users, and these datasets can still incur undesired potential distortion and cognitive load. So the question cannot be trivially answered based on spatial or non-spatial data.

Q17. *Defining immersion/immersive. I've heard these hints at a definition: (1) Immersion has to do with the experience. The person using a system is immersed in the process of analysing data. This, I think, relates to being in flow, and blocking out outside disturbances. Is there a difference between feeling immersed and being immersed? This might be thought about as immersed in analysis. (2) Immersion has to do with the technology, putting a focus on AR/VR.This might be thought about as the body being immersed. (3) Immersion has to with being inside/between the data as opposed to looking at it from the outside. This might be thought about as immersed in data. (4) Immersion has to do with being the social context.*

We hope that the questioner may find the definitions about the dimensions of VEs (Section 3) useful basis for improving the definitions proposed in **Q17**.

APPENDIX E FOUR LEVELS OF VISUALIZATION IN VES

Visualization tasks can be categorized into four levels according to the complexity of their search space [19]. In general, the more complex the search space, the more costly the visualization processes are expected to be. The costs typically increase in performing complex tasks because of the demand for more expensive resources, cognitive load, and/or damages due to more errors. In this appendix, we summarize our theoretical findings at each level, while providing our remarks (indicated by ▲) on new technical challenges.

Level 1: Disseminative Visualization. At this level, visualization serves as a presentational aid for disseminating information or insight to others. While the visualization providers do not purposefully search for new information in the data, it is desirable for the participants at the receiving end to gain as much information as possible. For a visualization provider, the complexity of the search space is thus $O(1)$, where $O()$ is the big-O notation in complexity analysis.

VEs can be used to maximize the attention of the participants through several dimensions of immersion and presence (e.g., *inclusion*, *surrounding*, *vividness*, and *sense of believing*). From an information-theoretic perspective, the benefit is achieved primarily through the reduction of potential distortion from the originally intended information, rather than through alphabet compression. (Otherwise, one would choose to deliver the intended information, for instance, through a list of bullet points.) Such VEs have a huge value in education and public engagement. There is a high infrastructural and operational cost to the providers and a high cognitive load to the participants. There must be continuing provision for the former as many VEs in the categories are providing excellent services to knowledge dissemination. The latter is incentivized by the novel experience to be gained by the participants, balanced by the demand for attention in an educational process, and rewarded by the amount of information delivered in the process.

▲ In addition to the financial costs, these VEs continuously face the challenges in delivering technical innovation and novel content. The need for accommodating a large audience is often in conflict with some dimensions of immersion and presence that emphasize the experience of individuals and small groups of participants.

Level 2: Observational Visualization. At this level, visualization enables intuitive and/or speedy observation of captured data. The complexity of the search space is at the level $O(n)$, where n is the number of data objects. It may be useful to note that here we do not include the complexity of analytical thinking in the big $O()$ measurement. For

visualization tasks involving observing a large amount of data, VEs equipped with large high resolution displays can bring more advantages to applications where datasets are less familiar to the users and there are routine requirements for observing such data (e.g., [40]). Such applications demand high-rate alphabet compression and low-rate potential distortion in almost every visualization session. The better utilization of humans' visual search capability and the provision of higher capacity of external memorization can potentially offset the higher costs than the commodity display screens.

▲ Our theoretical analysis suggests that it will be helpful to reduce the cognitive load caused by the frequent switching of attention across a wider field of view. A significant amount of head and body movement for enabling such switching also adds additional addition burden to already-limited working memory. We therefore hypothesize that medium size high-resolution displays may facilitate the reduction of such cognitive load 17. Further studies will be necessary to measure the cognitive loads related to displays of different sizes and different resolutions, and the impact of different levels of fidelity in modelling and rendering (e.g., stereoscopy displays).

For observational visualization tasks to be performed on mixed reality systems, our theoretical analysis confirms the necessity for utilizing parts of reality to reduce the costs and difficulties in capturing, processing, modelling, and rendering many objects in real-time in a real-world environment. Many mixed reality applications feature datasets that are unfamiliar to the users and requirements for rapid transformation from data to visualization, and then to decision making. Hence, the conditions for visualization tasks to benefit from VEs are similar to those for the class of "big data" applications.

▲ We recognize that the dimension of *match* poses a major technical challenge. We have identified the extra cognitive load in visual search due to unfamiliar visual representations and possible poor registration between virtual and real stimuli. We acknowledge that the existing mixed reality research has already made great effort in improving the accuracy of registration. We recommend reducing the cognitive load due to unfamiliar representation through innovative design of more "familiar" visual representations and introducing necessary training in improving the familiarity 10 11 12.

Level 3: Analytical Visualization. At this level, visualization is an investigative aid for examining and understanding complex relationships (e.g., correlation, association, causality, and contradiction). For a visualization task concerning relations involving up to k data objects (referred to as k-relations), a user normally needs to view at least k data objects in order to judge if a k-relation is of any interest. The complexity of the search space for relationships is typically at the level $O(n^k)(k \geq 2)$, where n is the total number of data objects.

While there have been many applications of VEs featuring such analytical visualization tasks, our study of the literature has not revealed any reports that confirm the relative advantages of VEs in supporting such visualization tasks over conventional desktop environments. In general, the more relationships there are to be observed, the more pixels will be required. However, visualizing a large number of connections across a large display would inevitably introduce a fair amount of cognitive load due to more head and body movement in visual search and more burdens on the limited working memory 17.

▲ The lack of concrete evidence does not imply that it is not feasible to use VEs to support analytical visualization. The high cognitive load in VEs does not imply low cognitive load with commodity computers and displays. Once we understand the challenge of the cognitive costs 17, we may be able to develop new visual representations and visualization techniques that can be effectively deployed in VEs. We believe that analytics-aided comparative visualization and visualization-aided causality analysis and predictive analytics are amongst those areas which may yield successful innovation, development, and deployment.

Level 4: Model-developmental Visualization. At this level, visualization is a developmental aid for improving existing models, methods, algorithms and systems, as well as for creating new ones. In this work, we have identified that VE-based training is a form of model-development, though the previous categorization in [19] considered only machine-centric models. The complexity of the search space for models is likely to be at the level of NP (non-deterministic, polynomial). With such a complex search space and without the knowledge of a specific pathway to develop a new model or recondition an existing model in humans' mind, VEs can provide more stimulus information to participants under training and enable them to find an appropriate pathway unthinkingly. The evidence in cognitive science suggests that human-centric models for motor coordination are more complex than most, if not all, current machine-centric models. Our theoretical analysis confirms the cost-benefit of VE-based training, and the evidence from practical applications also supports this finding overwhelmingly.

▲ There are continuing technical challenges to bring more reality into virtuality 19. While we develop new techniques to increase the dimensions of immersion and presence, we must also use model-developmental visualization to aid our understanding of the variables and pathways in the individual human-centric model under training (e.g., [24]) 13. The more understanding we gain, the more effective visualization that we can develop for VE-based training 18.

▲ Meanwhile, the use of VEs for developing machine-centric models is yet to be explored. The successful applications in training human-centric models suggest this potential. In addition to using VEs to control the visual stimuli for a machine-centric model, we can also potentially observe the evolution of a complex model such as a large neural network in a VE.