DOI: 10.1111/cgf.12915
Eurographics Conference on Visualization (EuroVis) 2016
K.-L. Ma, G. Santucci, and J. van Wijk
(Guest Editors)

Volume 35 (2016), Number 3

Similarity Voting based Viewpoint Selection for Volumes

Yubo Tao, Qirui Wang, Wei Chen†, Yingcai Wu, and Hai Lin

State Key Laboratory of CAD&CG, Zhejiang University, P. R. China

Abstract

Previous viewpoint selection methods in volume visualization are generally based on some deterministic measures of viewpoint quality. However, they may not express the familiarity and aesthetic sense of users for features of interest. In this paper, we propose an image-based viewpoint selection model to learn how visualization experts choose representative viewpoints for volumes with similar features. For a given volume, we first collect images with similar features, and these images reflect the viewpoint preferences of the experts when visualizing these features. Each collected image tallies votes to the viewpoints with the best matching based on an image similarity measure, which evaluates the spatial shape and appearance similarity between the collected image and the rendered image from the viewpoint. The optimal viewpoint is the one with the most votes from the collected images, that is, the viewpoint chosen by most visualization experts for similar features. We performed experiments on various volumes available in volume visualization, and made comparisons with traditional viewpoint selection methods. The results demonstrate that our model can select more canonical viewpoints, which are consistent with human perception.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

Different viewpoints may convey significantly different visual information about features. In order to effectively analyze and understand structures, it is better to suggest the optimal viewpoint for the initial exploration and overview of 3D volumes for users [ZAM11]. This is usually regarded as the problem of optimal viewpoint selection, which can avoid the use of the non-intuitive and trial-and-error viewpoint search process, and provide representative viewpoints for fast browsing through volume collections. Many volume visualization methods also consider viewpoint selection as an important step in their pipelines, such as image based transfer function design [WQ07] and proxy-image based interactive remote visualization [TCM10, CMP14].

Traditional viewpoint selection methods in volume visualization allow users to rotate objects freely and choose a good viewpoint based on their own perception. However, it would be difficult to search for a good viewpoint from scratch due to the high degree of freedom of 3D interaction for general users. Thus, several automatic viewpoint selection methods have been proposed to maximize the amount of the visible information of objects in 2D rendered images based on some deterministic measures of viewpoint quality, for example, surface area entropy [TFTN05], voxel entropy [BS05], opacity entropy [JS06], shape/detail measure [TLB*09], and gradient/normal variation [ZAM11]. Although these deterministic measures are suitable for some applications based on the mathematical soundness, it may be hard to express the familiarity and aesthetic sense of users, which are important in user-preferred viewpoints based on the psychophysical experiments [BTB99].

Visualization experts are more familiar with objects and their functionalities, and they can generally select better representative viewpoints. In addition, it is not necessary to ask experts to compare pairs of views or vote on the optimal viewpoints, as they have already shown viewpoint preferences in their 3D visualizations, such as rendered images and manually generated visualization results. For example, there are already many published papers on volume visualization, such as volume rendering and transfer function design. These papers usually leverage public volumes as examples to demonstrate their methods, and visualization experts generally select their preferred viewpoints for the volumes to express their intents in a clear manner. It may meet their aesthetic criteria, maximize the amount of information about interesting features in the rendered images, or highlight some particular features, such as tumors. These rendered images contain the viewpoint information, and they are usually representative viewpoints for the volume. The main motivation of our approach is to extract the viewpoint information from 3D visualizations and select the optimal viewpoint for relevant volumes.

In this paper, we propose an image-based viewpoint selection model for volume visualization. This model uses the images in vol-

† Corresponding Author: Wei Chen (chenwei@cad.zju.edu.cn)

ume visualization to investigate the viewpoint preference for volumes involved in the images. Images in volume visualization are first collected and classified into several categories/voter datasets, such as the head, fish, and engine. These images are cleaned and organized to compose a voter database, which provides a comprehensive coverage of all available volume categories.

Given a volume, we first select one voter dataset with similar features from the voter database. A shape-appearance based similarity measure for images from the voter dataset and rendered images of the volume is introduced to facilitate the image-based viewpoint selection. Based on this similarity measure, each image in the voter dataset casts votes to the viewpoints with the best matching. The optimal viewpoint for the volume is the viewpoint with the most votes from the images in the voter dataset. In this way, our viewpoint selection model tries to learn how visualization experts choose representative viewpoints for volumes with similar features. It may represent visual saliency that cannot be captured quantitatively in previous measures for viewpoint selection.

The paper is structured as follows. Related work is discussed in Section 2. Section 3 introduces the image-based viewpoint selection model. In Section 4, we describe the shape-appearance similarity measure and the voting scheme for viewpoint selection. In Section 5, we describe how to construct voter datasets, and compare out results to previous methods. Finally, we draw conclusions in Section 6.

2. Related Work

Viewpoint selection is a widely investigated research area in computer vision, computer graphics, and visualization. Existing methods in these areas can be broadly classified into three categories, namely, entropy based methods, feature based methods, and learning based methods.

2.1. Entropy Based Methods

Blanz et al. [BTB99] introduced the "canonical views" by computer-graphics psychophysics. These views are a small number of user-preferred viewpoints and have four attributes: goodness for recognition, familiarity, functionality, and aesthetic criteria. Based on this research, the optimal viewpoint generally provides the most information about 3D objects. Starting with the pioneering work of Vázquez et al. [VFSH01], many viewpoint selection methods use information theory to find the optimal viewpoint for meshes and volumes [Vio07, BRB*13].

Takahashi et al. [TFTN05] presented a decomposition based viewpoint selection method. The volume is decomposed into feature components, and the locally optimal viewpoint for each feature component is computed by the surface area entropy [VFSH01]. The global optimal viewpoint is obtained by combining all locally optimal viewpoints. The voxel entropy [BS05] is used to identify a minimal set of representative viewpoints. Each voxel is assigned a noteworthiness value, such as the opacity, and the optimal viewpoint is the one with voxel visibilities proportional to their noteworthiness. Ji and Shen [JS06] further investigated image-based metrics for viewpoint selection, such as opacity entropy, color entropy, and curvature information. These metrics prefer an even opacity and color distribution with a larger projected area and more perceived curvatures. Vázquez et al. [VMN08] applied multi-scale entropy and algorithmic complexity to select representative viewpoints. Tao et al. [TLB*09] introduced two view descriptors: the shape view descriptor evaluates the overall orientation of visible boundary structures, and the detail view descriptor measures the amount of visible details on boundary structures. An evaluation function based on the viewpoint entropy [ZW10] is introduced to measure how evenly the opacity and luminance value are distributed in the rendered image. Ruiz et al. [RBFS10] proposed an information channel to measure the visibility of each voxel with the set of visible viewpoints. The voxel mutual information is used to measure the informativeness of the viewpoint, which prefers various changes in visibility and more details. Entropy based methods generally do not consider the semantic information of features, and therefore they may not select the optimal viewpoints for semantic-meaning features.

2.2. Feature Based Methods

Feature based methods maximize the amount of the feature information in the images. Hong and Shen [HS07] applied the reflective symmetry to find optimal viewpoints by minimizing symmetric information of features in the volume presented in the images. Zheng et al. [ZAM11] presented a viewpoint suggestion framework, iView. In this framework, features are first clustered based on gradient/normal variation in high-dimensional space, and promising viewpoints are suggested during the volume exploration process based on what the user has already seen. Kim et al. [KUBS13] employed the Harris interest point detection algorithm to locate interest points as features in the volume, and found the optimal viewpoint by minimizing the occlusion between points by principal component analysis.

Generally, the optimal viewpoint for a given volume depends on intended applications. In medical applications, users are usually interested in small critical features, such as tumors and particular vessels in the head. Chan et al. [CQWZ06] defined a viewpoint selection framework for angiographic volumes. Visibility, coverage, and self-occlusion of features of interest are designed to search for the optimal viewpoint in the viewing sphere. The LiveSync interaction metaphor, proposed by Kohlmann et al. [KBKG07, KBKG08], synchronizes the 2D slice view with the volumetric view of medical datasets. All these images are 3D visualizations, and our approach can utilize these images to select the preferred viewpoints for similar volume datasets in specific applications.

2.3. Learning Based Methods

Recently, several viewpoint selection methods have been proposed based on data-driven learning in computer graphics. Laga and Nakajima [LN08] proposed a saliency learning based framework based on the light field descriptors for automatic viewpoint selection of 3D models. Intelligent design galleries proposed by Vieira et al. [VBP*09] train a classifier learned from the user interaction on viewpoints. This classifier is based on few users' viewpoint preferences and a large set of view descriptors, such as 2D image qualities and 3D feature visibilities. Secord et al. [SLF*11] conducted a large user study to collect the relative goodness of viewpoints based on

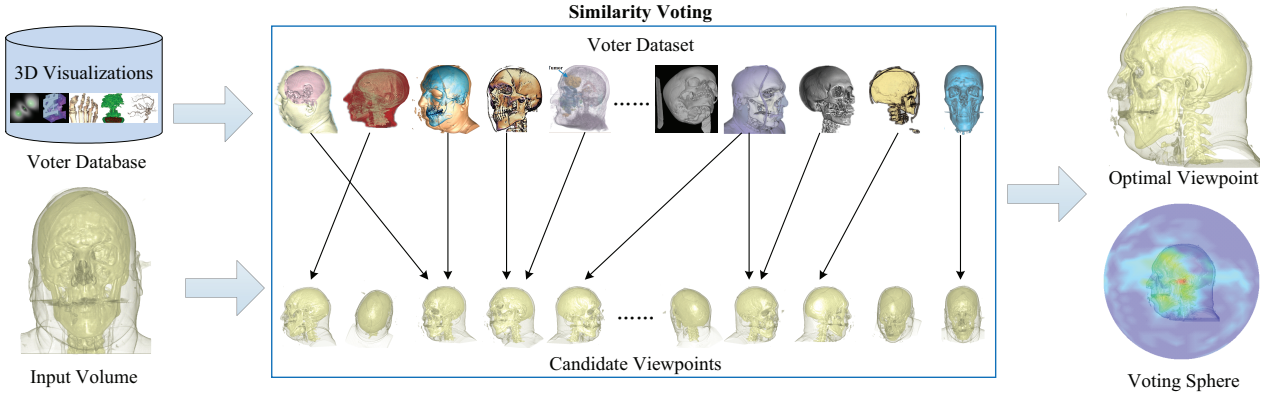


Figure 1: *Pipeline of the image-based viewpoint selection model. 3D visualizations are organized and stored in the voter database. The voter dataset with similar features in the input volume is chosen from the voter database, and each image casts votes to its most similar candidate viewpoints by similarity voting. A voting sphere records the votes of each viewpoint, and the optimal viewpoint is the one with the most votes.*

human preferences, and leveraged the result to train a linear model based on many previously proposed attributes of viewpoints. Liu et al. [LZH12] introduced a web-image voting method for viewpoint selection of 3D models. Each web-image votes on its most similar viewpoints based on the image similarity, and this performs better than previous view descriptors.

Our approach belongs to the data-driven learning based method. Previous methods largely use data produced by crowdsourcing or general users, such as users' preferences by user studies and images uploaded by general users. However, since visualization experts usually can provide more representative viewpoints for the volume investigated, this paper utilizes another data source, that is, 3D visualizations generated by visualization experts, to learn how they select viewpoints. To the best of our knowledge, our work is the first attempt so far to leverage 3D visualizations for viewpoint selection in volume visualization.

3. Image-Based Viewpoint Selection Model

Visualization experts generally choose the viewpoints carefully in their 3D visualizations to convey as much information as possible, to highlight important features, or to maximize some other metrics. 3D visualizations under these viewpoints are also often aesthetically pleasing. This paper attempts to extract the perceptual viewpoint information from the images generated by experts, and applies to select representative yet semantically meaningful viewpoints for relevant volumes. We propose an image-based viewpoint selection model. The inputs are a volume and relevant images with similar features in the volume. These images can be collected from many sources, such as published papers in the volume visualization literature and books with illustrations created by artists. The model learns from these relevant images by optimizing the best matching between relevant images and the rendered result of the volume from one viewpoint. Finally, it suggests an optimal viewpoint most similar to the one previously chosen by visualization experts.

The pipeline of our model is illustrated in Fig. 1. Given a volume as well as a transfer function, we first select one voter dataset, which contains similar features in the given volume, from the voter database. We leverage the images in the voter dataset to select the optimal viewpoint for the volume by similarity voting. Each image picks the most similar viewpoints by measuring the similarity between the image in the voter dataset and the rendered image of the volume from the viewpoint. We record the number of votes for each viewpoint in the voting sphere and suggest the optimal viewpoint with the most votes.

4. Similarity Voting

The voter dataset $D = \{I_1, I_2, ..., I_N\}$ is a collection of images, which have similar features in the given volume. I_i is the i-th voter image and N is the number of images. The volume can be observed from any viewpoint with arbitrary orientation. For simplicity, we sample the viewpoints on the viewing sphere [JS06] uniformly. All viewpoints are located on the surface of the viewing sphere, and the center of the viewing sphere coincides with the volume center. The viewing direction is from the viewpoint position to the volume center. These uniformly sampled viewpoints compose a set of candidate viewpoints as $V = \{v_1, v_2, ..., v_M\}$, where M is the number of viewpoints. We can generate an image for each viewpoint by volume rendering, and these images constitute the candidate dataset $C = \{R_1, R_2, ..., R_M\}$. Based on the image-based viewpoint selection model, each image in D will cast votes to its most similar images in C according to the image similarity measure.

The image similarity measure has been widely investigated in computer vision, especially image classification [LSP06] and image retrieval [PCI*07]. The well-known model is the bag of words (BoW) model [CDF*04], which is inspired by the success of BoW in text classification. The BoW model is based on the representation of affine invariant features extracted from local patches of an image. The most commonly used feature extraction methods are SIFT (Scale-Invariant Feature Transform) [Low04] and HOG (Histogram of Oriented Gradients) [DT05]. The codebook can be further generated by clustering these features, and each element is a visual word in the BoW model. Each feature is represented by the

code in the codebook, and all features in an image form a code vector. Based on the code vectors in the training dataset, an SVM with a nonlinear kernel can be used for image classification. Recently, researchers found that the spatial layout of features is also important for improving the performance of image recognition [LSP06].

In this paper, we focus on the similarity measure between images in the voter dataset and the candidate dataset. Generally, an image matches another image if their spatial shape and appearance distributions are both similar [BZM07, KSX14]. Therefore, we measure the image similarity from the similarity of the spatial shape and appearance distributions. The image similarity measure $s_{i,j}$ between the image I_i and the rendered image R_j generated from the viewpoint v_j is defined as follows

$$s_{i,j} = \alpha S_{shape}(I_i, R_j) + (1-\alpha) S_{appearance}(I_i, R_j), \qquad (1)$$

where S_{shape} and $S_{appearance}$ are the evaluation functions of the shape and appearance descriptors, respectively, and the parameter $\alpha \in [0,1]$ is the weight to control the trade off between the shape and appearance descriptors. The shape and appearance similarity values are normalized into [0, 1] separately before the combination.

The up-vector of the camera has a great influence on the spatial layout of shape and appearance of features. In this paper, we mainly focus on the viewpoint position. Thus, the influence of different up-vectors from the same viewpoint position should be reduced in the image similarity measure. Therefore, we need to transform the images both in the voter dataset and the candidate dataset before computing the similarity. Principal component analysis (PCA) is applied to positions of the pixels in the foreground of the image, which have a non-zero accumulated opacity. The origin of the transformed coordinate system is the central position of these pixels. The principle axis with the larger principle component is aligned with the y axis of the transformed coordinate system, while the other axis coincides with the x axis. Each image is rotated based on its principle axes, and resized so that the larger principle component becomes one. The resized image is cropped with the same size. In addition, features in the images of the voter dataset may have various colors, and these colors may be different from the one of features in the given transfer function. Thus, we cannot use the color information in the image similarity measure, and the shape and appearance descriptors are both evaluated on the intensity images.

4.1. Shape Descriptor

The shape information of features may differ significantly when observing the volume from distinct viewpoints. Since the images in the voter dataset contain similar features, we can use the shape difference to select the most similar viewpoints for each image in the voter dataset. Shape descriptors, such as the spatial distribution of edges, have been proved to be beneficial in image recognition [OPZ06], especially in the absence of color and texture information. Since edges in the image represent the geometrical characteristic of visible features, only edges are sufficient to discriminate different objects from each other when the boundaries of features are clearly rendered.

The image in the voter dataset may have different projective transformations, and the up-vectors of the camera are also approximately aligned. Compared to other shape descriptors, a Pyramid of Histograms of Orientation Gradients (PHOG) [BZM07] is less sensitive to small rotation and does not require strict spatial correspondence. Therefore, we chose PHOG as our shape descriptor, because PHOG is designed to handle varying degrees of spatial correspondence. In PHOG, the image is represented by its local shape and the spatial layout of its shape.

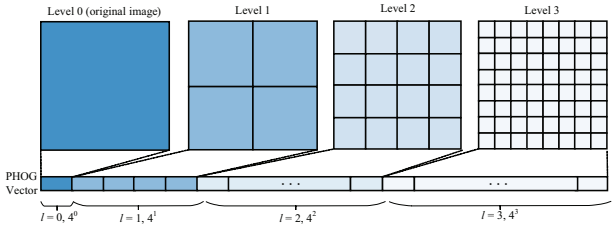


Figure 2: *A pyramid representation of the image and its PHOG vector.*

The local shape is defined by the distribution of edge orientations. Each image is first converted to an intensity image, and then the Canny edge detector is used to extract edge contours. The orientation gradient at each pixel on the edge contours is computed via a Sobel operator, and is discretized into K orientation bins, ranging from 10 to 80. The local shape corresponds to a histogram of edge orientations. For each contour point in the region, the contribution is weighted by its gradient magnitude. The generated histogram is a Histogram of Orientated Gradients, called a HOG vector, and can be considered a bag of words, where each word is a quantization on edge orientations.

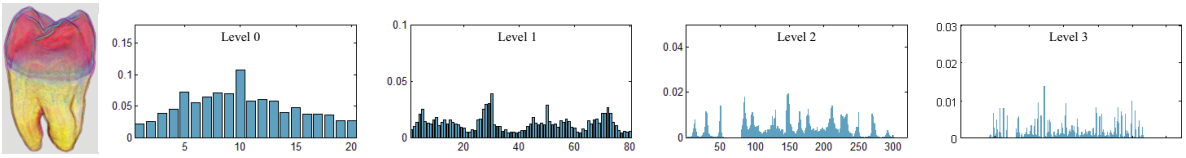
The spatial layout of its shape is captured by a pyramid representation of the image, as shown in Fig. 2. A HOG vector is computed for each region at each pyramid resolution level, and the PHOG descriptor of the image, namely the PHOG vector, is a concatenation of all HOG vectors, a vector with $K\sum_{l\in L} 4^l$ elements, where L is the number of levels and K is the dimension of the HOG vector (the number of bins in the histogram). In our implementation, the number of levels is limited to 3 ($L = 3$) to avoid over fitting. The PHOG vector is further normalized to avoid over-weighting images with more edges.

If the voter image I_i and the rendered image R_j are represented by the PHOG vectors H_i and H_j, respectively, the shape similarity between I_i and R_j can be computed by the PHOG similarity of H_i and H_j. The PHOG similarity is defined as

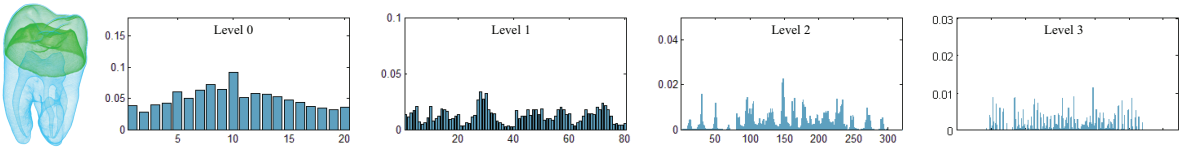
$$S_{shape}(I_i, R_j) = \sum_{l\in[1,L]} a_l d_l(H_i, H_j), \qquad (2)$$

where a_l is the weight at level l. The similarity between two vectors can be evaluated by many similarity metrics. Histogram intersection [LXW05] is widely used in image classification due to its simplicity and effectiveness. Thus, we use histogram intersection to compute the similarity between H_i and H_j at level l as

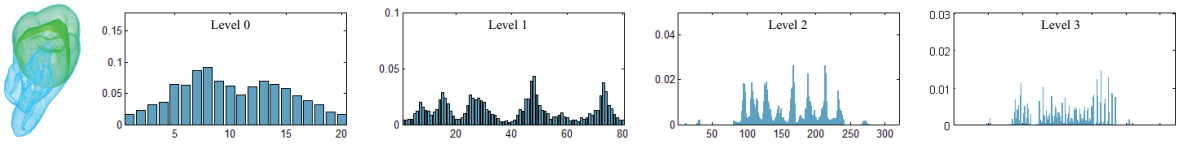
$$d_l(H_i, H_j) = \sum_{k=1}^{K4^l} min(H_i(k), H_j(k)). \qquad (3)$$



(a) A voter image and its HOG vectors.



(b) One rendered image and its HOG vectors.



(c) Another rendered image and its HOG vectors.

Figure 3: *Three images' HOG vectors of the tooth dataset. The histogram representations from levels $l = 0$ ($K = 20$) to $l = 3$ ($20 \times 4^3 = 1280$) for each image is shown from left to right, respectively. The histograms between (a) and (b) are more similar than the ones between (a) and (c), that is, the image in (b) shares more similar shape information with the voter image.*

Each level is generally weighted using $a_l = 1/2^{(L-l)}$, that is, histograms at finer resolutions are effective than those from coarser resolutions in the shape similarity measure [LSP06]. Level 0 is less distinctive for viewpoint selection, as it could not differ the front view from the back view for a symmetric feature. Thus, we only consider level 1 to level L in computing the similarity.

The distance between two PHOG descriptors shows the similarity of local shapes and their spatial layouts. A higher similarity between the images I_i and R_j means that the viewpoint v_j is more similar to the viewpoint in the image I_i. Fig. 3 shows three PHOG vectors ($L = 3$ and $K = 20$) for three images of the tooth dataset. The image in (a) is a voter image, and other images are two rendered images from different viewpoints. The shape similarity between the images in (a) and (b) is 0.84, while the shape similarity between the images in (a) and (c) is 0.70. Thus, the viewpoint in (b) is more similar to the one of the voter image, and this is consistent with human perception on observing these images.

4.2. Appearance Descriptor

Shape information alone is not sufficient to distinguish all types of images, especially with rich appearance information, such as textures. Thus, an appearance descriptor of the image is required to find a better matching. Similar to the shape descriptor, the image can be represented by its local appearance and the spatial layout of its appearance in the appearance descriptor.

SIFT is a widely used feature descriptor to detect and describe local patches in the image. Originally, SIFT descriptors are computed only at interest points. However, the recent comparative evaluation [LP05] shows that SIFT descriptors computed densely at a uniform grid work better than the one at sparse interest points for image classification. The dense appearance description is also necessary to describe low-contrast regions, such as the skin and bone. Therefore, this paper uses the dense SIFT descriptor to represent the local appearance.

SIFT descriptors of 16×16 pixel patches are computed at each pixel for each intensity image both in the voter dataset and the candidate dataset. In order to compare different SIFT descriptors effectively, a visual vocabulary is learned from the SIFT descriptors in the voter dataset by K-means clustering. Generally, a visual vocabulary consists of K visual words (K = 200 or K = 300). Each SIFT descriptor is quantized into a visual word in the visual vocabulary.

Similar to the shape descriptor, the spatial layout of its appearance is also captured by a pyramid presentation of the image, and the appearance descriptor used in this paper is called a Pyramid of SIFT (PSIFT) [LSP06]. A histogram of the visual words is computed for each region at each pyramid resolution level. Thus, level 0, the original image, is represented by a K-vector corresponding to the K visual words in the histogram. The appearance descriptor is a concatenation of all histogram, called the PSIFT vector. If the voter image I_i and the image R_j are represented by the PSIFT vectors S_i and S_j, respectively, the appearance similarity between I_i and R_j can be computed by the PSIFT similarity of S_i and S_j as

$$S_{appearance}(I_i, R_j) = \sum_{l \in [0,L]} a_l d_l(S_i, S_j), \tag{4}$$

where a_l is the same weight in Equation 2. Since the dense SIFT descriptor is used, the length of the SIFT vector at level 0 is larger than the length of the HOG vector at level 0. Thus, level 0 is included in computing the PSIFT descriptor, and the number of levels is 2 ($L = 2$) in our implementation. The similarity between S_i and S_j at

level l is also computed using histogram intersection in Equation 3. The PSIFT descriptor is a good description of approximate global appearance correspondence between two images.

For the three images in Fig. 3, the appearance similarity between the images in (a) and (b) is 0.71, while the appearance similarity between the images in (a) and (c) is 0.62. Thus, the viewpoint in (b) is also more similar to the one in the voter image from the appearance perspective.

4.3. Viewpoint Voting

Based on the image similarity measure above, each image I_i in the voter dataset casts votes to its most similar viewpoints as follows. The image similarity is first computed for each pair of the image I_i and each image in the candidate dataset. The images in the candidate dataset are sorted according to their similarity values with the image I_i. The image I_i selects its most similar images for voting, and the weight is the similarity value. In the implementation, the number of voted images is $log(|M|)$. Thus, the image I_i casts votes to the candidate viewpoints of its most similar $log(|M|)$ images. For example, if the image R_j is in the most similar images, the image I_i will vote to the viewpoint v_j with the weight $s(I_i, R_j)$. Using the similarity value as the weight, it can avoid the mis-vote by the images in the voter dataset, which have different features compared to features in the given volume.

After voting of each image in the voter dataset, we can summarize the voted weights to obtain the voting value for each candidate viewpoint in the viewing sphere. The viewpoints with a non-zero voting value are voted by the images in the voter dataset, and they are similar to the viewpoints selected by visualization experts for similar features. The optimal viewpoint is the one with the largest voting value, that is, the optimal viewpoint is the one selected by most images in the voter dataset. Besides the most voted viewpoint, we can also obtain several representative viewpoints by clustering viewpoints in the viewing sphere. Similar to the web-image voting [LZH12], mean-shift clustering can be applied to generate a group of clusters. These clusters are then sorted according to the voting value of the most voted viewpoint in each cluster. The representative viewpoints are selected from the most voted viewpoints or the cluster centers in sorted clusters.

5. Results and Discussion

5.1. Voter Dataset Construction

There are several large public volume collections, such as volvis[†] and The Volume Library[‡]. These volumes are widely used in the volume visualization research as experimented examples to validate the proposed method. In our experiments, the images in the voter database are collected from the volume visualization literature, especially visualization journals (TVCG, CGF, The Visual Computer) and conferences (IEEE Visualization, EuroVis, IEEE Pacific Vis, Volume Graphics) from 1999 to 2014. If 3D visualizations are available in other sources, we can also include them in the voter database for similarity voting.

Table 1: *Voter datasets and the number of images in each voter dataset*

Dataset	atom	engine	fish	foot	head	tooth	tree	vessel
Number	34	143	54	30	207	49	73	30

There are nearly 3K collected images, which are generated by volume rendering or illustrative visualization. Eight voter datasets are extracted from these collected images for experiments in this paper (about 2K collected images do not belong to these eight datasets). Table 1 lists eight voter datasets and their associated image numbers. The voter dataset of the head consists of several different volumes, and has 207 images, which display different features using various transfer functions under different viewpoints. Since one foot and two feet are significantly different, the foot voter dataset contains only images of one foot in our experiment, and the number of images is relatively small.

Our experiments were performed on a PC with 3.2 Hz Intel Core i5 CPU and 32GB memory. Each image is rotated and scaled according to its principal axes (0.2 seconds per image). The visual vocabulary for the appearance descriptor is trained for each voter dataset. The PHOG and PSIFT vectors of the images in the voter datasets are also computed in the preprocessing. Based on previous experiments on image classification, the number of bins in the histogram of the HOG vector is set to 20 and the number of the visual words in the visual vocabulary is 200 in our implementation. The preprocessing times of the PHOG vectors are 58.64 and 88.28 seconds for the engine and head voter datasets, respectively. The PSIFT preprocessing includes SIFT descriptor generation, dictionary construction and visual word quantization, and the times are 563.81 and 678.85 seconds for the engine and head voter datasets, respectively.

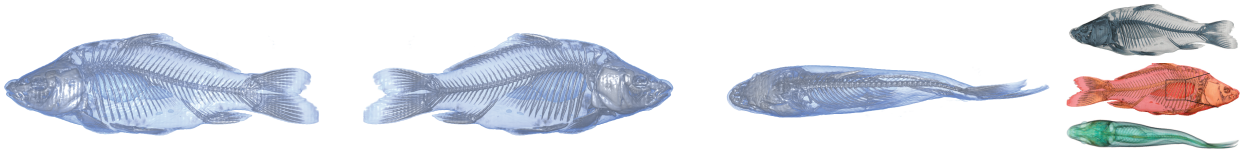
5.2. Viewpoint Selection Results

The HEALPix package [GHB*05] was used to generate uniformly distributed viewpoints on the viewing sphere. In our implementation, 1200 viewpoints were sampled on the viewing sphere to make voting more stable, and the size of the rendered image is 512×512. The weight α for the trade off the shape and appearance descriptors is 0.25 in our experiments, since the appearance is generally important in image recognition [BZM07].

We first perform an experiment to validate the feasibility of the image similarity measure in viewpoint selection using an NCAT phantom. 40 viewpoints in front and back of the phantom dataset were selected from the 1200 sampled viewpoints, and they are illustrated as points in Fig. 4(b) and (d). The rendered images from 40 viewpoints compose a voter dataset. Two voter images are displayed in Fig. 4(a) and (c). Each voter image casts votes to its most similar 10 rendered images out of 1200 viewpoints. The voting result is shown in Fig. 4(b) and (d). The color mapping from blue to red corresponds to the voting value from low to high. It is clear that the image similarity measure accurately selects viewpoints in front

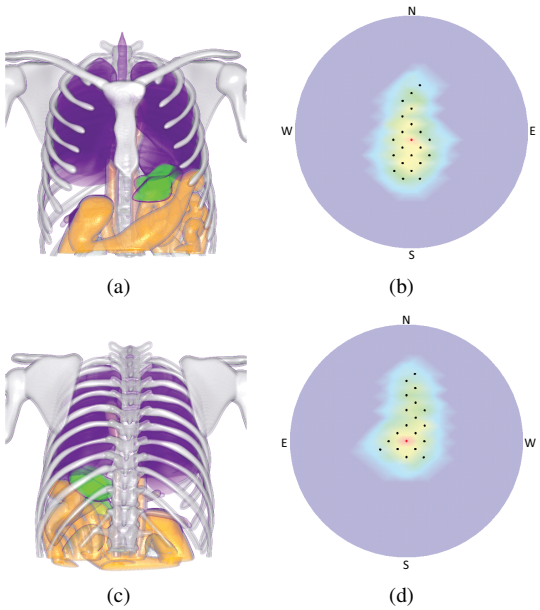
† http://volvis.org/

‡ http://www9.informatik.uni-erlangen.de/External/vollib/



(a) (b) (c) (d)

Figure 5: *(a)–(c) Three representative viewpoints of the carp dataset based on our method. (d) Three voter images with similar viewpoints from the voter dataset.*



(a) (b) (c) (d)

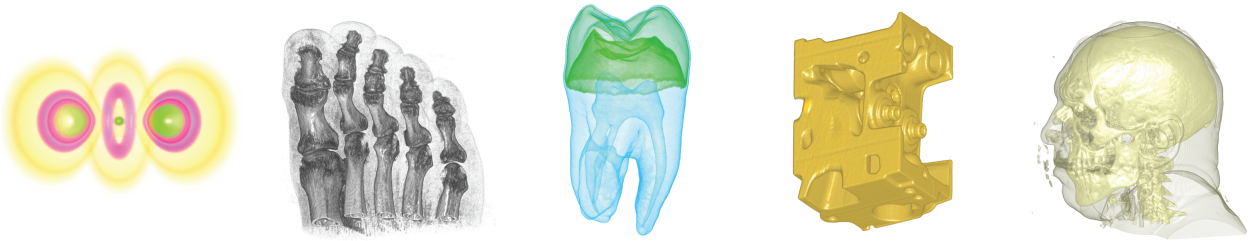
Figure 4: *The image similarity measure in viewpoint selection of the NCAT phantom. Two voter images of 40 selected viewpoints are shown in (a) and (c), respectively. The voting sphere is shown in (b) and (d). 40 viewpoints are displayed in black or red points, and the viewpoints of (a) and (c) are marked in red in (b) and (d), respectively.*

and back of the phantom. Original 40 viewpoints have higher voting values, and the largest value is 9. Voter images also cast votes to viewpoints around them. The larger the distance away from original viewpoints is, the smaller the vote is. Thus, our image similarity measure is less sensitive to small rotation and does not require strict viewpoint correspondence in similarity voting.

Three representative viewpoints for the carp volume selected by our method are shown in Fig. 5. These viewpoints are the most voted viewpoints in the sorted clusters. Most images select side viewpoints to avoid occlusion. As a result, our method selects two side viewpoints as the representative viewpoints. In addition, several images are visualized from the fish back to show the swimming pose of the fish. Thus, the viewpoint on the top is selected the third representative viewpoint. Fig. 5(d) gives three voter images similar to the three representative viewpoints, respectively.

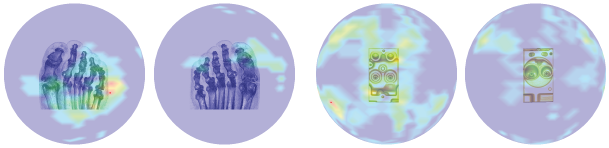
We also tested our method using other five datasets: the simulated electron density distribution in a hydrogen atom, the foot, the tooth, the engine, and the head of the visual male, as shown in Fig. 6. The voter dataset of the head contains several different volumes, and each volume reveals different features in its 3D visualizations. However, for other four voter datasets, there is only one volume in each voter dataset. The viewpoints in the atom, foot, and tooth voter datasets are relatively focused, as shown in the voting sphere in Fig. 7(a). Experts generally selected viewpoints with slightly oblique shift for the atom and foot datasets, which is consistent with aesthetic criterion. Thus, the optimal viewpoints selected by our method lean to the left and the right, respectively, as shown in Fig. 6(a) and Fig. 6(b). Similar to the opacity entropy [JS06], the viewpoint chosen for the tooth dataset shows the maximum information with a larger projection area. Inner structures of the tooth can be easily recognized in Fig. 6(c). The viewpoints in the engine voter dataset are quite diverse. As shown in Fig. 7(b), the most voted viewpoints are three-quarter views. The reason may be that the engine is not symmetrical, and different viewpoints reveal different aspects of the engine. As a result, there is no agreement in the optimal viewpoint for the engine. As shown in Fig. 6(d), the optimal viewpoint selected by our method is quite close to three-quarter views, which may be preferred by most users.

Head volumes are widely used in many published papers on volume visualization. There are 207 images in the voter dataset of the head, and the viewpoints are usually front views, side views, and three-quarter views. In our experiment, the transfer function is designed to display both the skin and bone, as shown in Fig. 6(e). The bones in the front view are visual clutter as shown in Fig. 1. The optimal viewpoint selected by our method is a side view and slightly over the ground, which agrees with what users choose for familiar features. About 17.4% voter images cast votes to this viewpoint. If we modify the transfer function to remove the skin, a different viewpoint selected by the same voter dataset is displayed in Fig. 9(a). We further use the Chapel Hill CT head to validate the effectiveness of our method using the same head voter dataset. Figure 8 shows three optimal viewpoints under different transfer functions, which show both the bone and skin (a), the bone only (b), and the skin only (c), respectively. Compared with the optimal viewpoints of the head of the visual male in Fig. 6(e) and Fig. 9(a), the optimal viewpoints are all side views, and therefore our method is relatively stable for similar features under different transfer functions. In addition, these optimal viewpoints are not the same viewpoint for different volumes and transfer functions. The optimal viewpoint is selected based on the similarity between



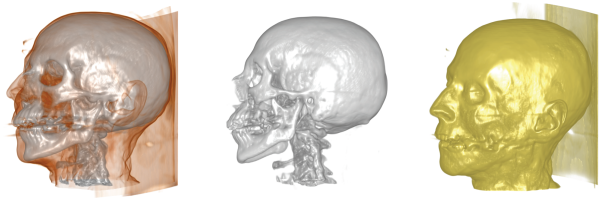
(a) (b) (c) (d) (e)

Figure 6: *The image-based viewpoint selection results of the simulated electron density distribution in a hydrogen atom, the foot, the tooth, the engine, and the head of the visual male from (a) to (e), respectively.*



(a) (b)

Figure 7: *The top and bottom views of the voting spheres of the foot (a) and the engine (b). The voted viewpoints are relatively focused in small regions for the foot, and the optimal viewpoint is indicated by the red point in Fig. 6(b). The voted viewpoints are quite diverse around the voting sphere for the engine, and the optimal viewpoint is indicated by the red point in Fig. 6(d).*



(a) (b) (c)

Figure 8: *The optimal viewpoints of the Chapel Hill CT head under different transfer functions, which highlight both the bone and skin (a), the bone only (b), and the skin only (c). The optimal viewpoints are slightly different under these transfer functions.*

the feature under the current transfer function and the feature in the voter images. Thus, our method can optimize the optimal viewpoint for different features.

We compare our method to the opacity entropy [JS06] and the shape-detail descriptor [TLB*09] using three datasets: the head of the visual male, the bonsai tree, and the aneurism, as shown in Fig. 9. The head and the bonsai tree have clear semantics meanings, especially the up direction. All selected viewpoints for the head dataset are side views. As shown in Fig. 9(a), the optimal viewpoint of our method is a low displacement under the ground, but other two methods select the viewpoint with a relatively large displacement under/over the ground, respectively. There are 73 images and at least three different volumes in the tree voter dataset. For the bonsai tree in Fig. 9(b), our method accurately recognizes the up direction, and the viewpoint is familiar to most users. Most images in the voter dataset have the same up direction as ours. Since the opacity entropy prefers an even opacity distribution and a larger projection area, the selected viewpoint is under the ground and hides most information of the bonsai tree. Similarly, the shape-detail descriptor prefers larger visible structures and local features as well as even relative orientation to the viewing direction. The viewpoint under the ground satisfies these requirements, and also has larger projection area. However, the viewpoint may not be acceptable in general applications.

For the aneurism, the users are generally interested in the aneurism, and avoid any occlusion to it. The images in the voter dataset reflect this preference, and the viewpoint selected by our method shows the aneurism more clearly. However, due to the lack of this semantic restriction, the viewpoints selected from other two methods both have occlusion to the aneurism, as shown in Fig. 9(c). Previous viewpoint selection methods generally depend on the predefined metrics of viewpoint quality. Currently, these metrics may be not complete for viewpoint selection in general applications, and it is also difficult to quantize some criteria on how users or experts select viewpoints. Our method directly learns how experts choose the viewpoint for similar features based on the image similarity, and it can better utilize implicit criteria in 3D visualizations, such as semantic information in different applications.

The computational time contains three parts: rendered image generation, PHOG similarity computation, and PSIFT similarity computation, and it largely depends on the number of candidate viewpoints and the number of images in the voter dataset. The times to generate all rendered images from the candidate viewpoints are 20 and 22 seconds for the $256 \times 256 \times 128$ engine and the $256 \times 256 \times 225$ head, respectively. The PHOG similarity computational times are 0.3 seconds and 0.4 seconds per image for the engine and the head, since the rendered images of the head have more edges. Most time is consumed in the PSIFT similarity computation (2.1 seconds per image for the engine and 2.2 seconds per image for the head), as each pixel in the rendered image requires to compute its SIFT descriptor, and be quantized into a visual word in the visual vocabulary. The efficiency can be further improved, since each rendered image is independent and can be parallel processing.

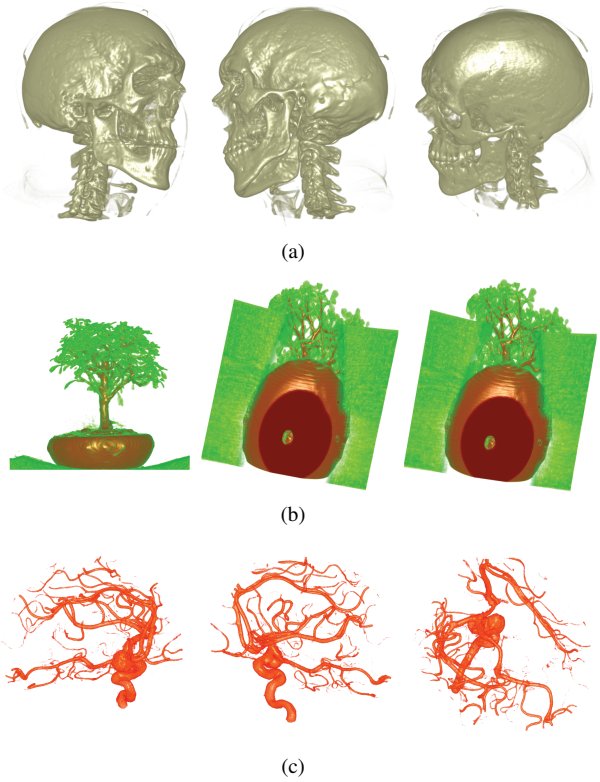


Figure 9: *Comparison of the optimal viewpoints for the head of the visual male, the bonsai tree, and the aneurism datasets obtained by three different methods: our method (left), the opacity entropy (middle), and structure-aware viewpoint selection (right).*

5.3. Discussion

When the given volume is rare in 3D visualizations, there will be few relevant images so that we cannot compute the optimal viewpoint for the volume. One possible solution is to ask domain experts or visualization experts to select a few optimal viewpoints, and these images are considered the voter dataset for this category of volumes. The required number of images generally depends on the complicity of interesting features. For example, if the feature is simple, about ten images are enough for similarity voting. After that, our method can select the optimal viewpoint as the initial exploration viewpoint for other volumes in this category.

In the meanwhile, since 3D visualizations are continually increasing each year, a new voter dataset can be constructed when enough images are discovered for another category. New images can also update previous voter datasets and be involved in similarity voting. Therefore, the optimal viewpoints of previous volumes will be further optimized with new images.

In the voter dataset construction, we manually classify each image into eight categories as shown in Table 1. The images not belonging to eight voter datasets are stored and can be classified into other voter datasets in future. Given a volume, we currently require the user to select the relevant voter dataset for similarity voting, which is a limitation of our method. Image classification techniques can be used to simplify this process. For example, we can use clustering methods to automatically group images into several categories. We can also train a classifier based on the images in the voter dataset using convolutional neural networks. Thus, only new images need to be further classified.

In the voter dataset, some "worst-case" viewpoints may exist in the collected images. We currently do not remove these images, since the number of these images is relatively small. For example, there are only three images in the "worst-case" viewpoints among 49 voter images in the tooth voter dataset. These "worst-case" viewpoints have little influence on viewpoint selection in our method, since we only assume that most of voter images are carefully selected by experts. In addition, it may be difficult to judge whether a viewpoint is the worst for different applications. For example, Fig. 5(c) may be considered a worst viewpoint according to some measures. However, this viewpoint can be also considered a representative viewpoint when exploring the carp dataset [ZAM11].

6. Conclusion

In this paper, we presented an image-based viewpoint selection model for volume visualization. Based on similarity voting, our model extracts and optimizes the viewpoint information from existing 3D visualizations with similar features in the volume. Each image/3D visualization tallies votes to the viewpoints with the best matching based on the image similarity measure, and the measure estimates the spatial shape and appearance similarity between the image and the rendered image from the viewpoint. Since 3D visualizations are generated by visualization experts and they are more familiar with the semantic meanings and functionalities of features, the viewpoint with the most votes from the images, that is, most used by visualization experts, is considered the optimal viewpoint for the volume. We collected images from the volume visualization literature and extracted several voter datasets from these images. Experiments on several volumes demonstrate that our method can select high-quality and semantically meaningful viewpoints.

We would like to improve the computational efficiency of similarity voting. Each image in the voter dataset requires to compute the similarity with the rendered image of each viewpoint. This can be accelerated by parallel processing such as GPU. Additionally, although our experiments only use 3D visualizations from visualization experts, our method can also incorporate them produced by general users or crowdsourcing. This would provide more images and voter datasets and extend the usefulness of our model in more applications.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments. This work was supported in part by National Natural Science Foundation of China No. 61232012, 61303141, 61472354 and the National Key Technology Research and Development Program of the Ministry of Science and Technology of China under Grant 2014BAK14B01.

References

[BRB*13] BRAMON R., RUIZ M., BARDERA A., BOADA I., FEIXAS M., SBERT M.: An information-theoretic observation channel for volume visualization. *Computer Graphics Forum 32*, 3pt4 (2013), 411–420. 2

[BS05] BORDOLOI U. D., SHEN H.-W.: View selection for volume rendering. In *Proceedings of the conference on Visualization'05* (2005), IEEE Computer Society, pp. 487–494. 1, 2

[BTB99] BLANZ V., TARR M. J., BÜLTHOFF H. H.: What object attributes determine canonical views? *Perception 28*, 5 (1999), 575–599. 1, 2

[BZM07] BOSCH A., ZISSERMAN A., MUNOZ X.: Representing shape with a spatial pyramid kernel. In *Proceedings of ACM International Conference on Image and Video Retrieval (CIVR'07)* (2007), pp. 401–408. 4, 6

[CDF*04] CSURKA G., DANCE C. R., FAN L., WILLAMOWSKI J., BRAY C.: Visual categorization with bags of keypoints. In *Proceedings of European Conference on Computer Vision (ECCV'04)* (2004), vol. 1, pp. 22–38. 3

[CMP14] CUI J., MA Z., POPESCU V.: Animated depth images for interactive remote visualization of time-varying data sets. *IEEE Transactions on Visualization and Computer Graphics 20*, 11 (Nov 2014), 1474–1489. 1

[CQWZ06] CHAN M.-Y., QU H.-M., WU Y.-C., ZHOU H.: Viewpoint selection for angiographic volume. In *Proceedings of the second International Symposium on Visual Computing'06* (2006), Springer-Verlag, pp. 528–537. 2

[DT05] DALAL N., TRIGGS B.: Histograms of oriented gradients for human detection. In *Proceedings of Computer Vision and Pattern Recognition (CVPR'05)* (June 2005), vol. 1, pp. 886–893 vol. 1. 3

[GHB*05] GÓRSKI K. M., HIVON E., BANDAY A. J., WANDELT B. D., HANSEN F. K., REINECKE M., BARTELMANN M.: Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal 622*, 2 (2005), 759. 6

[HS07] HONG Y., SHEN H.-W.: Parallel Reflective Symmetry Transformation for Volume Data. In *Eurographics Symposium on Parallel Graphics and Visualization* (2007), Favre J. M., Santos L. P., Reiners D., (Eds.), The Eurographics Association. 2

[JS06] JI G.-F., SHEN H.-W.: Dynamic view selection for time-varying volumes. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 1109–1116. 1, 2, 3, 7, 8

[KBKG07] KOHLMANN P., BRUCKNER S., KANITSAR A., GRÖLLER M. E.: Livesync: Deformed viewing spheres for knowledge-based navigation. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1544–1551. 2

[KBKG08] KOHLMANN P., BRUCKNER S., KANITSAR A., GRÖLLER M. E.: Livesync++: Enhancements of an interaction metaphor. In *Proceedings of Graphics Interface'08* (2008), Canadian Information Processing Society, pp. 81–88. 2

[KSX14] KIM G., SIGAL L., XING E.: Joint summarization of large-scale collections of web images and videos for storyline reconstruction. In *Proceedings of Computer Vision and Pattern Recognition (CVPR'14)* (June 2014), pp. 4225–4232. 4

[KUBS13] KIM H. S., UNAT D., BADEN S. B., SCHULZE J. P.: A new approach to interactive viewpoint selection for volume data sets. *Information Visualization 12* (July-October 2013), 240–256. 2

[LN08] LAGA H., NAKAJIMA M.: Supervised learning of salient 2d views of 3d models. *The Journal of the Society for Art and Science 7*, 4 (2008), 124–131. 2

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision 60*, 2 (2004), 91–110. 3

[LP05] LI F.-F., PERONA P.: A bayesian hierarchical model for learning natural scene categories. In *Proceedings of Computer Vision and Pattern Recognition (CVPR'05)* (2005), pp. 524–531. 5

[LSP06] LAZEBNIK S., SCHMID C., PONCE J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of Computer Vision and Pattern Recognition (CVPR'06)* (2006), vol. 2, pp. 2169–2178. 3, 4, 5

[LXW05] LEE S. M., XIN J. H., WESTLAND S.: Evaluation of image similarity by histogram intersection. *Color Research & Application 30*, 4 (2005), 265–274. 4

[LZH12] LIU H., ZHANG L., HUANG H.: Web-image driven best views of 3D shapes. *The Visual Computer 28*, 3 (2012), 279–287. 3, 6

[OPZ06] OPELT A., PINZ A., ZISSERMAN A.: Incremental learning of object detectors using a visual shape alphabet. In *Proceedings of Computer Vision and Pattern Recognition (CVPR'06)* (2006), pp. 3–10. 4

[PCI*07] PHILBIN J., CHUM O., ISARD M., SIVIC J., ZISSERMAN A.: Object retrieval with large vocabularies and fast spatial matching. In *Proceedings of Computer Vision and Pattern Recognition (CVPR'07)* (June 2007), pp. 1–8. 3

[RBFS10] RUIZ M., BOADA I., FEIXAS M., SBERT M.: Viewpoint information channel for illustrative volume rendering. *Computers & Graphics 34*, 4 (2010), 351–360. 2

[SLF*11] SECORD A., LU J., FINKELSTEIN A., SINGH M., NEALEN A.: Perceptual models of viewpoint preference. *ACM Transactions on Graphics 30*, 5 (Oct. 2011), 109:1–109:12. 2

[TCM10] TIKHONOVA A., CORREA C., MA K.-L.: Visualization by proxy: A novel framework for deferred interaction with volume data. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Nov 2010), 1551–1559. 1

[TFTN05] TAKAHASHI S., FUJISHIRO I., TAKESHIMA Y., NISHITA T.: A feature-driven approach to locating optimal viewpoints for volume visualization. In *Proceedings of the conference on Visualization'05* (2005), pp. 495–502. 1, 2

[TLB*09] TAO Y., LIN H., BAO H., DONG F., CLAPWORTHY G.: Structure-aware viewpoint selection for volume visualization. In *Proceedings of IEEE Pacific Visualization Symposium 2009* (April 2009), pp. 193–200. 1, 2, 8

[VBP*09] VIEIRA T., BORDIGNON A., PEIXOTO A., TAVARES G., LOPES H., VELHO L., LEWINER T.: Learning good views through intelligent galleries. *Computer Graphics Forum 28*, 2 (2009), 717–726. 2

[VFSH01] VÁZQUEZ P.-P., FEIXAS M., SBERT M., HEIDRICH W.: Viewpoint selection using view entropy. In *Proceedings of Vision Modeling and Visualization Conference (VMV'01)* (2001), pp. 273–280. 2

[Vio07] VIOLA I.: View selection in scientific visualization. Eurographics 2007 Tutorial #8 Applications of Information Theory to Computer Graphics, September 2007. 2

[VMN08] VÁZQUEZ P.-P., MONCLÚS E., NAVAZO I.: Representative views and paths for volume models. In *Proceedings of Smart Graphics* (2008), pp. 106–117. 2

[WQ07] WU Y., QU H.: Interactive transfer function design based on editing direct volume rendered images. *IEEE Transactions on Visualization and Computer Graphics 13*, 5 (Sept 2007), 1027–1040. 1

[ZAM11] ZHENG Z., AHMED N., MUELLER K.: iView: A feature clustering framework for suggesting informative views in volume visualization. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 1959–1968. 1, 2, 9

[ZW10] ZHANG Y., WANG B.: Optimal viewpoint selection for volume rendering based on shuffled frog leaping algorithm. In *Proceedings of IEEE Progress in Informatics and Computing 2010* (2010), vol. 2, pp. 706–709. 2