An Information-theoretic Framework for Visualization

Min Chen, *Member, IEEE*, and Heike Jänicke, *Member, IEEE*

Abstract—In this paper, we examine whether or not information theory can be one of the theoretic frameworks for visualization. We formulate concepts and measurements for qualifying visual information. We illustrate these concepts with examples that manifest the intrinsic and implicit use of information theory in many existing visualization techniques. We outline the broad correlation between visualization and the major applications of information theory, while pointing out the difference in emphasis and some technical gaps. Our study provides compelling evidence that information theory can explain a significant number of phenomena or events in visualization, while no example has been found which is fundamentally in conflict with information theory. We also notice that the emphasis of some traditional applications of information theory, such as data compression or data communication, may not always suit visualization, as the former typically focuses on the efficient throughput of a communication channel, whilst the latter focuses on the effectiveness in aiding the perceptual and cognitive process for data understanding and knowledge discovery. These findings suggest that further theoretic developments are necessary for adopting and adapting information theory for visualization.

Index Terms—Information theory, theory of visualization, quantitative evaluation.

✦

1 INTRODUCTION

Information theory is "the science of quantification, coding and communication of information" [44]. Visualization is concerned with visually coding and communicating information. Many aspects of a visualization pipeline feature events of a probabilistic nature, bearing a striking resemblance to a communication pipeline. For instance,

- data abstraction usually results in data compression;
- creating and viewing a visualization is usually an information discovery process;
- the messages in a visualization are not guaranteed to be received by a viewer;
- the quality of a visualization is often measured by probabilistic experiments; and so forth.

This suggests a strong connection between information theory and visualization. It is a reasonable assumption that the science of visualization should be built upon a number of theories established in different disciplines. It is also rational to consider information theory as one of these theories. This paper presents a theoretic study into a conceptual connection between information theory and visualization.

In the scientific world, a theory is a fact-based framework for explaining a set of observed phenomena or events. To examine the role of information theory in visualization, one can consider the following propositions: (a) information theory can explain all phenomena or events in visualization; (b) information theory can explain some but not all phenomena or events in visualization; (c) information theory can explain none of the phenomena or events in visualization.

For most people, proposition (b) is likely to be an instinctive hypothesis. In order to confirm or disprove (a), (b) or (c), one naive approach to examine exhaustively every phenomenon and event in visualization is to see whether or not it can be explained by any aspect of information theory. Such an approach would clearly be beyond the scope of this paper, if it were not impossible. Moreover, information theory is a scientific subject that is continuously being developed and broadened. Even if we cannot find an explanation of a visualization phenomenon in the current information theory, it does not necessarily prove (a) is false. We thereby adopt an approach to examine the major concepts of information theory through its taxonomy, and for each major concept, we appraise its applicability to visualization. If a concept contradicts with some observations in visualization, we consider this as evidence indicating that information theory cannot explain one particular phenomenon. As the theoretical development of information theory is more mature than that of visualization, it also makes sense to use the taxonomy of information theory as a basis and to view visualization from the perspective of information theory.

In the remainder of this paper, after a brief review of the literature, we examine the similarity and difference between a visualization system and a communication system in Section 3. In Section 4, we juxtapose the two subjects by outlining an information-theoretic taxonomy, and annotate the relevance in visualization. This is followed by a detailed examination of four major concepts of information theory in Sections 5-8. We then briefly discuss the role of user studies under an information theoretic framework in Section 9. We offer our concluding remarks in Section 10.

2 RELATED WORK

It is generally agreed that information theory was founded by Shannon [36, 38] and Wiener [53]. While the development of information theory has been focused on the fundamental limits of data compression and reliable communication [47], it has stimulated a wide spectrum of applications including biology, psychology, linguistics, game theory, and decision theory. Over the last two decades, information theory has been applied extensively to image processing and analysis, including quantization, compression, segmentation, registration, and object detection and recognition (e.g., [21, 28]).

Information theoretic measures have also been used in visualization and computer graphics, including scene and shape complexity analysis by Feixas *et al.* [15] and Rigau *et al.* [33], light source placement by Gumhold [20], view selection in mesh rendering by Vázquez *et al.* [46] and Feixas *et al.* [16], view selection in volume rendering by Bordoloi and Shen [2], and Takahashi and Takeshima [40], focus of attention in volume rendering by Viola *et al.* [48], multi-resolution volume visualization by Wang and Shen [49], feature highlighting in unsteady multi-field visualization by Jänicke and Scheuermann [23, 24], feature highlighting in time-varying volume visualization by Wang *et al.* [50], measuring aesthetics by Rigau *et al.* [34], and transfer function design in volume rendering by Bruckner and Möller [4]. Chen suggested several uses of information theory in visual analytics [7]. These works focused primarily on measuring the information in the data, and some attempted to optimize viewing coordinates using such measurements.

Yang-Peláez and Flowers proposed to use entropy-based measurements for evaluating the effectiveness of information visualization [54], providing metrics for four types of information contents. Part of our work is to consolidate their metrics by removing the assumption of uniform probability distribution for some metrics (see Section 5.1).

- *Min Chen is with Swansea University, E-mail: m.chen@swansea.ac.uk.*
- *Heike Jänicke is with Ruprecht-Karls-University Heidelberg, E-mail: heike.jaenicke@iwr.uni-heidelberg.de.*
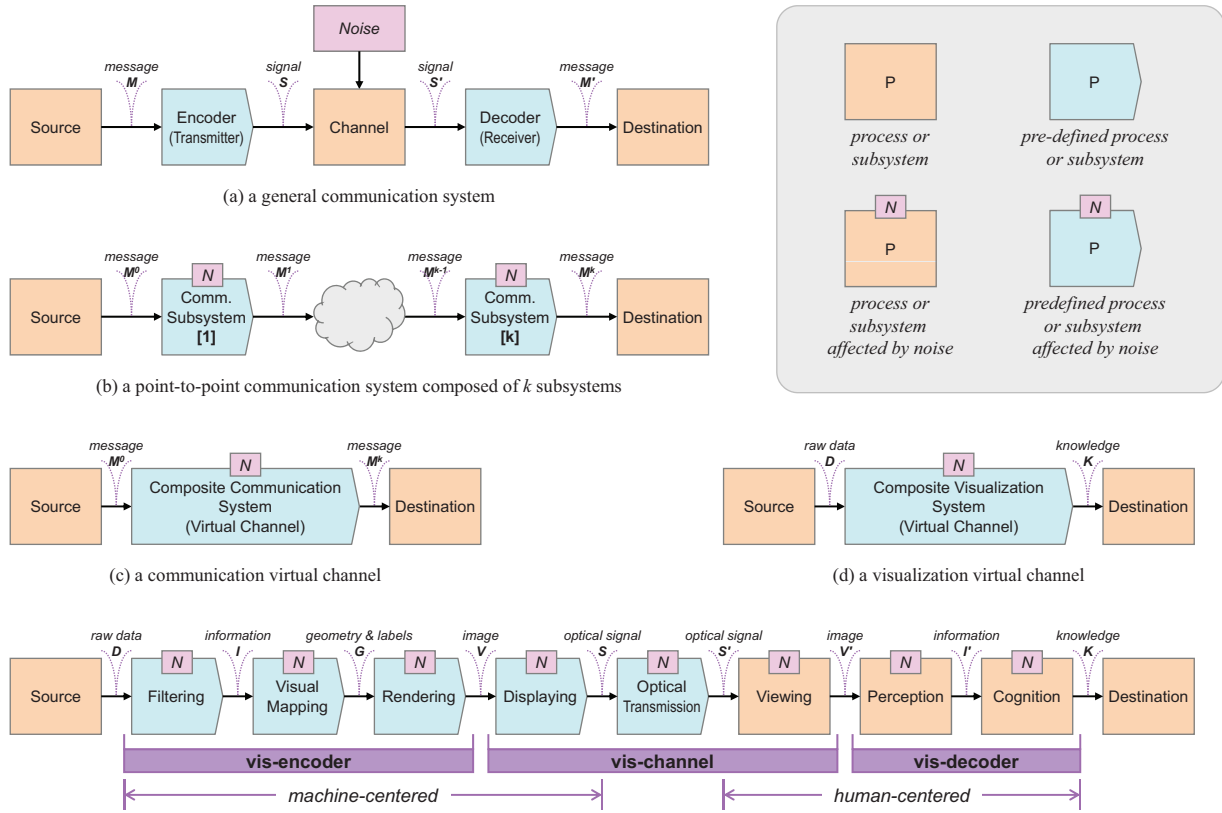
Fig. 1. This series of drawings illustrate the similarity between the models of communication and visualization. When some subsystems in the visualization model (e) are combined to form composite subsystems, vis-encoder, vis-channel, and vis-decoder, the model is similar to the general communication system in (a). When all subsystems, except the source and destination, in the visualization model (e) are amalgamated into a composite system (d), the composite system is similar to a virtual channel in communication (c).

In addition, our work examines the applicability of information theory to a broader spectrum of visualization (e.g., visual design, interaction, redundancy, and user studies).

In theoretic aspects of visualization, there have been significant advances in taxonomies for visualization. Wehrend and Lewis proposed perhaps the first classification based on visualization operations and types of objects and their attributes [52]. During 1990s and early 2000s, we saw a number of taxonomy proposals, including those based on data types by Shneiderman [39], display modes by Keim and Kriegel [26], interaction operations by Chuah and Roth [10], main types of analytical tasks and view manipulation by Buja *et al.* [5], visual analytical tasks by Zhou and Feiner [56] operational states of data by Chi [9], five factors (data, task, skill, context and interaction) by Pfitzner *et al.* [30], and attributes of data models (e.g., continuity, connectivity, dimensionality and variable types) by Tory and Möller [41]. A noticeable contribution of Tory and Möller's work is a common taxonomic framework for both scientific and information visualization.

In addition, there are a number of taxonomical studies on interaction in visualization (e.g., [42, 51, 55]). There are also several taxonomy proposals for specific classes of techniques and applications, e.g., [13, 14]. Brodlie proposed a notation for symbolic labeling visualization methods [3]. Card and Makinlay proposed a descriptive structure for visualization [6]. Duke *et al.* presented an argument to bring taxonomy and ontology together [12].

There are many forms of visualization pipelines. Upson *et al.* provided a generic abstraction of a pipeline with four main components, data source, filtering and mapping, rendering and output [43]. There are many variations, such as [6, 9, 18, 25]. Van Wijk proposed an abstract model that includes perception, cognition and knowledge as part of the visualization model [45]. This article inspired much discussion about the needs for theoretical frameworks for visualization. Liu *et al.* considered distributed cognition as a theoretic framework for informa-

tion visualization [27]. Purchase *et al.* examined three possible theoretic frameworks for information visualization, including data-centric predictive theory, information theory, and visualization process models [31]. Our work builds upon their outline of the need for measuring "information transfer, content, or loss at all stages of the pipeline".

3 MODELS OF COMMUNICATION AND VISUALIZATION

Information theory was first introduced in conjunction with the model of a basic communication system [36]. In this section we examine the similarity and difference between the models of communication and visualization from the perspective of information theory.

Fig. 1(a) shows a typical depiction of a basic communication system considered by Shannon and Weaver [38]. The source and destination of the message can be a person or a machine. The *encoder* (also referred to as transmitter) and *decoder* (also referred to as receiver) transform messages into signals and vice versa. Conceptually, the term "signal" is a generalization encapsulating messages represented by both digital and analog signals. In modern communication systems, we can simply consider both messages and signals as "data". Traditionally, the term "channel" refers to a transmission medium. In abstract, it is a function or process that operates on the input signal S and sometimes adds noise, resulting in the output signal S'.

By grouping the encoder, channel and decoder into a communication subsystem, we can build more complex communication systems. Fig. 1(b) shows a point-to-point communication system composed of k subsystems. In data communication, the term "virtual channel" is often used to denote such a composite communication system, as illustrated in Fig. 1(c). Let $\mathbf{D}$ be the set of all machine-representable data. Any communication system or subsystem is thus a function $F : \mathbf{D} \rightarrow \mathbf{D}$. A communication system may be affected by both internal and external noise. Historically information theory considered mainly the noise present in a channel, as most theoretic and algorithmic discussions as-

Table 1. Important information-theoretic concepts and applications, and their relevance in visualization.

Information Theory Taxonomy	Relevance in Visualization
Fundamental Concepts	**A possible mathematical framework that underpins the subject of visualization.**
Major Quantities and Properties	Quantitative measurements about the data and visualization space, and the relationship between input and output of a process or subsystem at different stages of a visualization pipeline.
Entropy	Measuring information content (see Section 5.1); salience in visualization.
Mutual Information	Uncertainty reduction in visualization (see Section 5.3); information-assisted visualization.
Major Theorems	Many theorems can be used to explain visualization phenomena and events.
Information balance (conservation law)	Given two visualizations, A and B, the amount of information about A contained in B is the same as that about B in A; overview + detail (see Section 5.3); multi-view visualization.
Data processing inequality	After visual mapping, the visualization normally cannot contain more information than the original data (see Section 6.3); information cannot be recovered after being degraded by some processes or subsystems in a visualization pipeline.
Channel Types	Providing a theoretical basis for classifying visualization subsystems (see Section 6).
Noiseless channel	Not common in practical visualization pipelines (see Section 3).
Noisy channel	Most visualization processes and subsystems can be affected by noise (see Section 3).
Channel Capacity	It can be adapted to define the maximum amount of information that can be visualized or displayed (see Section 5.1) but a major extension is necessary when considering channels with memory and interaction (see Section 6).
Redundancy	Efficiency of a visual mapping; error detection and correction (see Section 8).
Source Coding (for Noiseless Channels)	**Inspiration for developing new data abstraction and visual encoding techniques.**
Coding Schemes	Applicable, for example, to the following visualization algorithms:
Entropy coding (e.g., Huffman, arithmetic coding)	Logarithmic plots (see Section 7); importance-based visualization; information-assisted visualization; magic lens; illustrative deformation.
Dictionary-based coding	Legend design; icon design; visualization literacy; knowledge-assisted visualization.
Run-length encoding	Spatial clustering; cluttering reduction.
Differential encoding	Video visualization; time-varying data visualization.
Subband coding	Multi-resolution modeling; transfer function design in volume rendering.
Transform coding	perceptually-based visual encoding, frequency-domain volume rendering.
Quantization	Color mapping; multi-resolution modeling.
Multiplexing	Comparative visualization; volume rendering; multi-field visualization.
Channel Coding (for Noisy Channels)	**Inspiration for developing new fault-tolerant visual encoding techniques**.
Channel capacity of noisy channels	Understanding the limits of visual representations, displays, perception, and cognition.
Hamming Distance	Perceptual sensitivity (e.g., Weber's law); perceptually uniform color spaces.
Rate Distortion Theory	It may provide a theoretic basis for modeling the effects of noise in visualization.
Error Detection coding schemes	Multi-view visualization; uncertainty visualization.
Parity check	
Checksum	
Cyclic redundancy checks	
Error Correction Methods and Schemes	Exploratory visualization; computational steering.
Automatic repeat request	interactive visualization of complex 3D scenes.
Block coding	interactive visualization of volumetric models (see Section 8).
Convolutional coding	
Applications	**Providing tools for developing new visualization systems.**
Statistical Inference	User studies, visual analytics; knowledge-assisted visualization.
Pattern-Recognition	Feature extraction; salience in visualization; data reconstructability.
Game Theory and Decision Theory	Business visualization; user interface design; visual analytics.

sume that encoders and decoders are error-free. In principle, however, noise can also affect encoders and decoders. For example, JPG encoding introduces compression errors.

Fig. 1(e) depicts a general visualization pipeline, without human-computer, human-human or inter-process interaction. Interactive exploration is important in visualization but less so in communication. We will consider interactive visualization later in Sections 6 and 8. Note that almost every process in the pipeline can be affected by noise or errors. For example, the process for data *filtering* may cause information loss or distortion. The *visual mapping* process may introduce quantization errors due to limited bandwidth in geometrical space and attribute space of visual metaphors. The *rendering* process may introduce distortion and ambiguity due to projection, occlusion, and color and opacity mixing. The *displaying* process may introduce errors due to bandwidth limitation and incorrect calibration of a display device. Even the transmission between a display and human eyes may suffer from distance attenuation. The *viewing*, *perception* and *cognition* processes are three human-centered processes, which are much less "mechanical" or "algorithmic" than the earlier machine-centered (or mediated) processes. As Pettersson pointed out, these three processes may be affected by many factors. In terms of noise, it ranges from external distractions to the diversity between individuals [29].

Abstractly we can define a superset Ω as a union of the set of all machine-representable data, information and knowledge, and the set of all data, information and knowledge stored in human brains, despite that we are yet to have an adequate understanding about how information and knowledge are represented in the human brain. This generalization allows us to denote every visualization system or subsystem as a function $G : \Omega \to \Omega$. The similarity between G and the above-mentioned F for the communication system is palpable.

To make the comparison easier, we can group all the processes in Fig. 1(e), except the source and destination, into a single "virtual channel" as depicted in Fig. 1(d), which is juxtaposed with Fig. 1(c). We can also group the pre-defined processes, from *filtering* to *displaying*, into a machine-centered virtual channel, and the three processes, from *viewing* to *cognition* into human-centered processes, resulting in a model similar to that in Fig. 1(a). If we wish to place an emphasis on the visualization images as the intermediate "signals", we can group the eight processes, from *filtering* to *cognition* into three subsystems as in Fig. 1(e), which also results in a model similar to Fig. 1(a). We colloquially refer the three subsystems as *vis-encoder*, *vis-channel*, and *vis-decoder*. In the remainder of the paper, we use the three subsystem model as the main basis for our discussions.

While the models of communication and visualization are unmistakably similar, it is necessarily to recognize the difference between typical communication and visualization systems. For example, a communication system normally assumes that encoders and decoders behave in a deterministic manner. This cannot be said for any human-centered process in a visualization system. In most communication systems, the focus is usually placed on the compact representation of data transmitted through the systems. In most visualization systems, the priority is usually given to the requirements of viewing, perception and cognition, such as intuitiveness and clarity. When translating into the vocabulary of processes, this means that the priority is given to the speed, simplicity and accuracy of the *vis-decoder* subsystem rather than the *vis-encoder* and *vis-channel* subsystems. Despite such difference, we will see in the following sections that information theory can explain many phenomena and events in visualization systems, including the underlying reasons for the different focuses.

4 HOW INFORMATION THEORY RELATES TO VISUALIZATION

There is not much discussion about taxonomies in the literature of information theory, perhaps because it is a relatively mature subject. We thereby compiled a collection of major information theoretic concepts and applications based on a number of textbooks, including [1, 11, 19, 22, 35, 44, 47].

Table 1 lists these concepts and applications in a taxonomical manner (on the left), and provides our annotation as to their possible relevance to visualization (on the right). It is not intended to provide an exhaustive coverage of either information theory or visualization. This would be beyond the scope a single paper. For example, the textbook by Cover and Thomas [11] has some 170 theorems, each of which can potentially be applied to visualization. Nevertheless, these brief annotations suggest that there are strong connections between information theory and visualization in a broad spectrum.

From the table, we can observe that there are many existing visualization techniques that can be related to concepts and algorithms in information theory, especially in relation to source coding (i.e., coding for noiseless channels). With respect to channel coding (i.e., coding for noisy channels), interactive and multi-view visualization offers a means of error detection and correction. Uncertainty visualization may also be very relevant. In general, there has been limited development in visualization, with an explicit intention to deal with the noise in visualization pipelines. This is an area that we hope future visualization techniques can address.

5 QUANTIFYING VISUAL INFORMATION

To keep this article concise, our following technical discussions assume that readers are familiar with the most fundamental concepts of information theory. Detailed explanations of these concepts, and the related mathematical formulae, theorems and proofs, can be found in a number of textbooks (e.g., [1, 11, 19, 44]). We adopt the notation of Cover and Thomas's book [11] for this work. We use the base 2 logarithm wherever appropriate, so *bit* is the unit of the main information theoretic quantities such as entropy $\mathscr{H}$ and mutual information $\mathscr{I}$.

5.1 Entropy

Entropy is a fundamental measure in information theory. There are two common descriptions of entropy $\mathscr{H}(X)$. It can be thought of as the average uncertainty in a random variable X, or the minimal number of bits that are required on average to describe this variable. The random variable X takes a finite number of values $x_1, x_2, \ldots, x_m$, each with a probability $p(x_i)$. p is referred to as a *probability mass function*. Such a random variable can be used to describe the probabilistic attributes of a variety of entities, phenomena and events, such as all instances of data at a sample point, all datasets in a data space, and all visual representations used in an application.

Let us consider a simple, black-white time-series visualization as shown in Fig. 2(a). Assume that the graph plotting area is given as 256×64 pixels. The time series has 64 independent samples, and each sample has an integer value range between 0 and 255. Samples are taken at a regular temporal step. The visualization displays 64 pixels corresponding to the random samples, and the connecting lines between consecutive samples. The probability mass function of each sampling value is independent and identically-distributed, we have $p = 1/2^8$. Let S denote the random variable for a single sample, and Z denote the data space encompassing all time series with 64 samples. Fig. 2(a) shows an instance $z \in Z$. The entropy for this variable Z is calculated as:

$$\mathscr{H}(Z) = \sum_{t=1}^{64} \mathscr{H}(S_t) = -\sum_{t=1}^{64} \sum_{i=0}^{255} \frac{1}{2^8} \log_2 \frac{1}{2^8} = 512. \tag{1}$$

This is very much expected as the time series requires a minimum of 64 bytes (i.e., 512 bits) to encode. In theory, we can also display this binary code as shown in Fig. 2(b), which would require a huge perceptual and cognitive load to decode if not impossible. Fig. 2(b) actually uses more than 1 pixel for each "bit" box, as the figure would otherwise be unreadable by human eyes. If we use only 4×4 pixels per "bit" box, it would result in a total of 2^{13} bits. In practice, we choose to use a graphical (or visual) mapping $G(Z)$ as in Fig. 2(a), for which a plotting area with 256×64 black-white pixels (2^{14} bits) would be a minimal requirement. For this particular design of $G(Z)$, if it uses 2^{14} bits of display space, it is capable of depicting on average 512 bits of information.

If all of the samples take values only in the lower half of the value range, the entropy will be 448 instead of 512. Meanwhile, most users may sensibly halve the plotting area by remapping the y-axis from $[0 - 255]$ to $[0 - 127]$. The proportion of entropy reduction seems to differ
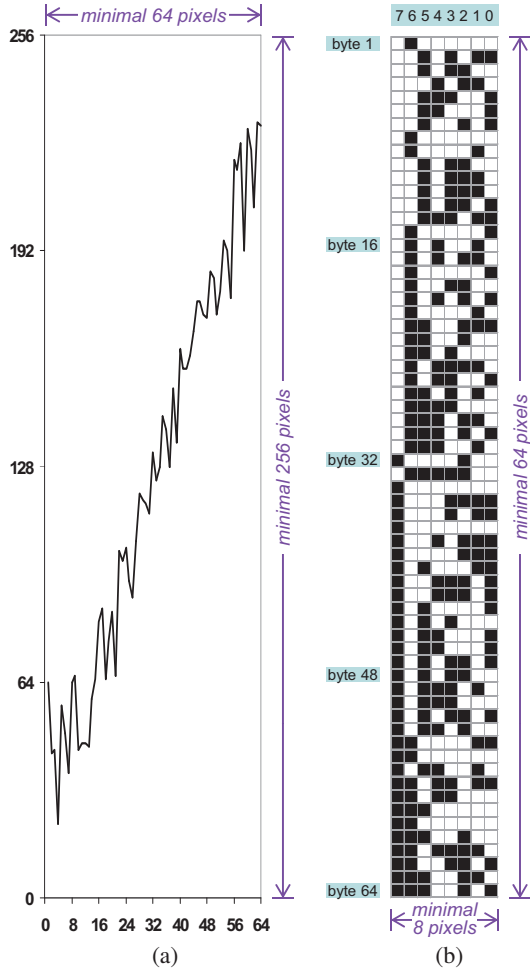
(a)   (b)

Fig. 2. (a) A 64-sample time series with possible integer values ranging between 0 and 255 is plotted as a line graph. A display space with a minimal of 256x64 pixels will be necessary to depict all information possibly contained in the data space. (b) The same time series can in theory be displayed using 64x8 pixels. In practice, more pixels are needed to make individual squares more distinguishable.

from that of space reduction. It is interesting to know how well entropy relates to the visual display. We now introduce two new quantities:

Visualization Capacity. Assume that the data space is sufficiently large for a static graphical mapping G, we define the average amount of information that G can depict as *Visualization Capacity*, $\mathcal{V}(G)$. G is usually constrained by a number of parameters, such as the required display space, the spatial partitioning of the display in relation to the data, use of colors, etc. Once these parameters are fixed, $\mathcal{V}(G)$ is the entropy of a random variable that takes all possible distinguishable outputs of this specific mapping G.

For a specific input data space X, we have $\mathcal{V}(G) = \min(\mathcal{V}(G(X)), \mathcal{H}(X))$. The "min" function indicates that a visualization *normally* cannot display more information than what is contained in the input data space X. This follows the principle of *data processing inequality* in information theory [11]. However, in visualization, it is often advantageous to break the conditions of this principle. We will discuss this further in Section 6.3.

Display Space Capacity. We define the display bandwidth available for visualization as *Display Space Capacity*, $\mathcal{D}$, which takes into account of the number of pixels in the display, and the depth of each pixel. Note that $\mathcal{D}$ is independent of a data space X or a graphical mapping G. It indicates the maximum entropy achievable by any graphical mapping within this display space.

Yang-Peláez and Flowers [54] proposed measurements similar to $\mathcal{H}(X)$, $\mathcal{V}(G)$ and $\mathcal{D}$, but theirs were based on a uniform probability

(a) overview   (b) detailed view

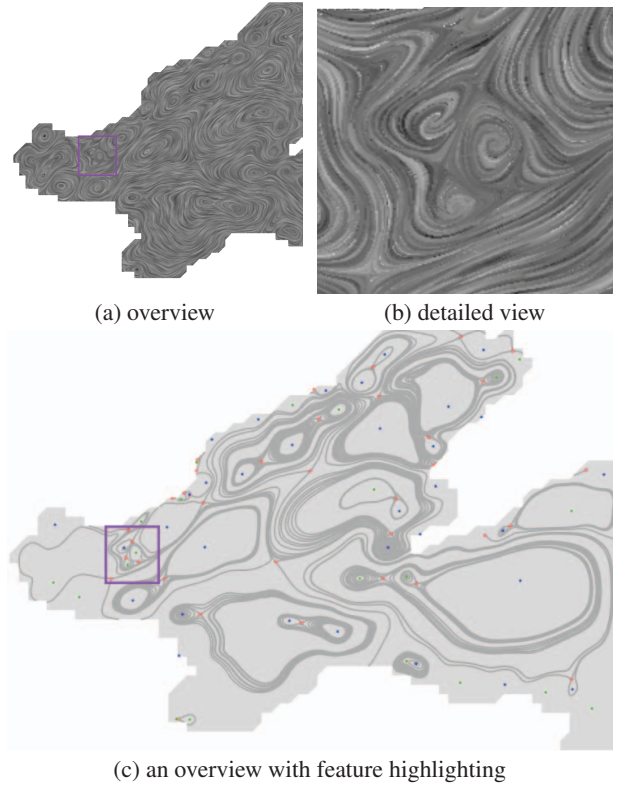(c) an overview with feature highlighting

Fig. 3. An example flow visualization, where the overview in (a) may not show enough information to encourage a user to explore the detailed view in (b). Using feature extraction and highlighting, the new overview in (c) contains more feature-related mutual information about (b).

distribution of X. Here we do not impose this restriction for $\mathcal{H}(X)$ and $\mathcal{V}(G)$, as a priori knowledge about the distribution is usually an essence of a design process in visualization. Later in Section 7, we will show an example how a non-uniform probability mass function leads to a commonly-used visual representation.

The quantities $\mathcal{V}$ and $\mathcal{D}$ have the same unit as entropy $\mathcal{H}$. This allows us to define the following measurement:

$$\text{Visual Mapping Ratio (VMR)} = \frac{\mathcal{V}(G)}{\mathcal{H}(X)}. \qquad (2)$$

$$\text{Information Loss Ratio (ILR)} = \frac{\max(\mathcal{H}(X) - \mathcal{V}(G), 0)}{\mathcal{H}(X)}. \qquad (3)$$

$$\text{Display Space Utilization (DSU)} = \frac{\mathcal{V}(G)}{\mathcal{D}}. \qquad (4)$$

Here we assume $\mathcal{H}(X) > 0$. When $\mathcal{H}(X) = 0$, there is no uncertainty or information in X [11]. Similarly we assume $\mathcal{D} > 0$.

Using the time series visualization in Fig. 2(a) as an example, we have $\mathcal{H}(Z) = 512 = 2^9$, $\mathcal{V}(G) = 2^9$, and $\mathcal{D} = 2^{14}$. Hence, VMR = 1, ILR = 0, and DSU = $2^{-5} = 0.03125$.

For the above example of remapping the y-axis from $[0 - 255]$ to $[0 - 127]$, we can obtain $\mathcal{H}(Z) = 448 = 7 \times 2^6$, $\mathcal{V}(G) = 7 \times 2^6$, and $\mathcal{D} = 2^{13}$. Hence, the entropy and visualization capacity reduce proportionally the same number of bits. Note that VMR, ILR and DSU are measurements of ratios, and thus unitless.

From the perspective of data compression, coding an 8-bit value using 2^8 bits for each sample seems totally insane. Nevertheless, it makes sense for visualization. The human visual system is a highly parallel system. It takes a single viewing step to determine that in Fig. 2(a) has a starting value of 64. It would take 8 viewing steps, together with much cognitive reasoning, to establish this fact from Fig. 2(b). There is thus no reason to be apprehensive about the apparently "inefficiency" of visualization from the perspective of data compression.

This suggests that application of information theory to visualization needs to accommodate and address a different emphasis.

However, there is often not 2^k bits of display space for every k-bit value as we are usually constrained by a limited number of pixels available in most practical applications. In such a situation, the above measurements provide us with a quantitative evaluation of a visual design G. For instance, let us reduce the display space from 64×256 pixel to 64×64 pixels. We denote the new Display Space Capacity as $\mathscr{D}'$, which is 2^{12} bits. The geometry mapping of the original visual design G has to be modified. The new design G' may not be able to depict the full amount of raw data.

Consider a simple data mapping function, M, that maps the above-mentioned time series data space Z to a new data space Z', where each value $j \in Z$ is mapped to $i \in Z'$ such that $i = \lfloor j/4 \rfloor$. We have:

$$\mathscr{V}(G') = \mathscr{H}(Z') = \mathscr{H}(M(Z)) = -\sum_{1}^{64}\sum_{0}^{63} \frac{1}{2^6} \log_2 \frac{1}{2^6} = 384. \quad (5)$$

Hence, $\text{VMR}' = 384/512 = 0.75$, $\text{ILR}' = (512-384)/512 = 0.25$, and $\text{DSU}' = 384/2^{12} \approx 0.094$. Fig. 5(a) shows a visualization of $M(z)$, where $z \in Z$ is the same time series as in Fig. 2(a). In Section 7, we will discuss the relative merits of several visual mappings for a different data space on the same 64×64 display.

5.2 Joint Entropy and Conditional Entropy

In visualization, it is common that many events are inter-related. For example, in interactive exploration, a user may first obtain an overview visualization $G_{overview}$ of a dataset, and then apply a zoom-in operation, resulting in one of the possible detailed views $G_{detail[i]}$. The information contained in $G_{detail[i]}$ is thus related to that in $G_{overview}$.

Let X and Y be two random variables with a joint probability mass function $p(x,y)$. $\mathscr{H}(X,Y)$ denotes the *joint entropy* of the two variables; $\mathscr{H}(Y|X=x)$ denotes the *conditional entropy* of Y given that X is known to be x; and $\mathscr{H}(Y|X)$ denotes the *conditional entropy* of Y for all possible events in X.

Consider that X and Y model the probabilistic attributes of $G_{overview}$ and $G_{detail[i]}$ respectively. Here X and Y represent two visualization spaces, $G_{overview}(Z)$ and $G_{detail[i]}(Z)$, where Z is the input data space shared by both views. We can apply some fundamental theorems in information theory to explain different phenomena in the overview-detail situation. Below are examples of applying two such theorems.

Rule 1. $\mathscr{H}(X,Y) \leq \mathscr{H}(X) + \mathscr{H}(Y)$ [1] — The joint uncertainty of the two views does not exceed the sum of the uncertainty exhibited by each individual view. In other words, having two views can reduce uncertainty. The equality is valid only when X and Y are independent, i.e., $G_{overview}$ and $G_{detail[i]}$ are not related to each other.

Rule 2. $\mathscr{H}(Y|X) \leq \mathscr{H}(Y)$ [19] — In the overview-detail situation, the possible variations of $G_{detail[i]}$ is strongly governed by those of $G_{overview}$. As illustrated in Fig. 3, the distribution and orientation of the visual primitives in the overview determine the overall trend of the distribution and orientation of those in the detailed view. If the event of the square box to be zoomed is known, the entropy of $G_{detail[i]}$ is reduced significantly. After having seen the overview, the viewer has a rough idea about the detailed view $G_{detail[i]}$, and hence is less uncertain about it. Information-theoretically, this means $\mathscr{H}(Y|X=x) < \mathscr{H}(Y)$.

In particular, this rule underpins the design principle for interactive exploration with overview, zoom and detailed views [39].

In situations where $\mathscr{H}(Y|X=x)$ is not as low as the viewer expected, the viewer would be either confused or surprised. For example, if a visualization system did not follow the basic design guideline that a zoom operation should ensure a succeeding view G_{k+1} has the same orientation of the preceding view G_k, the uncertainty about G_{k+1} will be significantly increased. In the case of Fig. 3, most viewers would be confused when encountering a rotated Fig. 3(b) (e.g., by $90°$).

Alternatively, a viewer may be surprised to see some vortices in Fig. 3(b) as there may not be a hint of its existence in Fig. 3(a). In such a situation, the instinctive expectation of a higher conditional entropy is disadvantageous, especially in very large dataset visualization.

Much research effort has been made to extract important features and highlight such features in higher level views (e.g., [23]).

5.3 Mutual Information

Mutual information $\mathscr{I}(X;Y)$ measures the amount of the reduction of uncertainty of one random variable X due to the knowledge of another Y. In information theory, there are a number of fundamental rules about mutual information, which are applicable to visualization events. For example,

Rule 3. $\mathscr{I}(X;Y) = \mathscr{I}(Y;X)$ [11] — This implies that the information about $G_{overview}$ in $G_{detail[i]}$ is the same as that about $G_{detail[i]}$ in $G_{overview}$. Undoubtedly, in most cases, the information about $G_{detail[i]}$ resides primarily in the corresponding window in $G_{overview}$. Let us partition $G_{overview}$ into n disjoint windows, each corresponding to a $G_{detail[k]}, k = 1, 2, \ldots, n$. We have:

$$\sum_{k=1}^{n} \mathscr{I}(G_{detail[k]}; G_{overview}) = \sum_{k=1}^{n} \mathscr{I}(G_{overview}; G_{detail[k]}) \quad (6)$$

The left-hand side represents the total mutual information about all detailed views in an overview, while the right-hand represents the total mutual information about the overview in all n detailed views. The former corresponds to one viewing step, but the latter n viewing steps. This confirms the principle of overview first and details after [39].

Mutual information can also be used to quantify the effectiveness of a type of visualization. Consider a simplified example, where a viewer makes a decision to zoom-in about the box in Fig. 3(a) based on whether there is a hint of vortices in the box. Let A be a random variable with two possible states, "show hints" and "show no hint" in that box. Similarly, let B be a variable about the detailed view of the box, with two possible states, "show at least one vortex" and "show no vortex". Table 2 shows an example joint probability mass function $p(a,b)$ in columns 3 and 4. We obtain $\mathscr{I}(B;A) \approx 0.147$, indicating the amount of uncertainty of Fig. 3(b) that can possibly be reduced by having visualized Fig. 3(a). Suppose that we introduce a feature highlighting technique to improve Fig. 3(a), as shown in Fig. 3(c). C is a random variable similar to A but corresponds to Fig. 3(c). The new probability mass function $p(c,b)$ is given in columns 6 and 7 of Table 2. The technique results in a 40% probability (instead of 25% previously) of showing a hint of any vortex in the same box area. Although the false positive has increased from 5% to 10%, the mutual information $\mathscr{I}(B;C)$ has risen to about 0.278. In other words, Fig. 3(c) can now tell more about Fig. 3(b) in terms of vortices.

Table 2. Two example joint probability mass functions $p(a,b)$ and $p(c,b)$.

	$p(b)$ ▼	$p(a,b)$ hint	$p(a,b)$ no hint	$p(b)$ ▼	$p(c,b)$ hint	$p(c,b)$ no hint
vortex	0.5	0.25	0.25	0.5	0.4	0.1
no vortex	0.5	0.05	0.45	0.5	0.1	0.4
	$p(a)$ ▶	0.3	0.7	$p(c)$ ▶	0.5	0.5

6 INFORMATION SOURCES AND COMMUNICATION CHANNELS

In information theory and its application of data communication, the terms of *information sources* and *communication channels* are formally defined and categorized. In this section, we propose our adaptation of these concepts from the perspective of visualization.

6.1 Information Sources

An *information source* is a process that produces a message or a sequence of messages to be communicated. A source is said to be "memoryless" if each message is an independent random variable and obeys the same probability distribution stochastically as the other messages. A source is said to be "stationary" if its probability distribution is spatially and temporally invariant. In information theory, these two properties are the preconditions of many theorems [19]. The preconditions are also commonly assumed by most communication systems and compression algorithms.

(a) noisy channel   (b) lossless channel   (c) deterministic channel   (d) noiseless channel   (e) useless channel
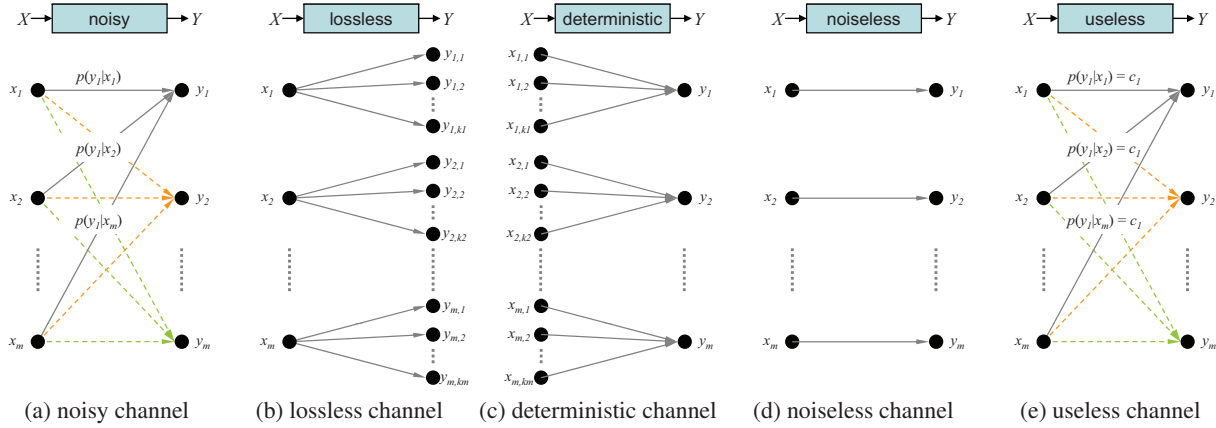
Fig. 4. Five common types of communication channels.

A visualization system may encounter three types of information sources, namely *input data*, *interaction*, and *pre-stored knowledge*. If we focus only on the input data, as in Section 3, the parallel between visualization and communication is apparent (Fig. 1). In most cases, assuming that the input is a memoryless and stationary source is a sensible abstraction for both theoretic and algorithmic development.

Interaction allows users to provide a visualization system with additional information, such as viewing parameters and mapping functions, resulting in different output data.

Pre-stored knowledge includes hard-coded knowledge (e.g., feature recognition) in vis-encoder, and human knowledge in vis-decoder (Fig. 1(e)). The former normally is not of a stochastic nature, but this may change in future systems with the introduction of knowledge-assisted visualization [8]. The latter is stochastic, especially when considering the whole population of potential users of a system.

There is still a major scientific gap in understanding the probabilistic properties of human interaction and human knowledge. Many existing theorems in information theory may not be readily applicable when such information sources are considered, and some adaptation and extension of information theory will be necessary.

6.2 Channels

A *channel* is the medium over which coded messages are transmitted from the encoder to the decoder. Here, we consider only discrete channels. In communication, unintended changes could be made to a coded message, resulting in errors in transmission. Such changes are referred to as *noise* or *perturbation*. A channel may have properties such as bandwidth, transmitted power and error rate.

Let X and Y denote the random variables corresponding respectively to the input and output of a channel. Fig. 4(a) shows a typical noisy channel, where an input message x_i may be received as y_j with a probability mass function $p(y_j|x_i)$.

In visualization, X and Y typically represent various input and output data spaces of each subsystem in Fig. 1(e), while x_i and y_i represent instances in each space. For example, considering a vis-encoder subsystem for the time series visualization in Fig. 2(a), we have $X = Z$ for the space of raw data and $Y = G(Z)$ for the space of visualization images. When we study a particular algorithmic component of a subsystem, e.g., a filtering or clustering function, X and Y can also represent the input and output spaces related only to this component.

Many processes in a visualization pipeline are noisy channels. In a vis-decoder subsystem, a graphical object depicted in a visualization can easily result in different interpretations by different viewers. Using the notation in Fig. 4(a), it means that an instance x_1, for example, may be probabilistically interpreted as different instances $y_i \in Y$, with different $p(y_i|x_1), i = 1, 2, \ldots, m$. When x_1 is intended to be seen as a specific y_k, we would like to maximize $p(y_k|x_1)$. This can be achieved by a better design of the visual mapping, or by helping the viewers to detect errors.

A channel is said to be "lossless" if every input message can be uniquely determined from an output message as illustrated in Fig. 4(b). Though it is a one-to-many mapping for each x_i, the mapped messages $\{y_{i,1}, y_{i,2}, \ldots, y_{i,k_i}\}$ are grouped into a set corresponding to x_i uniquely. Such a channel has a conditional entropy $\mathcal{H}(X|Y) = 0$ for all input distribution. The "lossless" channel, which introduces *redundancy*, provides a mean for error detection and correction in communication. We will examine this in Section 8.

A channel is said to be "deterministic" if every input message uniquely determines an output message, as shown in Fig. 4(c). Such a channel has a conditional entropy $\mathcal{H}(Y|X) = 0$ for all input distribution, and can facilitate abstraction. For example, quantization at different stages of the pipeline, such as color mapping, or the range mapping M in Section 5.1, behaves as a "deterministic" channel.

A channel is said to be "noiseless" if it is lossless and deterministic, resulting in a one-to-one mapping between input and output messages as illustrated in Fig. 4(d). In visualization, such a channel is only desirable when the input data space is small. For large scale data visualization, a totally noiseless visualization system may be neither practical nor helpful.

In visualization, an abstraction process often does not throw the original data away, and the abstraction is merely used to support visual mapping, e.g., assigning colors to data points in different clusters. In other words, such a process is a combination of "deterministic" and "noiseless" channels. This mechanism is commonly used in supporting visual categorization, and focus of attention.

A channel is said to be "useless" if every output message has an equal chance of resulting from any input message, as shown in Fig. 4(e). In terms of entropy, we have $\mathcal{H}(X|Y) = \mathcal{H}(X)$.

A discrete channel is said to be "memoryless" if the probability distribution of the output depends only on the input at that time, and is independent of previous inputs and outputs of the channel [11]. The channels in vis-decoder are certainly not memoryless, while interaction introduces historical dependency. This hinders the direct application of some major theorems in information theory, e.g., Shannon's channel coding theorem [36]. Nevertheless, the basic idea for handling errors in noisy channels is very much applicable to visualization.

6.3 On Data Processing Inequality

It is necessary to note that the actual size of a data set may increase after calling a visualization process. However, in principle, the visualization process does not generate more information than what is in the original data space. In communication, it is referred to as *data processing inequality* [11], which states: if random variables X, Y, Z form a Markov chain in the order of $X \rightarrow Y \rightarrow Z$, then we have the following inequality between their mutual information:

$$\mathcal{I}(X;Y) \geq \mathcal{I}(Y;Z) \qquad (7)$$

X, Y, Z are said to form a Markov chain if Z depends only on Y and is conditionally independent of X.
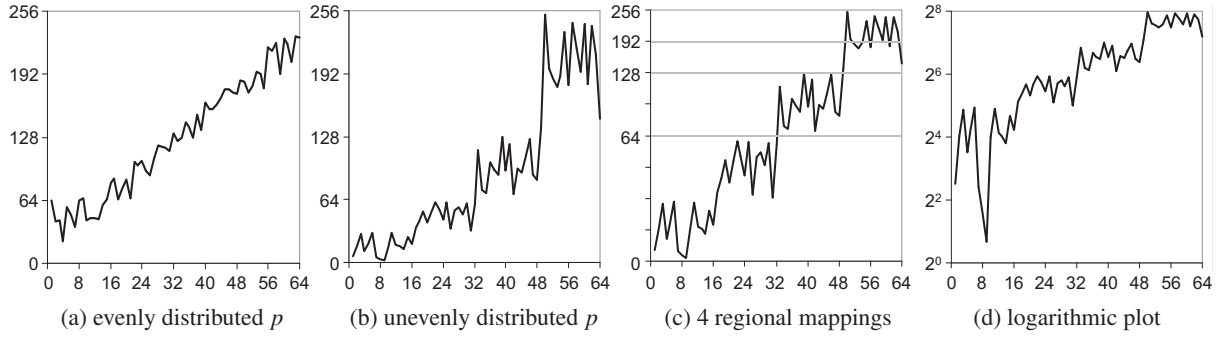
Fig. 5. (a) Reduced display may result in information loss. When displaying the same time series in Fig. 2(a) using only 64x64 pixels, the *visualization capacity* $\mathscr{V}$ is reduced from 512 to 384 bits, resulting in 25% information loss. (b) For a data space with a non-uniform probability mass function p, the *visualization capacity* $\mathscr{V}$ is further reduced to 368, while the entropy of the input data is now 496. (c) Using 4 regional spatial mappings, one can improve $\mathscr{V}$ and reduce the information loss. However, though this is entropy-based, the transitions between the 4 data ranges are not continuous. (d) One can replace (c) with a logarithmic plot to maintain the continuous spatial transition.

If we only allow data input as the information source for the channels in the chain, this inequality stands. However, this principle should not be naively applied to all visualization processes, because the pipeline is mostly not a Markov chain.

Firstly, many processes in visualization are interactive processes, so we cannot guarantee that Z solely depends on Y. Secondly, even if there is no interaction, we usually take into account our knowledge about the raw data (e.g., X), when we design an algorithm at a late stage of the chain (e.g., for $Y \to Z$). Z is not conditionally independent of X. The condition for a Markov chain is thus broken.

In fact, we should not be disappointed by the fact that the data processing inequality is not as ubiquitous in visualization as one would expect. This fact only implies that interaction and domain knowledge about the raw data are critical in breaking the bottleneck of "data processing inequality". This explains why most visualization systems are interactive systems, and supports the argument for knowledge-assisted visualization [8].

7 CODING IN NOISELESS CHANNELS

In data communication, coding schemes are broadly divided into two main categories, namely *source coders* and *channel coders*. A source coder focuses on the messages from the source, and tries to find the most compact representation of the messages. A channel coder focuses on the noise on the channel, and tries to find a cost-effective representation that can help detect or/and correct errors introduced by the channel. We study these two categories in the context of visualization in this section and Section 8 respectively.

As we can see from Table 1, there is a good collection of source coding schemes designed for noiseless channels. Many of these schemes correlate to some visualization techniques. Most importantly, they can inspire us to develop new data abstraction and visual encoding techniques. Here we give one example of such schemes to illustrate the relevance of source coding to visualization. We call this scheme *Entropy-based Spatial Mapping*.

Recall the time series example in Fig. 5(a). The original data space is Z, where $Z = \{z_{i,j} | i = 1, 2, \ldots, 64; j = 0, 1, \ldots, 255\}$. Consider a different data space W. Its value range $[0, 255]$ is divided into 2^d equal-sized sub-ranges, $R_1, R_2, \ldots, R_{2^d}$, where d may take an integer value between 0 and 8. Each sub-range thus has 2^{8-d} possible data values. Fig. 5(a) is an instance when $d = 0$, after a linear mapping of the value range from $[0, 255]$ to $[0, 63]$. When $d > 0$, the probability mass function, $p(w_{i,j})$ varies according to the sub-ranges. Suppose that there is a $1/2^k$ chance that the sample values will fall into sub-range $R_k, k = 1, \ldots, 2^d - 1$. The chance for sub-range R_{2^d} is the remainder of probability, i.e., R_{2^d} and R_{2^d-1} have the same probability.

For example, when $d = 2$, we have four sub-ranges, with probabilities, $1/2$, $1/4$, $1/8$ and $1/8$ respectively. Fig. 5(b) shows the visualization of such a time series, which uses the same linear mapping from $[0, 255]$ to $[0, 63]$ as in Fig. 5(a).

Table 3. Quantities and measures of different number of sub-ranges. When the number is $2^d, d > 1$, the probability mass function varies logarithmically in different sub-ranges.

# sub-ranges	1, 2	4	8	16	32
entropy $\mathscr{H}$	512	496	447	384	320
linear $\mathscr{V}(G_l)$	384	368	319	256	192
linear VRM_l	0.750	0.742	0.714	0.667	0.600
linear ILR_l	0.250	0.258	0.286	0.333	0.400
linear DSU_l	0.094	0.090	0.078	0.062	0.047
non-linear $\mathscr{V}(G_{nl})$	-	384	384	384	384
non-linear VRM_{nl}	-	0.774	0.859	1.000	1.200
non-linear ILR_{nl}	-	0.226	0.141	0.000	0.000
non-linear DSU_{nl}	-	0.094	0.094	0.094	0.094

Assume that $p(w_{i,j})$ within each sub-range is independent and identically-distributed. We thus have $p(w_{i,j}) = 1/2^7$ in R_1, $1/2^8$ in R_2, and $1/2^9$ in R_3 and R_4. The data space entropy $\mathscr{H}(W)$ is the sum of entropies of the four sub-ranges. The *visualization capacity* for the linearly mapped data is the sum of those of the sub-ranges. We have $\mathscr{H}(W) = 496$ and $\mathscr{V}(G_l(W)) = 368$.

Hence, we have visual mapping ratio $\text{VMR}_l = 368/496 \approx 0.742$, information loss ratio $\text{ILR}_l = (496 - 368)/496 \approx 0.258$, and display space utilization $\text{DSU}_l = 368/2^{12} \approx 0.09$. In comparison with a uniform distribution (Section 5.1), the *visualization capacity* $\mathscr{V}$ is slightly reduced, while there is slightly more information loss, poorer utilization of display space.

Let us consider a non-linear mapping function, which maps $R_1 : [0, 63]$ to $[0, 31]$, $R_2 : [64, 127]$ to $[32, 47]$, $R_3 : [128, 191]$ to $[48, 55]$, and $R_4 : [192, 255]$ to $[56, 63]$. It is not difficult to derive that $\mathscr{V}(G_{nl}(W)) = 384$ with $\text{VMR}_{nl} \approx 0.774$, $\text{ILR}_{nl} \approx 0.226$, and $\text{DSU}_{nl} \approx 0.094$. We have slightly improved the *visualization capacity*, and as well reduced information loss. When we plot out G_{nl} in Fig. 5(c), this is conceptually similar to a logarithmic mapping in Fig. 5(d).

If we increase the number of sub-ranges, for instance, for $d = 3, 4, 5$, the improvement becomes more significant and interesting as shown in Table 3. For $d = 3, 4, 5$, the six lower data ranges are mapped to six visualization ranges with 32, 16, 8, 4, 2, 1 pixels respectively. The remaining high-value data ranges are mapped to a single 1-pixel range. The non-linear mapping manages to maintain the *visualization capacity* at 384. The higher d is, the closer the distribution is to a logarithmic distribution. In other words, logarithmic plots are in effect a means to increase *visualization capacity* $\mathscr{V}$ when the distribution of the sample values follows a certain logarithmic pattern. This explains why logarithmic plots are commonly used in sciences and engineering.

Conceptually entropy-based spatial mapping bears a strong resemblance to *entropy coding* in data compression and communication. The latter is a family of coding schemes, which replace fixed-length codewords with variable-length codewords. The well-known entropy

Compositing along the ray causes loss of depth information

A viewer corrects errors in depth perception using the animation

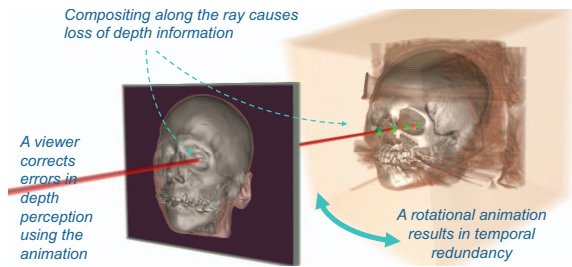*A rotational animation results in temporal redundancy*

Fig. 6. Motion parallax [17], facilitated by interaction and real-time rendering, is a form of error correction in visualization.

encoding schemes include Huffman encoding, Shannon-Weaver-Fano encoding and arithmetic encoding.

8 CODING IN NOISY CHANNELS

In visualization, *interaction* is the primary means for helping a viewer to detect and correct errors. For example, in medical visualization, a volumetric model (e.g., a CT scan dataset) is often displayed using direct volume rendering. In a resulting visualization, as illustrated in Fig. 6, samples at different depth along the same ray are projected onto the same pixel, and the colors and opacities of these samples are combined according to the volume rendering integral, resulting in a pixel color capturing information from many samples.

The process itself is very similar to frequency-division multiplexing in communication, where different signals are transmitted in several distinct frequency ranges over that same medium. However, in the case of volume visualization, we cannot assure the distinctive separation between different frequency ranges, though a good transfer function may provide more visual cues to alleviate the problem. Furthermore, there is a substantial loss of 3D information in a 2D projection. In other words, viewers are expected to make mistakes in determining the shapes depicted in such a visualization.

Typically, a viewer interactively rotates the volume, receiving multiple visualizations for the same model. From these interactively generated visualizations, the imprecise models perceived initially gradually converge to a more accurate 3D model in the viewer's mind. Conceptually, this is similar to "backward error correction" (or automatic repeat request) in communication, which requests for retransmission of erroneous data. The changes of viewing positions spread errors across different parts of the model, making error detection and correction possible for each individual part. This is conceptually very similar to multidimensional parity-check coding, which is a type of block-based error correction schemes.

Multi-view visualization is another means for error detection and correction, especially in visualizing non-spatial data, where errors are often due to the perceptual load of visual search, change detection and attention. Multi-view visualization allows the same information to be found in different views, increasing the probability of locating the information. This is very similar to repetition coding schemes in data communication.

This naturally leads to the issue of *redundancy*. All error detection or correction coding schemes cause redundancy. Rheingans and Landreth studied the benefits of redundancy in visualization [32]. They found that (i) different parameters of a visual mapping convey different types of information with different efficiency, (ii) multiple display parameters can overcome visual deficiencies; (iii) multiple display parameters reinforce each other. Information theory can provide support to their conclusions.

Considering that the *vis-decoder* part of the pipeline in Fig. 1 is highly noisy, it is a great challenge to design visual mappings with built-in error detection and correction mechanisms.

9 THE ROLE OF USER STUDIES

In previous sections, we have shown that information theory provides a quantitative measure about information in a visualization context. We do not however suggest that such quantitative measurements might replace user studies. On the contrary, the above discussions have naturally led to the question as to what is the role of user studies from the perspective of information theory, since users studies produce statistics about phenomena and events in visualization.

In an information-theoretic framework, user studies have a much bigger role than what is in the state of the art of visualization. A fundamental component of any information-theoretic measure is the *probability mass function*. Perhaps we may estimate such a function for an input variable X based on our domain knowledge about the application concerned, or we may obtain this by placing a data flow monitor in the *vis-encoder* part of the visualization pipeline (Fig. 1(e)). However, we simply do not have sufficient knowledge about human perception and cognition in order to estimate such a function yet.

Recall the overview+detail example in Section 5.3, the two joint probability mass functions in Table 2 are synthetic data to demonstrate a mathematical concept. However, such data can be collected through user studies, and to a certain extent, may also be collected seamlessly through users' interaction with the system. The challenges will be our understanding of what probabilistic attributes are fundamental and generic in visualization, so we can estimate a finite set of probability mass functions to be used in practical applications of information theory. For example, in language processing, we have statistics about probability of the appearance of each English letter, the conditional probability about one letter after another, the redundancy in printed English, and so on. Such statistically estimated probability mass functions have been used effectively in applications such as data compression and hand-writing recognition. If we have such fundamental statistical findings from visualization user studies, we can transfer information theory to practice in visualization.

10 CONCLUSIONS

In this paper, we have reported our theoretic findings on whether information theory can become one of the theoretic frameworks of visualization. Our contributions are:

- We have presented an information theoretic view of a visualization pipeline (Fig. 1).

- We have examined information theory and its major applications taxonomically, and established connections between information theory and visualization in a broad spectrum (Table 1).

- We have applied the information-theoretic measures to several aspects of visualization, and we have consolidated and extended the measures proposed in [54]. We have used examples to show that these quantities can explain some phenomena and events in visualization (e.g., visual mapping, and overview+detail).

- We have shown that the major concepts of information theory, ranging from information sources to channel coding theory, and from data processing inequality to redundancy, are all relevant to visualization. In some cases (e.g., channel coding theory), adaptation and extension are necessary. In other cases (e.g., data processing inequality), visualization exhibits features that are not commonly seen in data compression and communication.

- We have also studied the parallel of source coding and channel coding in the context of visualization, and the role of user studies in the framework, suggesting challenges in dealing with noisy channels in visualization, and in organizing user studies from an information theoretic perspective.

Based on these, we can state that information theory can explain a very large collection of phenomena and events in visualization. It can provide visualization with a theoretic framework, underpinning many aspects of visualization, including but not limited to visual mapping, interaction, user studies, quality metrics, and knowledge-assisted visualization. Having a theoretic framework is only a start. For future work, we quote Shannon's words, it will be a "slow tedious process of hypothesis and experimental verification" [37].

REFERENCES

[1] R. Ash. *Information Theory*. John Wiley & Sons, 1965.

[2] U. Bordoloi and H.-W. Shen. View selection for volume rendering. In *Proc. IEEE Visualization*, pages 487–494, 2005.

[3] K. Brodlie. A classification scheme for scientific visualization. In R. A. Earnshaw and D. Watson, editors, *Animation and Scientific Visualization*, pages 125–140. Academic Press, 1993.

[4] S. Bruckner and T. Möller. Isosurface similarity maps. *Computer Graphics Forum*, 29(3):773–782, 2010.

[5] A. Buja, D. Cook, and D. F. Swayne. Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics*, 5:78–99, 1996.

[6] S. Card and J. Mackinlay. The structure of the information visualization design space. In *Proc. IEEE Information Visualization*, pages 92–99, 1997.

[7] C. Chen. An information-theoretic view of visual analytics. *IEEE Computer Graphics and Applications*, 28(1):18–23, 2008.

[8] M. Chen, D. Ebert, H. Hagen, R. S. Laramee, R. van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver. Data, information and knowledge in visualization. *IEEE Computer Graphics and Applications*, 29(1):12–19, 2009.

[9] E. H. Chi. A taxonomy of visualization techniques using the data state reference model. In *Proc. IEEE Information Visualization*, pages 69–75, 2000.

[10] M. C. Chuah and S. F. Roth. On the semantics of interactive visualizations. In *Proc. IEEE Information Visualization*, pages 29–36, 1996.

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

[12] D. J. Duke, K. W. Brodlie, D. A. Duce, and I. Herman. Do you see what I mean? *IEEE Computer Graphics and Applications*, 25(3):6–9, 2005.

[13] G. Ellis and A. Dix. A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1216–1223, 2007.

[14] N. Elmqvist and P. Tsigas. A taxonomy of 3d occlusion management for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(5):1095–1109, 2008.

[15] M. Feixas, E. del Acebo, P. Bekaert, and M. Sbert. An information theory framework for the analysis of scene complexity. *Computer Graphics Forum*, 18(3):95–106, 1999.

[16] M. Feixas, M. Sbert, and F. González. A unified information-theoretic framework for viewpoint selection and mesh saliency. *ACM Transactions on Applied Perception*, 6(1):1–23, 2009.

[17] S. H. Ferris. Motion parallax and absolute distance. *Journal of Experimental Psychology*, 95(2):258–63, 1972.

[18] D. P. Groth and K. Streefkerk. Provenance and annotation for visual exploration systems. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1500–1510, 2006.

[19] S. Guiaşu. *Information Theory with Applications*. McGraw-Hill, 1977.

[20] S. Gumhold. Maximum entropy light source placement. In *Proc. IEEE Visualization*, pages 275–282, 2002.

[21] M. Gurban and J.-P. Thiran. Information theoretic feature extraction for audio-visual speech recognition. *IEEE Transactions on Signal Processing*, 57(12):4765–4776, 2009.

[22] D. Hankerson, G. A. Harris, and P. D. Johnson, Jr. *Introduction to Information Theory and Data Compression*. CRC Press, 2003.

[23] H. Jänicke and G. Scheuermann. Visual analysis of flow features using information theory. *IEEE Computer Graphics and Applications*, 30(1):40–49, 2010.

[24] H. Jänicke, A. Wiebel, G. Scheuermann, and W. Kollmann. Multifield visualization using local statistical complexity. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1384–1391, 2007.

[25] T. Jankun-Kelly, K.-L. Ma, and M. Gertz. A model and framework for visualization exploration. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):357–369, 2007.

[26] D. A. Keim and H.-P. Kriegel. Visualization techniques for mining large databases: A comparison. *IEEE Transactions on Knowledge and Data Engineering*, 8(16):923–936, 1996.

[27] Z. Liu, N. Nersessian, and J. Stasko. Distributed cognition as a theoretical framework for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1173–1180, 2008.

[28] J. A. O'Sullivan, R. E. Blahut, and D. L. Snyder. Information-theoretic image formation. *IEEE Transactions on Information Theory*, 44(6):2094–2123, 1998.

[29] R. Pettersson. *Visual Information*. Educational Technology Publications, 1993.

[30] D. Pfitzner, V. Hobbs, and D. Powers. A unified taxonomic framework for information visualization. In *Proc. Asia-Pacific Symposium on Information Visualisation*, pages 57–66, 2003.

[31] H. C. Purchase, N. Andrienko, T. Jankun-Kelly, and M. Ward. Theoretical foundations of information visualization. In A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization: Human-Centered Issues and Perspectives*, Springer LNCS 4950, page 4664, 2008.

[32] P. Rheingans and C. Landreth. Perceptual principles for effective visualizations. In G. Grinstein and H. Levkowitz, editors, *Perceptual Issues in Visualization*, pages 59–74. Springer-Verlag, 1995.

[33] J. Rigau, M. Feixas, and M. Sbert. Shape complexity based on mutual information. In *Proc. IEEE Shape Modeling and Applications*, 2005.

[34] J. Rigau, M. Feixas, and M. Sbert. Informational aesthetics measures. *IEEE Computer Graphics and Applications*, 28(2):24–34, 2008.

[35] K. Sayood. *Introduction to Data Compression*. Morgan Kaufman, 1996.

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

[37] C. E. Shannon. The bandwagon. *IRE Transactions on Information Theory*, 2(1):3, 1956.

[38] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.

[39] 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.

[40] S. Takahashi and Y. Takeshima. A feature-driven approach to locating optimal viewpoints for volume visualization. In *Proc. IEEE Visualization*, pages 495–502, 2005.

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

[42] L. Tweedie. Characterizing interactive externalizations. In *Proc. ACM CHI*, pages 375–382, 1997.

[43] C. Upson, T. Faulhaber, Jr., D. Kamins, D. H. Laidlaw, D. Schlegel, J. Vroom, R. Gurwitz, and A. van Dam. The application visualization system: A computational environment for scientific visualization. *IEEE Computer Graphics and Applications*, 9(4):30–42, 1989.

[44] M. J. Usher. *Information Theory for Information Technologists*. MacMillan, 1984.

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

[46] P.-P. Vázquez, M. Feixas, M. Sbert, and W. Heidrich. Automatic view selection using viewpoint entropy and its application to image-based modelling. *Computer Graphics Forum*, 22(4):689–700, 2004.

[47] S. Verdu and S. W. McLaughlin, editors. *Information Theory: 50 Years of Discovery*. IEEE Press, 1999.

[48] I. Viola, M. Feixas, M. Sbert, and M. E. Gröller. Importance-driven focus of attention. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):933–940, 2006.

[49] C. Wang and H.-W. Shen. LOD Map - a visual interface for navigating multiresolution volume visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1029–1036, 2005.

[50] C. Wang, H. Yu, and K.-L. Ma. Importance-driven time-varying data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1547–1554, 2008.

[51] M. O. Ward and J. Yang. Interaction spaces in data and information visualization. In *Proc. Eurographics/IEEE TCVG Symposium on Visualization*, pages 137–145, 2004.

[52] S. Wehrend and C. Lewis. A problem-oriented classification of visualization techniques. In *Proc. IEEE Visualization*, pages 139–143, 1990.

[53] N. Wiener. *Cybernetics*. John Wiley & Sons, 1948.

[54] J. Yang-Peláez and W. C. Flowers. Information content measures of visual displays. In *Proc. IEEE Information Vizualization*, pages 99–103, 2000.

[55] J. S. Yi, Y. ah Kang, J. T. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1224–1231, 2007.

[56] M. X. Zhou and S. K. Feiner. Visual task characterization for automated visual discourse synthesis. In *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 392–399, 1998.