Quantitative Effectiveness Measures for Direct Volume Rendered Images

Yingcai Wu Huamin Qu Ka-Kei Chung Ming-Yuen Chan * Hong Zhou †
The Hong Kong University of Science and Technology Shenzhen University

ABSTRACT

With the rapid development in graphics hardware and volume rendering techniques, many volumetric datasets can now be rendered in real time on a standard PC equipped with a commodity graphics board. However, the effectiveness of the results, especially direct volume rendered images, is difficult to validate and users may not be aware of ambiguous or even misleading information in the results. This limits the applications of volume visualization. In this paper, we introduce four quantitative effectiveness measures: distinguishability, contour clarity, edge consistency, and depth coherence measures, which target different effectiveness issues for direct volume rendered images. Based on the measures, we develop a visualization system with automatic effectiveness assessment, providing users with instant feedback on the effectiveness of the results. The case study and user evaluation have demonstrated the high potential of our system.

Index Terms: I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques

1 INTRODUCTION

Direct Volume Rendering (DVR) is a powerful and flexible visualization technique for exploring volume data sets. To promote wider adoption of volume rendering in practice, recent research mainly focuses on the intuitiveness [24] and efficiency [6] of volume rendering. However, even if ideally intuitive and efficient volume rendering can be developed, end users may still feel unsatisfied because some ambiguous information such as artifacts could be introduced in volume rendering. Effective volume rendering, on the other hand, has the capability of effectively presenting the results that can be accurately interpreted by perceivers [11].

In many volume visualization applications, the context and the spatial relations between structures are usually important for users to analyze the data and gain insight. Merely displaying important structures one by one automatically is not sufficient. Hence, there is often no automatic approach for delivering effective visualization and user interaction with the system is needed. For instance, to visualize an MRI volume, a predefined transfer function is often needed to be fine-tuned to adapt to the data and modality. Nevertheless, the interactions by end users are often error prone and may introduce misleading information resulting in unreliable conclusions. Therefore, in visualization users often have to rely on their expertise and judgment to determine whether the information has been faithfully revealed or not, thus leading to an inefficient and ineffective visualization. In this scenario, the efficiency is affected because the burden of the evaluation rests on the users. The effectiveness is also limited since the subjective evaluation is usually unreliable and inconsistent. Thus, an effective visualization system should not only generate impressive images for users, but also evaluate the effectiveness of these images.

*e-mail: {wuyc, huamin, kkchung, pazuchan}@cse.ust.hk
†e-mail: hzhou@szu.edu.cn

IEEE Pacific Visualisation Symposium 2010
2 - 5 March, Taipei, Taiwan
978-1-4244-6686-3/10/$26.00 ©2010 IEEE

In recent years, researchers have become increasingly interested in assessing the usability of visualization [13]. Previous approaches usually validate the effectiveness of visualization with formal user studies [7]. These approaches evaluate the effectiveness at a high level and human perception issues are often involved. They are promising and helpful for the visualization community to improve the quality of their systems based on the feedback from users or experts. Nevertheless, previous research has mainly focused on helping visualization experts design and improve their systems rather than providing quick effectiveness feedback on the results to end users. In addition, these approaches cannot be easily quantified and incorporated into visualization systems. This strongly motivates us to develop a new visualization system that can evaluate the effectiveness of a volume rendered image quantitatively and provide instant feedback to users.

In this paper, we investigate the effectiveness issues specifically for *direct volume rendered images* (DVRIs) generated by full *Direct Volume Rendering* (DVR) based on ray casting, since full DVR is more general and complex. To evaluate the effectiveness of DVRIs, two different classes of effectiveness criteria - sufficient criteria and necessary criteria - could be developed. In visualization, if the sufficient criteria are satisfied, the visualization goal can be achieved. For the necessary criteria, if they are not satisfied, the visualization goal cannot be achieved. However, even if these necessary criteria are satisfied, there is still no guarantee that the visualization goal can be achieved. Sufficient effectiveness criteria are the holy grail of volume visualization and may not be possible for many applications in the near future. Therefore, in this paper we just deal with a set of necessary criteria which we believe that most visualization results (DVRIs) should satisfy.

The necessary criteria include four independent effectiveness measures which evaluate the effectiveness of a DVRI from four effectiveness aspects, *i.e.,* the distinguishability of features, the clarity of contours, the edge consistency between DVRIs and the volume, and the depth coherence between features. To the best of our knowledge, these are the first generic quantitative effectiveness measures for DVRIs. Based on the proposed measures, a new visualization system is developed by integrating the quantitative effectiveness evaluation into the visualization pipeline. With the help of our system, various effectiveness issues in DVRIs can be identified in visualization automatically. We believe that it is a first step towards automatic effective DVR.

2 PREVIOUS WORK

Visualization Evaluation has been viewed as one of the top research challenges by many famous visualization scientists in the joint NSF-NIH Fall 2004 Workshop on Visualization Research Challenges [13]. Van Wijk [22] proposed an economical model of visualization and exploited it to assess the value of visualization from the viewpoint of technology, innovation, art, design, and science, respectively. In our paper, we evaluate the effectiveness in the visualization process mainly from the technology viewpoint. Recently, user studies have attracted much attention from the community and have been used to examine the effectiveness of visualization results [7]. However, it often takes a long time and many resources for researchers to design and conduct appropriate user

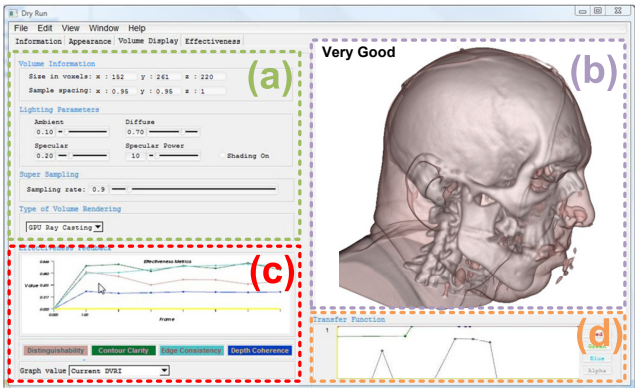

Figure 1: User interface for our system: (a) Region for adjusting parameters for DVR; (b) Region for presenting the DVRI; (c) Region for providing effectiveness feedback; (d) Region for specifying the transfer function.

studies and analyze the results [10]. Tory and Möller proposed to exploit expert evaluation when user studies fail [20]. Compared with the previous methods that validate the effectiveness for the visualization experts to improve their systems, our measures provide quick quantitative effectiveness feedback to end users automatically.

Volume Exploration by automatically selecting good views is also closely related to the effectiveness of visualization, as good viewpoints may deliver informative results that can effectively reveal the features in a volume [1, 19]. Our framework addresses a different problem and focuses on automatic effectiveness evaluation of the DVRIs rendered from a set of viewpoints which are either generated automatically by these methods or selected manually by users. Transfer function (TF) plays an important role in volume exploration [21]. Kaufman and Mueller provided an excellent survey on TF design [8]. Despite the great advance in the research of TF design, automatic TF specification is still very difficult and user interaction is usually needed [14]. However, the user interaction in specifying TFs may introduce errors and is less reliable. Thus, quantitative effectiveness evaluation and feedback will be very helpful in guiding users through a visualization process.

Image Quality Assessment is a well investigated problem in image and video processing. A comprehensive review on this is beyond the scope of this paper and interested readers can refer to [2] for further details. We review only some closely related work in volume visualization. Meißner *et al.* [12] compared four volume rendering algorithms by considering image quality and computational complexity. Wang *et al.* [23] developed an image-based quality measure to facilitate interactive level-of-detail exploration for large volumes. Their measure assesses the quality of images by efficiently measuring the contribution of multi-resolution data blocks to the final images. An automatic image enhancement approach was introduced by Chan *et al.* [3] using a quality measure for evaluating the image contrast based on the information obtained from the images as well as the volumetric data. Compared with the previous quality assessment approaches, our measures focus on the generic and important effectiveness issues of DVRIs which end users can easily understand. Some ambiguous and misleading information can be identified by our measures as well. To the best of our knowledge, these issues have not been thoroughly addressed before in volume rendering.

3 System Overview and User Interface

Our system is a comprehensive DVR system equipped with quantitative and automatic effectiveness evaluation as well as intuitive and quick effectiveness feedback. Fig. 1 shows the user interface of our system which can be divided into four parts: Region (a) is for adjusting parameters such as lighting settings and sampling rates; Region (b) presents the resulting DVRI; Region (c) is used to provide quantitative effectiveness feedback to users; Region (d) is for specifying the transfer function. The effectiveness evaluation mechanism built upon our proposed quantitative effectiveness measures can assess the effectiveness of a DVRI or a whole visualization process.

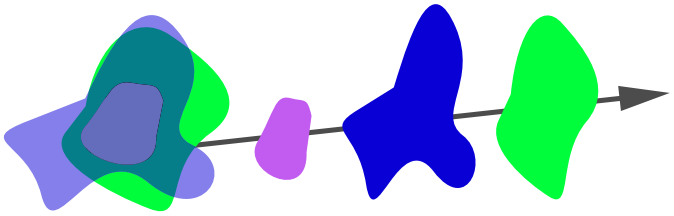

Figure 2: Three features in purple, blue, and green are composited with opacity 0.5, 0.5, and 1. The purple feature cannot be distinguished from the blue feature after composition.

Following a well-known InfoVis principle - "Overview first, zoom and filter, then details-on-demand" proposed by Shneiderman, the system provides the effectiveness feedback to users at different levels of detail, *i.e.*, the semantic level, the middle level, and the detail level based on users' demands. At the semantic level, our system shows a semantic term ("very good", "fairly good", "not good", or "bad") summarizing the overall effectiveness in the left top corner of Region (b) in Fig. 1, such that users can have a basic idea of the effectiveness with minimum distraction. The system estimates the overall effectiveness by a weighted average of effectiveness values computed by different effectiveness measures. The system treats all measures equally important and thus set their weights to be 1. At the middle level, the system provides users with moderate details about the effectiveness through an effectiveness graph (see Region (c) in Fig. 1). In the graph, the horizontal axis represents the DVRIs that we have explored and the vertical axis indicates the effectiveness values. The curves in different colors on the graph represent the effectiveness values computed by different measures. The effectiveness graph can be updated almost in real time. Whenever users stop at a certain viewpoint to explore the volume in detail, the system can utilize the idle time for updating the graph. At the detail level, it allows users to backtrack to any visualization step by clicking on the effectiveness graph. All settings will be adapted subsequently to the step and the DVRI shown in Region (b) will be updated accordingly. It also enables the detailed effectiveness examination by directly highlighting the detected abnormality in the DVRI (see Fig. 4(c) and (e)).

4 Effectiveness Measures

In this section, we introduce four independent effectiveness measures - distinguishability, edge consistency, contour clarity, and depth coherence measures - to address the effectiveness issues of a DVRI from different perspectives.

4.1 Distinguishability Measure

The distinguishability measure evaluates how well features in a DVRI can be visually differentiated from one another by estimating the color similarity between the features. It is particularly useful in DVR. Because of the blending effect in DVR, different features may appear to be visually similar even if they are assigned distinct colors in the transfer function (see Fig. 2 for an example).

The measure works as follows: Given a DVRI, the measure first segments it into a number of regions by an efficient graph-based

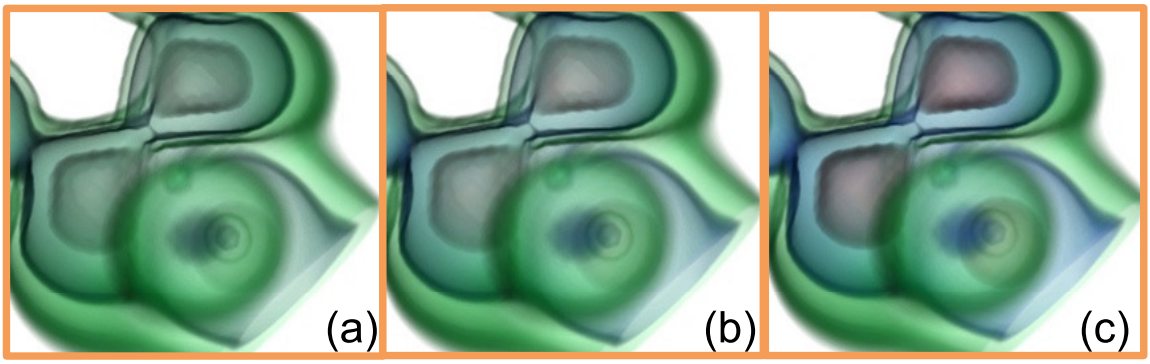

Figure 3: Three DVRIs with varying distinguishability values. The computed effectiveness values for (a), (b), and (c) were 0.17 and 0.43, and 0.85, respectively.

image segmentation algorithm [5]. It then builds a histogram for each region to show the scalar value distribution of the samples on the rays cast from the region into the volume. The samples within homogeneous areas (*i.e.,* the samples that satisfy the condition: $g < \varepsilon$ where g is the gradient value of a sample and ε is a predefined threshold and $\varepsilon = 0.05$ in our experiments) will be filtered out when building the histogram. This can ensure that small features will not be overwhelmed by larger ones in the histogram. For any two regions, our measure estimates whether they belong to the same feature by calculating the similarity of their histograms. We employ the cross correlation method to measure the degree of similarity between histograms as follows:

$$r = \frac{\sum_{i=1}^{m}(H_1(i)-\overline{H_1})(H_2(i)-\overline{H_2})}{\sqrt{\sum_{i=1}^{m}(H_1(i)-\overline{H_1})^2\sum_{i=1}^{m}(H_2(i)-\overline{H_2})^2}} \quad (1)$$

where

- $\overline{H_1} = \sum_{i=1}^{m} H_1(i)/m$ and $\overline{H_2} = \sum_{i=1}^{m} H_2(i)/m$
- m is the number of scalar values in each histogram
- $H_1(i)$ and $H_2(i)$ are the values of the histograms at i

If any two regions with a similar color (the similarity is computed as the color distance in *L*a*b** color space) have a weak correlation, *i.e.,* the pair is likely to be of different features but has a similar color, they should be viewed as an indistinguishable pair. The effectiveness value can then be derived as the ratio of the distinguishable pairs over the total number of the pairs.

Fig. 3 presents three DVRIs generated from a neghip protein molecule dataset ($64 \times 64 \times 64$). All DVRIs reveal three important isosurfaces assigned similar opacity but different colors in the transfer functions. However, because of the blending effect introduced by DVR, the isosurfaces can hardly be distinguished in Fig. 3(a) and can barely be differentiated in Fig. 3(b). In contrast, we can easily discriminate the isosurfaces in Fig. 3(c). The computed effectiveness values for Fig. 3(a), 3(b), and 3(c) were 0.17 and 0.43, and 0.85, respectively, demonstrating that our measure can successfully detect the ambiguity.

4.2 Edge Consistency Measure

Edge consistency is another important effectiveness issue for DVRIs. According to the Gestalt principle of continuity, humans are often inclined to observe objects as smooth and continuous [25]. Any false discontinuity (or false edges) shown in DVRIs but actually not existing in the corresponding volumetric data would likely confuse or mislead users. In volume visualization, false edges are usually caused by various artifacts such as filtering artifacts, classification artifacts, and shading artifacts [4]. These results are of course ineffective for conveying information to users, and should be identified in the visualization process. To address the issue, we propose an edge consistency measure to find any possible false discontinuity presented in the DVRIs.

Algorithm 1: ConsistencyAlg

```
1  Estimate the edge image E of the input DVRI;
2  for Each edge pixel (x,y) in E do
3  |   Cast a ray into the volume data and do ray composition
   |   like volume rendering;
4  |   Let c ← the opacity weighted color contribution of all the
   |   feature lines passed through by the ray;
5  |   if c > t then                    /* t is a threshold */
6  |   |   Record E(x,y) as a correct edge pixel;
7  |   else
8  |   |   Record E(x,y) as a false edge pixel;
9  |   end
10 end
```

Our edge consistency measure works as follows: For each pixel showing discontinuity (*i.e.*, edges) in the DVRI, our measure checks whether the ray shoot from this pixel passes through any feature lines such as silhouettes and valley creases. If the ray does not pass through any feature lines, then the discontinuity indicated by this pixel is false and users should be alerted. Algorithm 1 is the pseudo-code of the measure. The edge image E is extracted by Canny edge detector and the threshold t is 0.05 in our experiments.

Our measure considers only the most important feature lines - silhouettes, ridges, and valley creases - in the data for simplicity and efficiency. For each sample point along a ray, it is straightforward to know whether it belongs to a silhouette or not by checking $|v \cdot n|$ where n is the normal and v is the ray direction at the sample. If $|v \cdot n| < t$ where t is a certain threshold, the sample is on a silhouette. We employ an approach proposed by Kindlmann *et al.* [9] to detect the ridge and valley creases, based on the principal curvatures κ_1 and κ_2 which are evaluated as follows:

$$\kappa_1 = \frac{T + \sqrt{2F^2 - T^2}}{2}, \kappa_2 = \frac{T - \sqrt{2F^2 - T^2}}{2} \quad (2)$$

where T and F are the trace and Frobenius norm of G, respectively, and $G = -PHP/|g|$ and $P = I - nn^{\mathrm{T}}$, and $n = -g/|g|$ where g is the gradient and H is the Hessian matrix of the sample point. We can then compute the edge consistency effectiveness value of the input DVRI as $e = n_1/(n_1 + n_2)$ where n_1 and n_2 are the numbers of the correct edge pixels and false edge pixels, respectively.

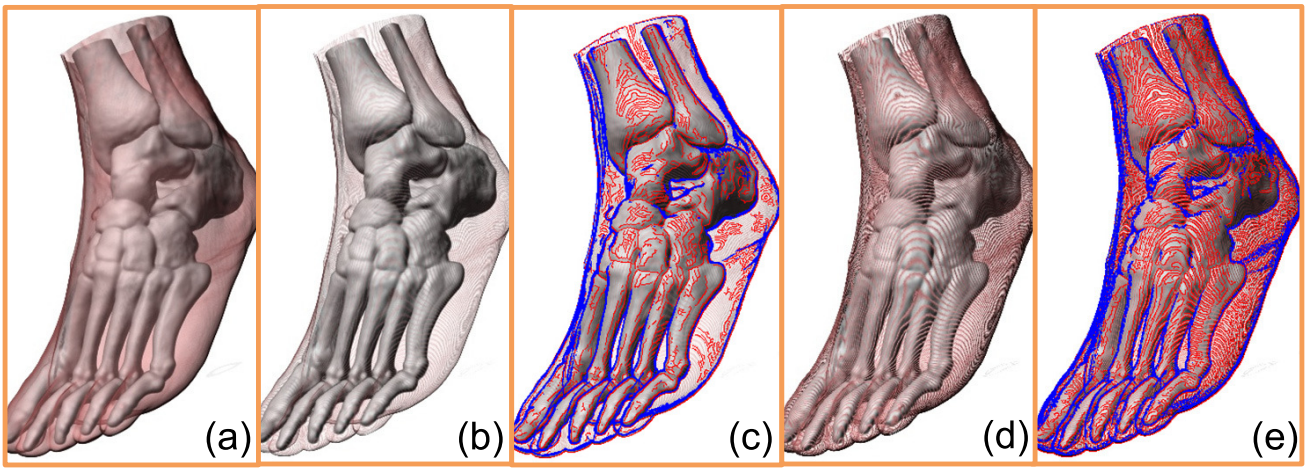

Figure 4: The measured edge consistency values for (a), (b), and (d) are 0.95, 0.7, and 0.3, respectively; (c) and (e) highlight the false edges of (b) and (d), respectively, using red strokes.

Fig. 4(a), (b), and (d) show three DVRIs generated from a CT human foot volume data set ($152 \times 261 \times 220$) with different amounts of false edges. The false edges are mainly caused by the classification artifacts introduced by inappropriate transfer functions. The measured edge consistency values for the three DVRIs are 0.95, 0.7, and 0.3, respectively. Fig. 4(c) and (e) highlight the false edges of Fig. 4(b) and (d), respectively, using red strokes. The blue strokes represent the correct edges. We have also tested the measure on DVRIs with false edges caused by other types of artifacts such as filtering artifacts, and found that the measure also worked well and could identify the false edges effectively.

4.3 Contour Clarity Measure

The clarity measure evaluates how clear the contours of important isosurfaces are presented in a DVRI. In volume rendering, isosurfaces are often set to be semi-transparent and are blended with one another, and thus their contours may become fuzzy and unclear in DVRIs. The unclear contours could confuse or even mislead users as the contours are important clues that help people perceive 3D structures correctly [25]. Thus, the measure is useful and necessary for evaluating the effectiveness of DVRIs. The measure first identifies some important isosurfaces from the DVRI, and then estimates the contour clarity of the isosurfaces in the DVRI.

The isosurfaces can be estimated by taking into account the edges of the DVRI as well as the volumetric information, since the isosurfaces can usually be represented by the edges in the image. The algorithm of detecting the important isosurfaces works as follows: It extracts the edges in the DVRI by the Canny edge detector, and builds a histogram of the scalar values of the samples in the volume on the rays cast from the edge pixels. In addition, the samples must be on the feature lines in the volume (see Section 4.2 for the detection of the feature lines) to filter out irrelevant samples. Finally, the important isovalues can be detected by estimating the peaks of the histogram, and so the important isosurfaces can be detected accordingly. The important isosurfaces that can be detected by the measure are not necessarily identifiable for end users.

After the isosurfaces are obtained, the measure renders the isosurfaces separately into an isosurface DVRI. Two edge images are created to store the edges extracted from the original DVRI and the isosurface DVRI by the Canny edge detector, respectively. The measure then compares the edge images to determine the clarity of the isosurfaces' contours in the DVRI. We propose a new method called *Block Correspondence measure* (BCM) extended from *Pixel Correspondence measure* (PCM) [15] for the comparison.

The PCM transforms the image comparison problem into an optimal matching problem which is to find a matching with a minimum accumulated cost of each match in a bipartite graph G. The bipartite graph can be defined as follows: Let $G(V,E)$ be a graph where V and E represent the set of the vertices and the set of edges, respectively. A bipartite graph is a graph $G(V,E)$ where V can be divided into two disjoint sets V^+ and V^- such that the end-points of each edge should belong to different sets of vertices. A matching $\mathscr{M}$ is a subset of E in which no two edges share a vertex. A vertex is matched if it is incidental to an edge of $\mathscr{M}$ or otherwise unmatched.

Suppose $f(i,j)$ and $g(k,l)$ are two pixels in two edge images f and g, respectively. PCM treats the pixels in g as V^+ and the pixels in f as V^-. The set of edges (E) contains all the edges defined as follows: An edge between $f(i,j)$ and $g(k,l)$ exists if $\max(|k-i|,|l-j|) \leq \eta$ where η is the maximum localization error and $\eta = 3$ in our implementation. In other words, $f(i,j)$ should be matched from the neighbors of pixel (i,j) in g within radius η. The cost of a match between $f(i,j)$ and $g(k,l)$ is defined as:

$$C(f(i,j),g(k,l)) = 1 - S((i,j),(k,l))(1 - \frac{|f(i,j) - g(k,l)|}{Z}) \quad (3)$$

where Z is the maximum pixel value in g and f, C is between 0 and 1 where 0 indicates a perfect match of the two pixels, and separation S is defined as:

$$S((i,j),(k,l)) = W(max(|k-i|,|l-j|)), \quad (4)$$

where $W(d)$ is a normalized function defined as $W(d) = 1, 0.9, 0.69, 0.5 | d = 0...3$. It penalizes the distance in a 2D image space between the pixels. The similarity between two edge images f and g can be computed as:

$$\mathrm{PCM}_\eta(f,g) = 1 - \zeta(\mathscr{M}_{opt}(f,g))/|f \bigcup g|, \quad (5)$$

where $\mathscr{M}_{opt}$ is an optimal matching between f and g, $\zeta(\mathscr{M}_{opt}(f,g))$ is the accumulated cost of the matches in $\mathscr{M}_{opt}$ between f and g, and $|f \cup g|$ indicates the number of the non-zero pixels in f or in g. Note that if there is any pixel (*i.e.*, unmatched vertices in graph G) in g and f left unmatched, PCM adds its pixel value to ζ. Finally, we can apply an algorithm [17] to find the optimal matching.

Although PCM adopts an approximation to the optimal matching, it is still time-consuming. Thus, we propose BCM to accelerate

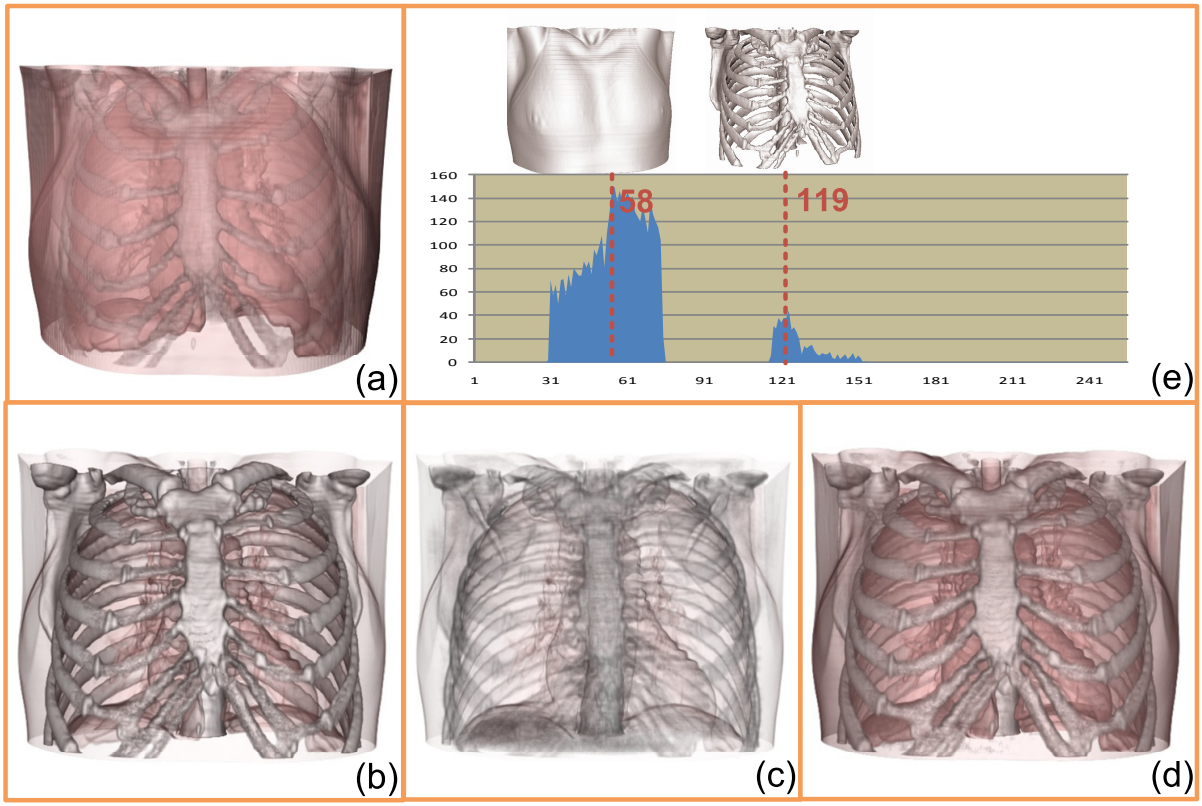

Figure 5: (a)-(d) Four DVRIs with varying contour clarity (0.47, 0.56, 0.14, and 0.92, respectively); (e) A histogram built by the clarity measure to estimate the isosurfaces presented in (a).

the performance. BCM is based on blocks instead of pixels, and so we first divide each edge image with the size $X \times Y$ into a set of blocks. Let $f(i_b, j_b)$ and $g(k_b, l_b)$ be two blocks in f and g, respectively, and $i_b, k_b \in [0, \frac{X}{m})$ where m is the height of the block, and $j_b, l_b \in [0, \frac{Y}{n})$ where n is the width of the block. In our experiments, we set $m = n = 2$ and found that our algorithm runs much faster than PCM while maintaining accuracy. BCM computes the cost of a match $C(f(i_b, j_b), g(k_b, l_b))$ between two blocks $f(i_b, j_b)$ and $g(k_b, l_b)$ as:

$$1 - S((i_b, j_b), (k_b, l_b))(1 - \frac{\|f(i_b, j_b) - g(k_b, l_b)\|_F}{Z|f(i_b, j_b) \cup g(l_b, k_b)|}) \qquad (6)$$

where $\|f(i_b, j_b) - g(k_b, l_b)\|_F$ is the Frobenius norm - the most frequently used matrix norms in numerical linear algebra. $|f(i_b, j_b) \cup g(l_b, k_b)|$ is the number of non-zero pixels in block $f(i_b, j_b)$ or $g(k_b, l_b)$. Finally, the BCM similarity value between two edge images f and g can be computed as:

$$\mathrm{BCM}_\eta(f, g) = 1 - \zeta(\mathscr{M}_{opt}(f, g))/|f \bigcup g|_b, \qquad (7)$$

where $|f \bigcup g|_b$ is the number of non-zero blocks in f or in g.

Fig. 5(a)-(d) show four DVRIs generated from a CT Chest volume ($384 \times 384 \times 240$) with varying contour clarity. A histogram in Fig. 5(e) is built by the measure to estimate the important isosurfaces in Fig. 5(a). The identified isovalues are 58 and 119. With the important isosurfaces, the measure performs the BCM algorithm to quantify the contour clarity of Fig. 5 (a), and the measured effectiveness value is 0.47. Similarly, the contour clarity values for Fig. 5 (b), (c), and (d) are 0.56, 0.14, and 0.92, respectively. This is quite consistent with our observation, demonstrating that the measure can successfully evaluate the contour clarity of DVRIs.

4.4 Depth Coherence Measure

Coherence properties, such as depth coherence, spatial coherence, and image coherence, have been widely exploited in computer graphics. Among those coherence properties, the depth coherence, which denotes that the depth at a surface point changes gradually to its neighborhood, is often assumed by users when they are inspecting the surface of a specific feature. The incoherence between the depth of the feature perceived by users in 2D images and the real depth in 3D volumes usually results in ambiguity. For example, in vascular volume visualization, separate branches may appear as one branch in the 2D image because of the depth incoherence. To address the problem, we design a new algorithm to automatically detect the depth incoherence in DVRIs.

The measure works as follows: It first obtains a layered depth image (LDI) [18] for the DVRI and then checks the depth coherence of different features. The system casts rays into the volume and records the depth values of the surface layers (which can be determined by gradient magnitude [16]) that each ray passes through into a LDI. After that, the system performs a low-level over segmentation using the watershed algorithm for the DVRI to get a number of atomic regions. An efficient high-level graph-based segmentation [5] is also performed to simulate how users group or merge the atomic regions. If two adjacent atomic regions (A and B) are expected to be merged by the high-level segmentation algorithm, which means users tend to perceive these two regions as one feature, the system will check whether depth coherence exists for these two regions. To do so, for each pixel i on A's boundary adjacent to B, we compare its stored set of depth values to the set of depth values of its neighboring pixel on B's boundary. If these two sets of depth values do not have any match, the pixel i is viewed as an incoherent pixel. If most of the pixels on A's boundary adjacent to

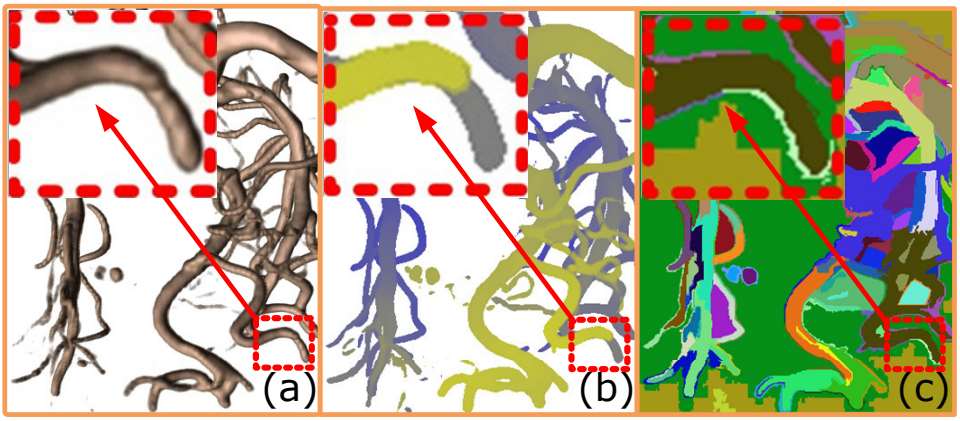

Figure 6: A DVRI and its depth image and overly segmented image where some depth incoherence exists.

B are coherent pixels, A and B are coherent in depth, or otherwise incoherent.

The depth coherence value e of a certain DVRI can be estimated as the ratio of the number of coherent pairs over the number of all adjacent pairs examined by the above algorithm. The pseudo-code of the measure is shown in Algorithm 2.

Algorithm 2: CoherenceAlg

```
1  Cast rays into the data and obtain a LDI;
2  Perform a watershed algorithm to obtain a overly segmented
   image I_0 from an input DVRI;
3  Perform a graph-based algorithm to obtain a segmented image
   I_1 from the input DVRI;
4  Scan I_0 and build a 2D array A with the size of n × n, where n
   is the number of atomic regions in I_0 and A[i][j] stores the sets
   of depth values obtained from the LDI of the boundary pixels
   of region i adjacent to region j;
5  Let c ← 0 and num ← 0;
6  for Each region k in I_1 do
7      for Each pair (m,n) of the adjacent regions of I_0 in k do
8          if A[m][n] is different from A[n][m] then
9              Record (m,n) as incoherent pairs;
10         else
11             c ← c+1;
12         end
13         num ← num+1;
14     end
15 end
16 Output the effectiveness value: e = c/num;
```

Fig. 6(a) is a DVRI generated from a CT aneurism volume ($256 \times 256 \times 256$) where depth incoherence exists. For example, two separated small branches in the red rectangle appear like one branch. Fig. 6(b) is the depth image of Fig. 6(a) where a yellow-to-blue color encoding scheme is used to represent the depth from close to far. Fig. 6(c) is the overly segmented image of 6(a). This abnormal case can be detected by our coherence measure successfully and the measured coherence value is 0.77.

4.5 Effectiveness for A Whole Visualization Process

Besides the effectiveness assessment for each DVRI, our system can also evaluate the effectiveness of a whole visualization which consists of a set of DVRIs. For each measure, we collect all viewpoints that have been explored so far. For each explored viewpoint, the system would record a highest effectiveness value for each measure and treat it as the measure value at this viewpoint for the whole process. For example, if an unclear contour in a DVRI becomes clear in a new DVRI from the same viewpoint, we believe the contour is well revealed from this viewpoint and the system would store the effectiveness value of the new DVRI for this viewpoint.

Among all explored viewpoints' recorded effectiveness values, we can treat the lowest effectiveness value as the measure's effectiveness value of the process. For instance, as long as a contour is not well revealed from one viewpoint, the whole process should not be considered effective as an important piece of information is still missing for this viewpoint. To improve the effectiveness value for the whole process, users should go back to the viewpoint and fix the problem if this viewpoint is important, or delete the cached DVRIs of this viewpoint if this viewpoint is not important.

5 Experiments and Discussion

In this section, we demonstrate the usefulness of our measures in an informal user study and a case study. The system was tested on an IBM laptop T61p (2.4GHz Intel Core 2 Duo, 2GB RAM, NV Quadro FX570M (256MB)). Each measure can be computed quite fast. Together, our system needs one or two seconds to measure the effectiveness of a DVRI.

5.1 User Evaluation

Six participants (four doctors, one biologist, and one medical imaging expert) from different fields were invited to evaluate our system. The four doctors are from two hospitals and one medical school. They have used commercial medical systems such as "Mimics Medical Imaging Software" for pre-surgery planning of orthopedical surgeries, "Volume Viewer Plus" at GE's Advantage Workstation to do radiation therapy planning, and Siemens's "Syngo Somaris/5 VA47C" for surgery planning. One doctor in the cardio-respiratory department of a hospital uses DVRIs for daily medical diagnoses. The biologist and the medical imaging expert have used some VTK-based volume visualization systems for their own research.

5.1.1 Evaluation Procedure

The user evaluation includes three steps. We first introduced our effectiveness measures to the subjects, and then asked them to rate their understanding and the usefulness of each measure in their areas. After that, the subjects used our system to freely explore four typical volumetric datasets, including a CT Foot dataset ($152 \times 261 \times 220$), a CT visible full human body dataset ($512 \times 512 \times 626$), an MRI knee dataset ($512 \times 512 \times 87$), and a CT Aneurism dataset ($256 \times 256 \times 256$); Finally, the subjects were asked to give feedback about the performance of each measure and the intuitiveness

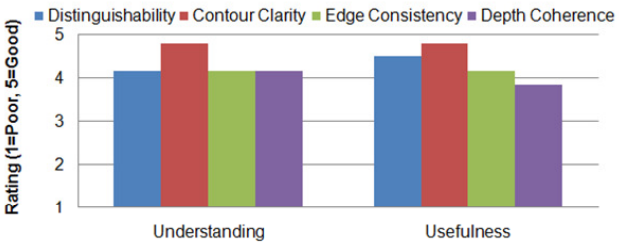

Figure 7: Average ratings of the understanding and usefulness of the measures.

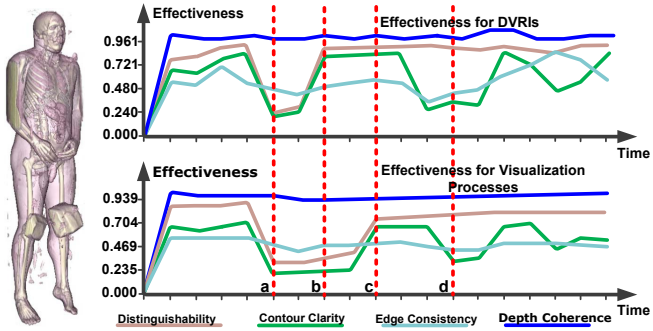

Figure 8: Effectiveness graphs for a typical visualization. : upper graph shows the effectiveness values for each DVRI and the bottom graph is the corresponding accumulated effectiveness graph. Color encodes the measure types.

and the usefulness of the system as well as any suggestion for future improvements. It took each subject on average one and a half hours to finish the evaluation.

5.1.2 Evaluation Results

Fig. 7 shows the average ratings on the understanding and usefulness of the measures from the user evaluation. Qualitatively, we can see that all measures were received well and were rated relatively high by the participants. All participants had a good understanding (above 3.9 in average) of all the effectiveness measures, and did not have any problems differentiating four measures. They also felt that the measures would be useful for different applications. Among the four measures, some of them were especially appreciated by the participants. One doctor from the Cardio-respiratory Department especially liked the depth coherence measure, as the measure could help him fix the depth ambiguity which usually occurs in vascular volume visualization. One radiologist from the department of radiation oncology appreciated the edge consistency measure. He often noticed some suspicious artifacts using his current systems, but he was not sure whether they were really artifacts or just data features. He was glad that the system can automatically tell him whether these are artifacts or not. They confirmed that some misleading information could be detected with the help of our system. Our level-of-detail effectiveness feedback mechanism was also received quite well. They all believed our approach has high potential and wanted to see these features in their systems because these measures and the built-in assessment made them feel more confident in the visualization results.

5.1.3 Suggestions

The participants provided some helpful suggestions for future improvements. These included:

- A more detailed description about a sudden change of an effectiveness value in the effectiveness graph to clarify the reason of the abnormal change.
- A mechanism to inform users when the effectiveness of the whole visualization is sufficiently good or warn the user if it is bad.
- An automatic approach to correct the detected errors or tell users how to improve the effectiveness.

We also received some unexpected feedback. Our effectiveness measures were previously named as effectiveness metrics. However, the biologist pointed out that the term "metric" sounds too technical and may have different meanings in other fields and thus suggested we call it measure instead of metric. Our system previously provided an additional measure called entropy measure which can evaluate the correlation field of volume entropy and DVRI entropy based on information theory. We originally thought it would be quite useful for data exploration because it can measure whether the entropy in a volume region matches the entropy in the corresponding DVRI region and thus provide a high level effectiveness evaluation. However the feedback from the participants was disappointing. Most of the participants had problems to understand the concept and also did not know how to improve the rendering if the measure indicates a low effectiveness value. Therefore, we finally decided to remove it from our system.

Overall, all participants gave us very positive feedback, which indicates that our framework is useful and promising for volume visualization. The user evaluation confirmed the usefulness of the measures and identified several avenues for future study and development. However, as the user evaluation was just an initial study and was not conducted in real applications, it has some limitations. In future, we will integrate our framework into a real medical system to perform further testing.

5.2 Case Study

In the case study, we analyze the effectiveness feedback for a typical visualization, selected from the user evaluation, in which the user freely explored the CT human full body volume. The goal of the case study is to demonstrate the usefulness of our system for providing quantitative effectiveness feedback to users. Fig. 8 shows two effectiveness graphs - the upper graph shows the effectiveness values of each measure for each frame, and the bottom graph presents the effectiveness values from the beginning of the visualization to each frame.

From the upper effectiveness graph, we can see that the depth coherence values are around 0.95 in the visualization, indicating that the structures in the data are coherent in depth at any time frame in the visualization. The distinguishability and contour clarity curves are similar. This is reasonable as a feature that is easy to distinguish usually has clear contours. However, in the upper graph for DVRIs around time frame *d* highlighted by a red vertical dashed line, the distinguishability and contour clarity values are different. This case has also been demonstrated in Fig. 3(a) where the distinguishability value is 0.17 and clarity value is 0.8, which is fairly consistent with our observation. The effectiveness values of the edge consistency are not as good as others and are below 0.5 for most of the DVRIs. After investigation into the visualization, we found that the problem was mainly caused by the classification artifacts introduced by the improper transfer functions.

The effectiveness graph in the bottom describes the lowest effectiveness values of the visualization process, which keep track of the viewpoints where the data were not well revealed. For example, although the contour clarity value of the DVRI at frame *b* is high (in the upper graph), the corresponding effectiveness value of the whole process (at the bottom graph) is still so low, as from a typ-

ical viewpoint (from which the DVRI at frame *a* was created) the contours of the features are unclear and this is not fixed in the following several DVRIs. After being notified by the graph, the user went back to the same viewpoint at frame *a* and made the contours clearer by adjusting the opacity transfer function, which greatly improved the effectiveness of the whole process (as shown at position *c* in the bottom graph). The visualization graphs also indicate that, in the whole visualization process, the contours of the important feature were not quite clear but still over 0.5 in average and there were some false discontinuities.

5.3 Discussion

The experiments have clearly demonstrated the usefulness of our measures. Each measure targets a quite unique effectiveness problem which cannot be addressed by any other measures. There is usually no correlation between different measures and they are independent from one another, as demonstrated in the case study. Thus, they are all useful for some applications. However, not all measures are useful in all applications. The usability of a specific measure depends on the underlying data as well as the applications. For example, the coherence measure becomes useless when there are only a few features and their spatial relations are clear in the data.

The effectiveness of DVRIs is a complicated issue and depends on tasks, algorithms, and systems. Formal user studies and expert reviews are still indispensable for effectiveness evaluation at a high level. We only focus on the generic effectiveness measures that can be quantitatively computed and readily integrated into typical visualization systems, providing users with the feedback on the integrity and correctness of the presented information. These measures are especially valuable for end users when they need to interactively fine tune the results. Automatically generating DVRIs that satisfy these measures is very difficult and it involves transfer function design, viewpoint selection, lighting design, and anti-aliasing. Because of the huge search space, typical optimization methods may fail. We will investigate this issue in future.

6 Conclusions and Future Work

In this paper, we present four effectiveness measures: distinguishability, contour clarity, depth coherence, and edge consistency measures for evaluating the effectiveness of a DVRI. Based on our measures, a new visualization system with built-in automatic effectiveness assessment has been developed. Users will be quickly informed about the effectiveness of the DVRIs in visualization. Ambiguity like depth incoherence in DVRIs could be identified easily. We believe that our measures can facilitate the wider adoption of visualization techniques in some real applications. Our system focuses on detecting the effectiveness problems but cannot fix the problems automatically. Automatic generation of effective results based on our effectiveness measures is worth further study. Another avenue for our future study is to conduct a formal task-based user study to assess the appropriateness of each measure in different applications. We also plan to study the influence of the human perception and cognitive factors to the effectiveness of DVRIs, and integrate computational cognitive models to our measures.

Acknowledgment

This work was supported in part by grant HK RGC CERG 618706 and GRF 619309. We would like to thank Anbang Xu for his contributions to the contour clarity metric.

References

[1] U. Bordoloi and H.-W. Shen, "View selection for volume rendering," in *IEEE Visualization*, 2005, pp. 487– 494.

[2] A. C. Bovik, *Handbook of Image and Video Processing*, 2nd ed. Academic Press, 2005.

[3] M.-Y. Chan, Y. Wu, , and H. Qu, "Quality enhancement of direct volume rendered images," in *Volume Graphics*, 2007, pp. 25–32.

[4] K. Engel, M. Hadwiger, J. M. Kniss, C. Rezk-Salama, and D. Weiskopf, "Improving image quality," in *Real-Time Volume Graphics*, 1st ed. A K Peters, 2006.

[5] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[6] N. Fout and K.-L. Ma, "Transform coding for hardware-accelerated volume rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1600–1607, 2007.

[7] J. Giesen, K. Mueller, E. Schuberth, L. Wang, and P. Zolliker, "Conjoint analysis for measuring the perceived quality in volume rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1077–2626, 2007.

[8] A. Kaufman and K. Mueller, *The Visualization Handbook*, 1st ed. Academic Press, 2005, ch. Overview of Volume Rendering, pp. 127–174.

[9] G. Kindlmann, R. Whitaker, T. Tasdizen, and T. Möller, "Curvature-based transfer functions for direct volume rendering: Methods and applications," in *IEEE Visualization*, 2003.

[10] R. Kosara, C. G. Healey, V. Interrante, D. H. Laidlaw, and C. Ware, "User studies: Why, how, and when?" *IEEE CG&A*, vol. 23, no. 4, pp. 20–25, 2003.

[11] J. Mackinlay, "Automating the design of graphical presentations of relational information," *ACM TOG*, vol. 5, no. 2, pp. 110–141, 1986.

[12] M. Meißner, J. Huang, D. Bartz, K. Mueller, and R. Crawfis, "A practical evaluation of popular volume rendering algorithms," in *IEEE Symposium on Volume Visualization*, 2000.

[13] NIH/NSF, "Position papers," Workshop on visualization research challenges, 2004.

[14] H. Pfister, B. Lorensen, C. Bajaj, G. Kindlmann, W. Schroeder, L. S. Avila, K. Martin, R. Machiraju, and J. Lee, "The transfer function bake-off," *IEEE CG&A*, vol. 21, no. 3, pp. 16–22, 2001.

[15] M. S. Prieto and A. R. Allen, "A similarity metric for edge images," *IEEE PAMI*, vol. 25, no. 10, pp. 1265 – 1273, 2003.

[16] C. Rezk-Salama and A. Kolb, "Opacity peeling for direct volume rendering," *Computer Graphics Forum*, vol. 25, no. 3, pp. 597–606, 2006.

[17] H. Saip and C. Lucchesi, "Matching algorithms for bipartite graphs," Universidade Estadual de Campinas, Tech. Rep., 1993.

[18] J. Shade, S. Gortler, L. wei He, and R. Szeliski, "Layered depth images," in *SIGGRAPH*, 1998, pp. 231–242.

[19] S. Takahashi, I. Fujishiro, Y. Takeshima, and T. Nishita, "A feature-driven approach to locating optimal viewpoints for volume visualization," in *IEEE Vis*, 2005, pp. 495 – 502.

[20] M. Tory and T. Möller, "Evaluating visualizations: Do expert reviews work?" *IEEE CG&A*, vol. 25, no. 5, pp. 8–11, 2005.

[21] F.-Y. Tzeng, E. B. Lum, and K.-L. Ma, "An intelligent system approach to higher-dimensional classification of volume data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 3, pp. 273–284, 2005.

[22] J. van Wijk, "Views on visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 4, pp. 421–432, 2006.

[23] C. Wang, A. Garcia, and H.-W. Shen, "Interactive level-of-detail selection using image-based quality metric for large volume visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 1, pp. 122–134, 2007.

[24] Y.-S. Wang, T.-Y. Lee, and C.-L. Tai, "Focus+context visualization with distortion minimization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1077–2626, 2008.

[25] C. Ware, *Information Visualization: Perception for Design*. Morgan Kaufmann, 2004.