

3D Saliency from Eye Tracking with Tomography

Bo Ma, Eakta Jain, and Alireza Entezari, *Senior Member, IEEE*

Abstract— This paper presents a method to build a saliency map in a volumetric dataset using 3D eye tracking. Our approach acquires the saliency information from multiple views of a 3D dataset with an eye tracker and constructs the 3D saliency volume from the gathered 2D saliency information using a tomographic reconstruction algorithm. Our experiments, on a number of datasets, show the effectiveness of our approach in identifying salient 3D features that attract user's attention. The obtained 3D saliency volume provides importance information and can be used in various applications such as illustrative visualization.

Index Terms—Tomographic reconstruction, Eye Tracking, Saliency, Direct Volume Rendering

1 INTRODUCTION

Direct Volume Rendering (DVR) is commonly used for visualization of volumetric datasets generated by scanners in biomedical imaging (e.g. CT/MR) or by simulations in scientific computing. In DVR, a transfer function is used to classify features by assigning optical properties (i.e., color and opacity) to the scalar field. The visual appearance of various points along a viewing ray is influenced by local geometric attributes such as scalar values and its first and higher-order derivatives. However, as users' interests might be non-uniformly distributed spatially over the volume, traditional DVR can be ineffective in identifying regions of interest. For example, given a rendered image of human feet, doctors may be interested in the details of joints while non-expert users might be interested in the overall feet structure. Therefore, it is useful to understand where people focus in the 3D volume visualization and use such information to better highlight the regions of interest (ROI). In other words, we want to detect the saliency regions of volume datasets.

In this paper, we introduce an approach to detect the 3D saliency regions for a dataset based on the eye tracking data from multiple viewing angles that are assembled in 3D using tomographic reconstruction. As the eye tracking data is generated in 2D (i.e., image space), we collect user's saliency information from multiple projection images of the volume. Then, we locate the salient regions by constructing the 3D saliency volume from back projected 2D saliency maps. The produced 3D map assigns saliency values to the voxels of the original data, which is useful for various applications such as transfer function design and illustrative visualization.

2 RELATED WORK

Several approaches have been developed to compute visual saliency by analyzing the measurements derived from the data without visual feedback. Itti et al. [4] used an image processing technique, based on the center surround operators, to calculate the saliency map of a 2D image. Lee et al. [9] proposed a model of mesh saliency using center-surround filters with Gaussian-weighted curvatures. Kim et al. [8] provided an evaluation of the computational model of mesh saliency using eye tracking data. Other approaches located visual saliency in 3D space using eye tracking glasses [13, 14, 15]. They strove to estimate 3D gaze fixations based on the ray/object intersection method. The assumption of these approaches is the first object that is hit by the viewing ray is the target of the fixation. In contrast, in DVR users might focus on the interior structure while the viewing rays pass through the transparent exterior structure. Therefore, the simple ray/object intersection method cannot be applied to locate saliency in DVR images.

Eye tracking has been widely used in graphics and visualization for implicit data collection from users. Unlike traditional methods based on mouse clicks that will bring significant burden on interaction, eye tracking enables the analysis of users attention. Santella et al. [16] used a perceptual model together with eye tracking data to produce abstracted painterly renderings. Jain et al. [5] used eye tracking data to track readers' attention in comic book cartoons and assess the artist's success in directing the flow of attention. Burch et al. conducted experiments to evaluate traditional, orthogonal, and radial tree diagrams [2]. Participants were asked to find the least common ancestor of a given set of leaf nodes and eye tracking was used to record their exploration behavior. Eye tracking was also used in video re-editing to expose the important parts of the video [6]. Mantuk et al. [12] improved the accuracy of headmounted eyetracker in 3D scenes using both eye tracking data and prior knowledge of the environment. Instead of tracking in 2D image space, our work adds a new dimension by enabling the construction of 3D eye tracking profile that can be used for various purposes including saliency detection. The constructed saliency volume can be directly employed to the illustrative volume rendering [1, 18] which uses non-photorealistic rendering techniques to enhance important features or filter irrelevant details out. The saliency volume is also useful in other visualization applications, such as progressive visualization [11] and volume visualization enhancement [7].

Lu et al. [10] used eye tracking to identify the salient points on individual isosurfaces from a volume and used it for parameter selection in direct volume rendering. They visualized isosurfaces (with transparency) to track the user's attention to various regions on an individual isosurface while the volume is rotating. Users explore the features of the volume by selecting different isovalues while viewing the rotating volume. In this approach, a 3D focus point is identified by finding the intersection of two rays generated from a pair of fixation points from two frames corresponding to consecutive viewing angles. The 3D focus points can be located from fixation points on multiple consecutive frames if the user maintains viewing at the same 3D position. While this technique identifies salient regions on individual isosurfaces, it is challenging to track users attention among different isosurfaces specially since different isosurfaces are presented at different times during the exploration process. In our approach, we also collect the eye tracking data while the users view a rotating volume; however, we use tomographic reconstruction to construct 3D saliency map, for the entire classified volume, based on the acquired saliency on each viewing angle. We discuss the basic principles in tomographic reconstruction in Section 4.

3 SALIENCY INFORMATION ACQUISITION

Since the locating the ROI in a volume could be subjective and vary among users, we collect saliency information from users using an eye tracker. We are interested in two types of eye movement: fixations and saccades. Fixations are eye movements that stabilize the retina over a static object of interest. Saccades are rapid eye movements used in

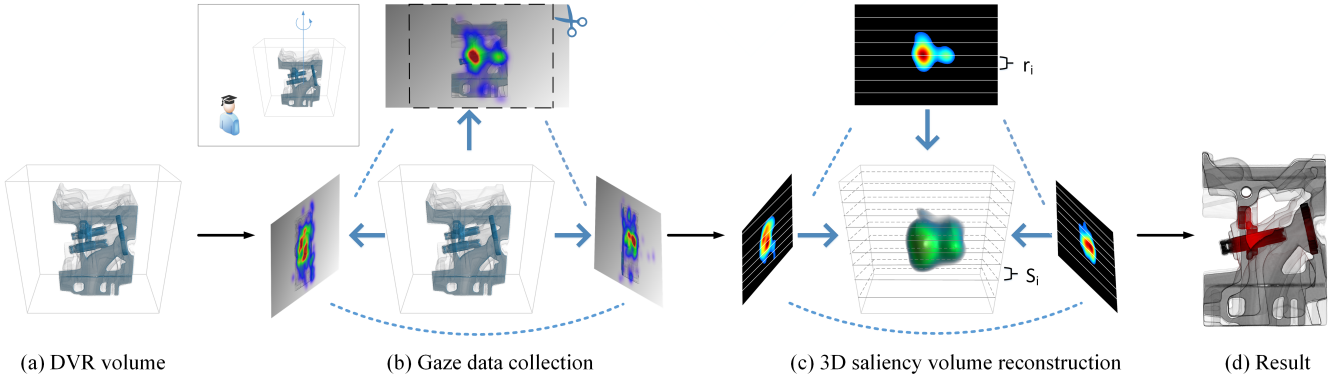


Fig. 1. The process of constructing the 3D saliency volume.

repositioning the fovea to a new location in the visual environment [3]. Fixations are the eye movements that best indicate the location of the viewer’s visual attention, thus identifying the fixations is the key to finding ROI. As locating the saliency regions in 3D requires additional information besides the 2D saliency information (i.e., image space), we leverage the 2D gaze data from multiple viewpoints

Initially, we render the volume (Fig. 1 (a)) based on a pre-defined transfer function. The transfer function can be specified manually or obtained from automatic transfer function design techniques [19, 17]. Obviously the choice of transfer function influences what features of the volume are visualized. Therefore, a proper transfer function is needed that can reveal the relevant features and details from the volume. While the choice of transfer function influences the visibility of features and impacts the eye tracking results, currently we assume a proper transfer function, with semi-transparent features, that reveal the relevant structures is given. We envision our 3D eye tracking system will be useful in providing feedback to the transfer function process (interactively or off-line). Next, as shown in Fig. 1 (b), we present the rotating volume to users so that they can freely view the interesting features. Under each viewpoint, we collect the gaze data which indicate the most attractive regions on the rendered image. The rotation axis is selected so that the respective projections can best reveal the features of the data. To collect enough gaze data, we rotate the volume 12 degrees every 4s which yield a set of 2D saliency information from 30 viewpoints (projections).

After the data collection, we analyze the fixations for each 2D projections image separately. We employ two off-line strategies for different requirement: 1. We analyze the fixations obtained from an individual, the gathered saliency information reveals the personal interest of the volume data. This is useful for a domain expert who knows the underlining data and desire to explore a particular part of the data. 2. We process the fixations from multiple users. For a volume, the users general interest is studied, and the regions that catch most of the attention can be identified. The generated 3D saliency volume can, in turn, inform the transfer function design in DVR (discuss in Section 6). With either strategy, we have collected a set of 2D fixations that are the sources of our saliency information.

4 SALIENCY VOLUME CONSTRUCTION

Tomography has been widely used in medical (e.g., X-ray CT) and other scientific fields, including physics, chemistry, astronomy. It allows for the reconstruction of an object from its projections (e.g., shadows). Modern tomography involves gathering projection data via scanning the object from multiple directions using a specific X-ray source. Then, the original object can be reconstructed by feeding the projections into tomographic reconstruction algorithms.

Fig. 2 shows an example of tomography that uses parallel beams to scan a 2D object $f(x,y)$ at a specific direction. For a given angle θ , the 1D projection of the 2D object is made up of a set of line integrals (shown as blue lines). The data collected at the sensor s_i is the line integrals of the beam b_i which represent the total attenuation of b_i as

it travels through the object. $p_\theta(s)$ is the 1D projection of the 2D object $f(x,y)$ which is formed by combining a set of line integrals at angle θ . The extent of $p_\theta(s)$ is determined by the bounding rectangle of the object. Mathematically, $p_\theta(s)$ can be understood as the Radon transform of f :

$$p_\theta(s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \delta(x \cos \theta + y \sin \theta - s) dx dy \quad (1)$$

From Fourier slice-projection theorem, we can reconstruct the original object $f(x,y)$ if we have projections $p_\theta(s)$ for many angles. The reconstruction is done by inverting the transformation in Eq (1) known as inverse Radon transform. In practice, a set of projections $p_\theta(s)$ from a finite set of directions can be obtained as outputs from a scanner where a rotating object is being scanned. These projections serve as the input for the tomographic reconstruction algorithms, such as filtered back projection and iterative reconstruction algorithms, that compute the inverse Radon transform and approximate the absorption density of the original object.

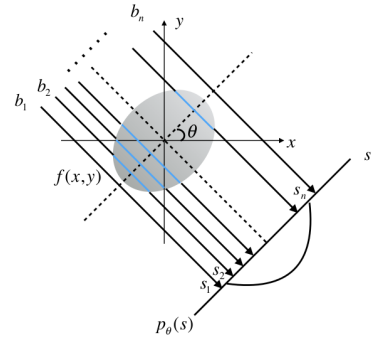


Fig. 2. Parallel beam tomography. Each projection is made up of the set of line integrals through the object

In our data acquisition process, we track the users viewing a rotating volume which is similar to the tomographic scan where rays are used to scan a rotating object. The saliency information are recorded as fixations in the rendered images that are generated from parallel projections of the volume. Therefore, the 2D saliency map in the image space can be back projected, for each projection angle, as the tomographic scans under parallel beam geometry. Once we collect the gaze data for a particular angle, we build the 2D saliency information for that angle which is then back projected to the 3D volume in the tomographic reconstruction algorithm. In our method, we use the commonly-used filtered-back-projection (FBP) to construct the 3D saliency volume from a set of 2D saliency maps. The FBP algorithm filters (i.e., Ram-Lak filter) each projection image before back projection – compensating for the concentration of information (samples) in

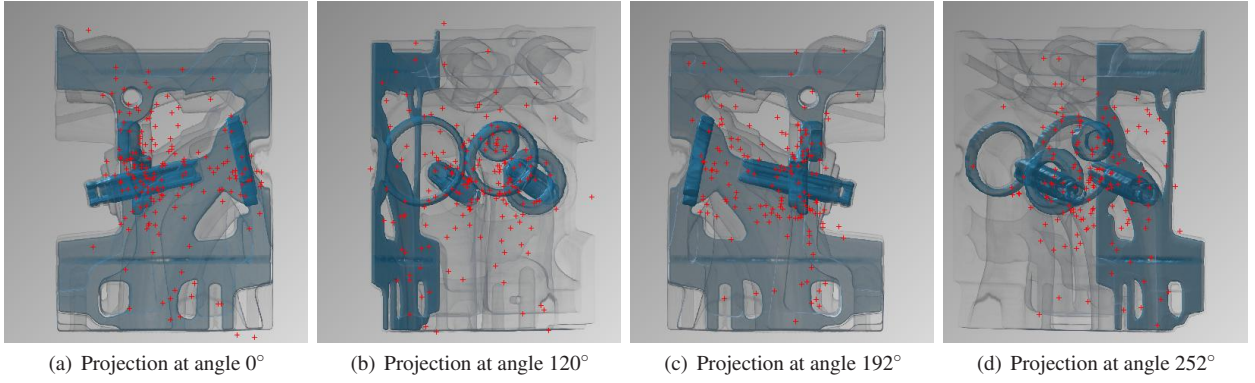


Fig. 3. Plot the fixation locations on the respective 2D projection images. The original images are 1920×1080 . The images shown here are zoomed in to see the fixations.

the center of the volume where all projections intersect and the lack of information away from the center. More accurate reconstructions can be obtained by using more projections that, in turn, increase the acquisition time. The necessary number of projection angles depend on the complexity of the 3D image that is being acquired. As discussed in Section 3, we have collected the fixations of 30 projection angles which provided a stable reconstruction, via FBP, of the 3D saliency map. Our preliminary investigation showed negligible improvements from increasing the number of projection angles.

The construction of 3D saliency volume involves the following three steps:

1. **Generate 2D saliency map:** we construct the 2D saliency map from gathered gaze data by convolving the map of fixation locations with a Gaussian kernel. We calibrated subjects to < 0.5 deg visual angle. We removed noise in the saliency map after data collection by thresholding saliency values $< 20\%$ of the maximum value. The size of the constructed saliency map is bounded to the size of the projection image which is shown with high resolution in order to locate user’s gaze data accurately. Therefore, the underlying saliency map contains redundant background pixels that need to be removed before being fed into the reconstruction algorithm. As an example shown in the top images of Fig. 1 (b) and (c), we crop the 2D saliency maps according to the bounding box of the volume which makes it easier to map the constructed 3D saliency volume to the original volume (step 3).
2. **Construct 3D saliency volume:** Fig. 1 (c) shows the process of constructing the 3D saliency volume slice by slice from the rows of the 30 saliency maps. Each slice S_i is located in 2D space and the respective rows r_i of saliency maps are the 1D tomographic projections. We employ the Matlab function `iradon` that uses the filtered back projection algorithm to compute the inverse Radon transform to construct each slice of the saliency volume.
3. **Resample 3D saliency volume:** Since the reconstructed volume from FBP does not necessarily align with the source volume, we use a resampling step to obtain saliency values for the voxels in the original volume.

After all the steps, we generate the 3D saliency volume that indicate the saliency of the voxels in the original volume.

5 EXPERIMENT

We conducted experiments on three volume datasets: an engine block, a CT human head and a carp fish. The volumes are rendered using a semi-automatic transfer function design approach that identifies a set of distinct representative isosurfaces from a dataset[17]. In order to understand users general interest to these data, 12 participants (9

male, 3 female, age from $21 \sim 28$) were recruited. The participants are campus students and were compensated with class credit for their participation.

Participants viewed the rotating volume on a 17-inch monitor (1920×1080). The rendered images were placed on the screen as large as possible to increase the accuracy of locating saliency regions. The participants were asked to adjust their chair height and distance to the monitor to make themselves comfortable. Then, the system was calibrated and the rotating volume was present to the participants.

For eye tracking, we used The Eye Tribe eye tracker with a cloud-based analytical platform. It can locate the gaze locations on the calibrated screen with less than 20 ms at 60 Hz recording rate. The raw data was converted into fixations and saccades inside EyeProof platform which can be exported as a spreadsheet for off-line analysis.

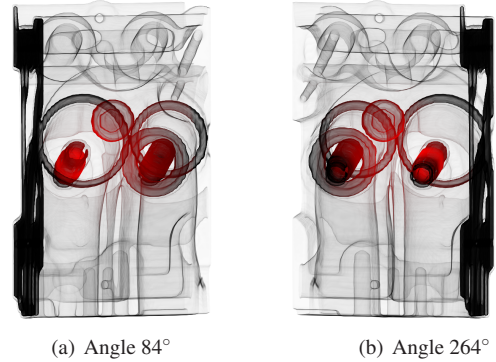


Fig. 4. Saliency volume visualization using red channel at different view angles: higher saliency voxels are assigned color with larger red value while the green and blue channels are absent.

In our first experiment, we show the effectiveness of our approach on a CT engine block ($256 \times 256 \times 128$). The structure of the engine is relatively simple, and we anticipated that users pay more attention to the two internal cylinders (see Fig. 1.(a)). After collecting the gaze data (see Fig. 1.(b)), we generated the fixation locations using EyeProof platform. Fig. 3 (a) shows some examples of overlaying the fixations (from 12 participants) on the respective projection images. A similar pattern can be observed from these images, that is fixations distributed densely at the cylinder region. Thus, the 2D saliency regions indicated by the fixations matches our assumption. Then, we convolved the fixations with the Gaussian kernel ($\sigma = 50$) to generate the 2D saliency maps (1920×1080). The saliency maps were thresholded and cropped to size 1531×1080 . It took 79s to construct the saliency volume ($1080 \times 1080 \times 1080$) on a standard PC. The constructed volume is further resampled to match the original volume. The middle

image of Fig. 1 (c) shows a DVR of the 3D saliency volume where saliency values from low to high are mapped to color blue, green and red. Comparing to Fig. 1 (c), the saliency volume generally have high saliency values at the internal cylinder region. Specifically, the left green ball corresponds to the cylinder parts camshaft and timing gears while the right green ball corresponds to the cylinder parts flywheel. Fig. 1 (d) and Fig. 4 displays with the visualizations of engine volume by modifying the color of the voxels based on the saliency volume. For each voxel, a higher red value is assigned when it has a larger saliency value while the blue and green channel are absent. We can clearly see that the internal cylinders are highlighted as red. The ROI indicated by the saliency volume matches our 2D saliency inspection, which further confirmed our assumption and also approve our reconstruction approach.

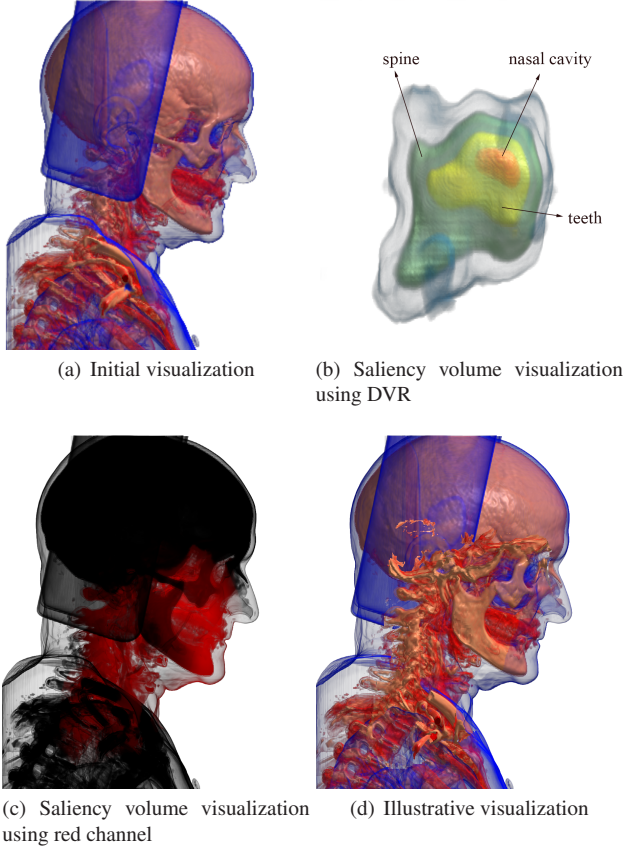


Fig. 5. Visualizations of CT human head at angle 270° : (a) Initial volume rendering based on a pre-defined transfer function. (b) Saliency volume visualization using DVR: saliency values from low to high are mapped to colors blue, green, yellow, red. (c) Saliency volume visualization using red channel: higher saliency voxels are assigned color with larger red value while the green and blue channels are absent. (d) Illustrative visualization: uninteresting regions (with low saliency values) are removed when they occlude interesting regions.

We conducted our second experiment on a more complex dataset: a CT human head ($256 \times 256 \times 230$). Fig. 5 (a) shows an initially rendered image where a number of features are presented, such as side props, skin, skull, ribs, spine, vessels, and teeth. The 2D saliency maps (1920×1080) were generated by smoothing the fixations with a Gaussian kernel ($\sigma = 50$), then were thresholded and cropped to resolution 1249×1080 . By feeding the 2D saliency maps into the reconstruction process, it takes 49s to construct the 3D saliency volume ($880 \times 880 \times 1080$) which is further resampled to match the original volume. Fig. 5 (b) shows the DVR of the generated 3D saliency volume where saliency values from low to high are mapped to colors

blue, green, yellow, red. The image indicates that instead of the big structures (e.g. skull and ribs) users are more interested in the internal structures such as nasal cavity (shown as red) and teeth (shown as yellow). More interestingly, the constructed 3D saliency volume models the shape (vertical stick) of the partial spine with high saliency which further confirmed the effectiveness of the reconstruction process. The saliency regions can be more clearly perceived in Fig. 5 (c) where higher saliency voxels are mapped to color with larger red value while the green and blue channels are absent. Fig. 5 (d) shows a simple illustrative volume rendering that makes use of the 3D saliency volume to remove less important parts of the volume to generate cut-away views. In the traditional DVR image (Fig. 5 (a)), the high saliency spine region is occluded by the low saliency side props and skin. In the illustrative rendering, the spine are expressed more clearly by removing part of the side props and skins that occlude it.

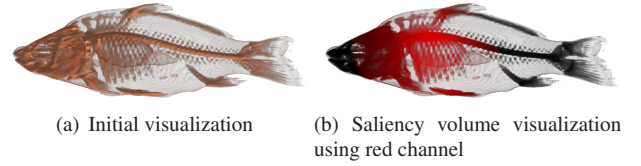


Fig. 6. Visualizations of carp fish at angle 0° : (a) Initial volume rendering. (b) Higher saliency voxels are assigned color with larger red value while the green and blue channels are absent.

To further validate our approach, we conducted another experiment on a carp fish dataset ($128 \times 128 \times 128$). As shown in Fig. 6 (a), the initial rendering solidly exposed the fish head while kept the internal lung and bones clearly visible. We were curious whether user's attention will be attracted to the prominently rendered fish head or the internal details. With a similar procedure, the reconstruction process took 31s to construct the saliency volume for this dataset. By visualizing the saliency volume using red channel, Fig. 6 (b) shows that the saliency regions are located at the front fish bone (bright red) while the fish head is dark with low saliency. This observation along with the above-mentioned experiments suggest that users tend to focus on internal structures of the volume even though an outstanding exterior structure is shown. But, we need more experiment to verify it.

6 DISCUSSION

We have presented an approach to construct the salient regions in a volume dataset. Our approach has shown to be effective in producing saliency volumes from the eye tracking data for various datasets. In our process, the saliency information is obtained from users by rendering the volume from multiple viewpoints. The constructed saliency volume can in turn customize the transfer function to better highlight regions of interest.

For the purpose of exploratory visualization, a current limitation of our approach is that the detected saliency regions heavily depend on the choice of transfer function. The volume rendering based on the transfer function determines the areas that catch more user's attention and different transfer functions might result in different saliency volumes. For example, in Fig. 5 (a), vessels and teeth were rendered with color red that were stand out easily. If we assign them another color such as the color similar to the skull, then users focus might switch to other regions. However, a visual stimuli is needed to obtain saliency information from users, and it make sense to have different saliency regions for different stimuli. In our future work, we plan to use the saliency volume in a feedback loop to evaluate the design of transfer functions. We plan to render a volume with various transfer functions that are to reveal same set of features. Then the constructed saliency volumes can be used to evaluate which transfer functions successfully attract users attention to the specified set of features. Further experiments could include the study of the differences between expert and non-expert viewers.

REFERENCES

- [1] S. Bruckner and E. Gröller. Style transfer functions for illustrative volume rendering. *Computer Graphics Forum*, 26(3):715–724, 2007.
- [2] M. Burch, N. Konevtsova, J. Heinrich, M. Hoferlin, and D. Weiskopf. Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2440–2448, Dec 2011.
- [3] A. Duchowski. *Eye tracking methodology: Theory and practice*, volume 373. Springer Science & Business Media, 2007.
- [4] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(11):1254–1259, Nov 1998.
- [5] E. Jain, Y. Sheikh, and J. Hodgins. Inferring artistic intention in comic art through viewer gaze. In *Proceedings of the ACM Symposium on Applied Perception*, SAP ’12, pages 55–62, New York, NY, USA, 2012. ACM.
- [6] E. Jain, Y. Sheikh, A. Shamir, and J. Hodgins. Gaze-driven video re-editing. *ACM Trans. Graph.*, 34(2):21:1–21:12, Mar. 2015.
- [7] Y. Kim and A. Varshney. Saliency-guided enhancement for volume visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):925–932, Sept 2006.
- [8] Y. Kim, A. Varshney, D. W. Jacobs, and F. Guimbretière. Mesh saliency and human eye fixations. *ACM Trans. Appl. Percept.*, 7(2):12:1–12:13, Feb. 2010.
- [9] C. H. Lee, A. Varshney, and D. W. Jacobs. Mesh saliency. In *ACM SIGGRAPH 2005 Papers*, SIGGRAPH ’05, pages 659–666, New York, NY, USA, 2005. ACM.
- [10] A. Lu, R. Maciejewski, and D. S. Ebert. Volume composition and evaluation using eye-tracking data. *ACM Trans. Appl. Percept.*, 7(1):4:1–4:20, Jan. 2010.
- [11] R. Machiraju, J. E. Fowler, D. Thompson, B. Soni, and W. Schroeder. Evitaefficient visualization and interrogation of tera-scale data. In *Data mining for scientific and engineering applications*, pages 257–279. Springer, 2001.
- [12] R. Mantiuk, B. Bazyluk, and R. K. Mantiuk. Gaze-driven object tracking for real time rendering. *Computer Graphics Forum*, 32(2):163–173, 2013.
- [13] L. Paletta, K. Santner, G. Fritz, H. Mayer, and J. Schrammel. 3d attention: measurement of visual saliency using eye tracking glasses. In *CHI’13 Extended Abstracts on Human Factors in Computing Systems*, pages 199–204. ACM, 2013.
- [14] T. Pfeiffer. Measuring and visualizing attention in space with 3d attention volumes. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA ’12, pages 29–36, New York, NY, USA, 2012. ACM.
- [15] T. Pfeiffer and P. Renner. Eyesee3d: A low-cost approach for analyzing mobile 3d eye tracking data using computer vision and augmented reality technology. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA ’14, pages 195–202, New York, NY, USA, 2014. ACM.
- [16] A. Santella and D. DeCarlo. Abstracted painterly renderings using eye-tracking data. In *Proceedings of the 2Nd International Symposium on Non-photorealistic Animation and Rendering*, NPAR ’02, pages 75–ff, New York, NY, USA, 2002. ACM.
- [17] S. Suter, B. Ma, and A. Entezari. Visual analysis of 3d data by isovalue clustering. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, R. McMahhan, J. Jerald, H. Zhang, S. Drucker, C. Kambhamettu, M. El Choubassi, Z. Deng, and M. Carlson, editors, *Advances in Visual Computing*, volume 8887 of *Lecture Notes in Computer Science*, pages 313–322. Springer International Publishing, 2014.
- [18] I. Viola, A. Kanitsar, and M. E. Groller. Importance-driven volume rendering. In *Proceedings of the conference on Visualization’04*, pages 139–146. IEEE Computer Society, 2004.
- [19] J. Zhou and M. Takatsuka. Automatic transfer function generation using contour tree controlled residue flow model and color harmonics. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1481–1488, Nov 2009.