Gaze-Based Annotations: Labels on Demand

P. Saalfeld[1, 2], A. Mewes[2], C. Hansen[2], B. Preim[1]

[1] Otto-von-Guericke University, Visualization Group, Magdeburg, Germany
[2] Otto-von-Guericke University, Computer-Assisted Surgery Group, Magdeburg, Germany

Contact: saalfeld@isg.cs.uni-magdeburg.de

Abstract:

We present an approach that tracks the gaze position of a user to determine a structure of interest in a medical planning model. This structure is automatically and dynamically annotated with an external label. The approach considers aspects from hand-crafted illustrations and interactive applications. In general, labels are simultaneously shown for all structures in interactive applications. However, this leads to visual clutter and is costly regarding computation time. Both aspects could be neglected if annotations are only shown regarding user's demands, i.e., only one label at a time. We use an eye tracker to obtain the user's gaze positions and, thus, the structure of interest. We address problems due to imprecise input with smoothing and increasing the selection area around each structure. The prototype was evaluated with an unstructured interview, which confirmed the suitability of our approach, e.g., during interventions where surgeons benefit from sterile interaction.

Keywords: Labels, Annotation, Gaze-based, Medical Imaging

1 Problem

Labeling is the process of annotating structures with textual annotations [1] and has its origins in cartography [2]. Nowadays, annotations exist in a wide range of domains and are commonly used in the medical area. Here, they enrich medical structures in anatomy atlases and medical textbooks to primarily support education. These hand-crafted static illustrations follow rules to support legibility and clarity, such as *clear graphic association* and *overlapping avoidance* [2].
The same rules are applied in interactive applications, where automatic label layouts are calculated with heuristics to address the positioning problem which is not deterministically computable in polynomial time (NP). Beneath rules from static visualization, *frame coherency* and *interactive calculation rates* are important aspects in interactive labeling applications [3, 4, 5]. Many automatic placement methods, such as [6], neglect that interactive applications do not have to show all labels at the same time. Each annotation can be shown and hidden based on user's demands. By showing only one label at a time, the space to position an annotation is larger, i.e., a label can be placed more freely in empty space and also occlude other structures, which are currently not interesting. Furthermore, the user or layout algorithm does not have to take care of crossing connection lines or overlapping labels. Medical applications, such as treatment planning, image-guided interventions, medical training and education benefit from the reduced visual clutter.
The challenge is to identify the structure of interest in a *reliable* (in terms of user intention), *fast* and *robust* manner. Beneath using mouse interaction to manually set labels beneath structures, such as in the Voxel-Man project (VOXEL-MAN, Hamburg, Germany), other hands-free input modalities are possible and efficient [7]. With the rise of affordable eye tracking hardware, gaze-based selection of structures is a promising approach. The gaze position delivers not only a reliable information about the user's structure of interest, but it is also fast compared to other selection techniques, such as mouse-based selection [8]. Due to tremble of our eyes and inaccuracies caused by the tracking hardware, focusing small objects, e.g., thin vessels or small tumors, in a robust way is challenging. Other work in the medical domain uses eye tracking to analyze what surgeons see during interventions [9]. Here, the gaze-based annotations can supply surgeons with a hands-free and thus, sterile interaction to show additional information such as volumes of tumors or radii of vessels.
According to the taxonomy of [1], we visualize surfaces of arbitrary shapes, where less ambiguity is desired. Their guidelines suggest to use external labels, i.e., textual boxes placed in free space connected with a line to an anchor point of the corresponding structure. This labeling technique is appropriate due to two reasons: external labels prevent occlusion of the connected structure and the disadvantages caused by solving complex computational layout problems are dissolved by showing only one label. An additional advantage is the possibility to handle longer multiline texts.
This work introduces a gaze-based labeling approach to annotate structures of current interest. Here, problems caused by imprecise input are addressed with two approaches: first, smoothing the gaze positions and secondly, creating structure-

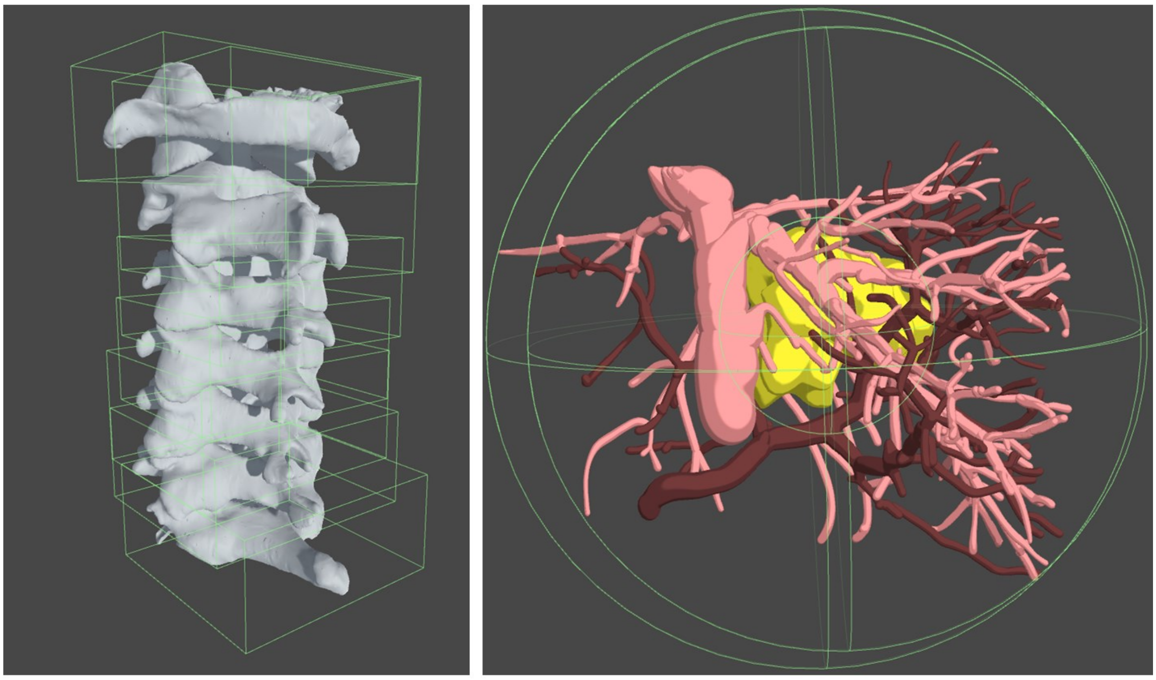
specific larger bounding objects to select a structure. The prototype was evaluated with an unstructured interview, which confirms the usefulness during interventions.

2 Materials and Methods

This section gives an overview about the hardware used for eye tracking, gaze-based selection, and describes the label placement as well as visualization. We evaluated the prototype with an unstructured interview. We stated questions regarding the general usability of eye tracking for structure selection as well as questions about advantages and drawbacks while interacting with the prototype. The results are described in the results section (Sec. 3).

2.1 Technical Setup

The commercial eye tracking system Tobii EyeX (Tobii AB, Danderyd, Sweden) is used to track the user's gaze point on the screen. The eye tracker incorporates near-infrared microprojectors, optical sensors and image processing. The microprojectors create reflection patterns on the eyes, which are registered in real time by the image sensors together with the user's head and eyes. The user's features and the projected patterns on his eyes are obtained with image processing. Finally, the eye positions and the gaze point are calculated. The visual angle error of the Tobii EyeX tracker ranges below 0.5°. The working distance is between 45 and 80 cm. At 70 cm distance, the head of the user can be detected on a 48x39 cm plane with a tracking frequency of 30 Hz [10]. Our prototype is implemented in the game engine Unity (Unity Technologies, San Francisco, USA). The API of the eye tracking system provides an interface for Unity such that the user's gaze point can be assessed.

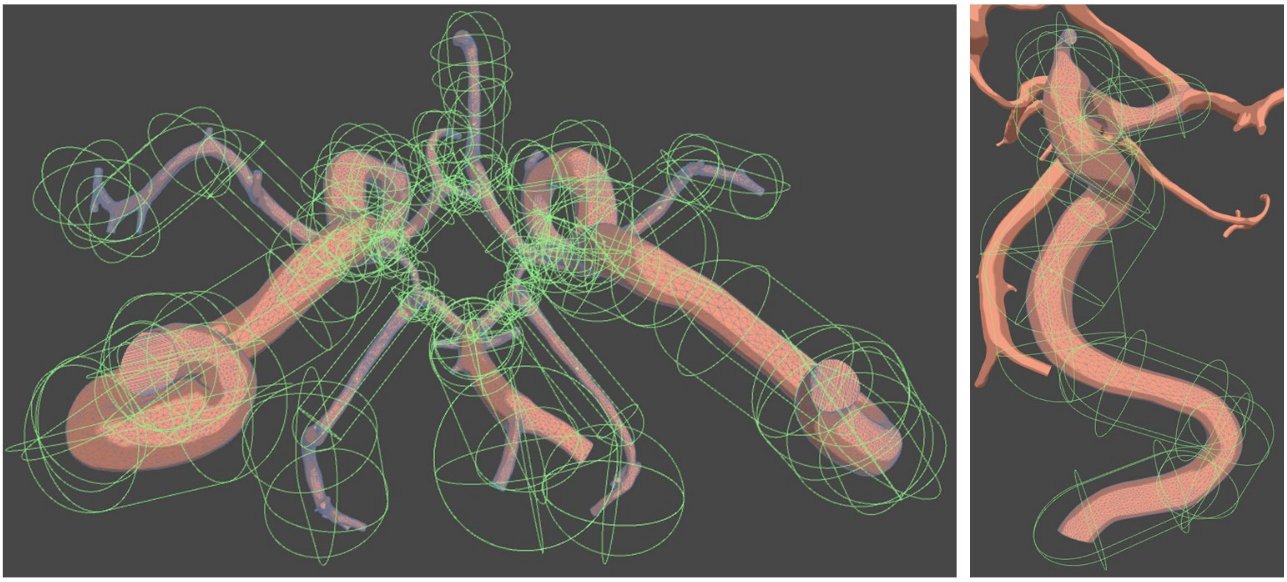


(a) (b)

Figure 1: Bounding primitives are appropriate to separate compact and non-interweaving structures, such as vertebras (a) but are insufficient for overlapping or thin, elongated objects. (b) shows overlapping bounding spheres for hepatic vessel systems, i.e., the portal vein (dark red) and hepatic vein (light red).

2.2 Gaze-Based Structure Selection

The gaze point is calculated by tracking the eye position and the viewing direction of the user. The intersection between this vector and the display can be calculated to obtain the gaze point in screen coordinates. Due to the limited tracking accuracy and trembling of the eyes, the gaze point is unsteady and varies within a range of several pixels. The noisy input is processed with *exponential smoothing,* an approach to forecast or, in our case, smooth the received gaze data. Here, the weight of past observations is decreased over time, i.e., older gaze points have a lower weight. Yet, some structures, such as vessels, are thin and it is hard to set a focus on them. Therefore, an area around each object is necessary to support the user to fixate it. The convex hull or larger bounding primitives are only appropriate for compact structures, such as vertebras, but insufficient for smaller vessels (see Fig. 1). Therefore, we manually build a bounding object around each structure by connecting multiple primitives (see Fig. 2).



(a) (b)

Figure 2: Illustration of a more complex bounding object around each selectable structure by combining multiple bounding primitives. (a) shows 16 structures of the Circle of Willis and (b) the Carotis interna, a single selectable structure.

Although this approach allows selecting objects in a wider range, it still leads to ambiguities w.r.t. overlapping objects or nearby structure transitions. If more than one bounding object is hit by the gaze point, the structure center points are compared with the gaze point. The object with the shortest distance is selected.

2.2.2 Label and Structure Visualization

After the user focused on a structure, the label is faded in on the described position. The label remains visible as long as the gaze point lies on:

- the structure,
- the boundary object,
- the label or
- an empty area.

If the user looks at another structure, the old label is quickly faded out (200ms) and the new one appears. It is important to hide the old label as unobtrusively as possible. Otherwise, the attention and thus, the gaze point could be drawn to the old label.

The label bounds are dynamically calculated based on the contained text, which is wrapped on whitespaces. To provide a clear view of the text and to show objects behind the label, the label boundary is transparently filled. Depending on the placement, the text inside the label is aligned on the left or right side. The structure of interest is highlighted with a silhouette (see Fig. 4).

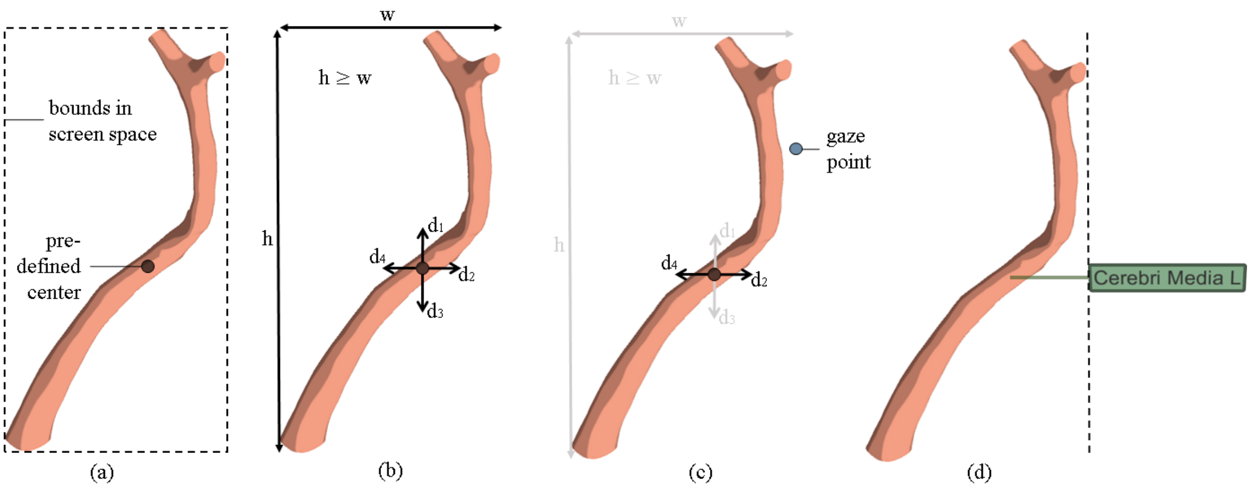


Figure 3: Based on the screen space bounds and the pre-defined center of an object (a) the direction of the label is calculated. The width and height are compared to decide if a label is placed vertically (w>h) or horizontally (h>=w) (b-c). The entering position of the gaze point is used to decide if the label is placed on the left or on the right (d). The exact position of the label is determined by intersecting the chosen direction with the screen space bound (d).

3 Results

The result of our work is a prototype to allow an interactive gaze-based selection of structures of a medical planning model with arbitrary shapes. Selected structures are annotated with an external label. The approach considers aspects such as clarity as well as prevention of occlusions. The method is easy and fast regarding computation, which allows interactive frame rates. External labels dynamically adapt their size depending on the content and allow a frame-coherent interaction.

We assessed the quality of our prototype qualitatively with an unstructured interview. It was carried out with an expert in the medical application domain, who has eleven years of experience with medical applications and three years with annotations. The expert confirmed the usefulness of gaze-based structure selection as well as intuitive interaction and pointed out the following difficulties and possibilities to improve the approach:

- In the state of a first exploration, only one label would be insufficient. Showing all labels at the beginning would allow the user to get an overview over the number of structures and its names. This follows Shneiderman's Information-Seeking Mantra: overview first, details-on-demand [11].
- The external labels are sometimes too far away. Then, it is disturbing to first focus on the structure and afterwards on the label.
- If there are objects between the first and second structure of interest, labels for all structures in-between are shown due to the gaze point causing intersections. This leads to visual clutter.
- The connection line is not necessary because the structure of interest is highlighted and only one label is shown.

Beneath these points of criticism, the expert confirmed the usefulness of the approach in sterile environments and suggested possibilities to improve it, which are described in the next section (Sec. 4).

4 Discussion

Some aspects of our approach could be automatized and improved. The pre-defined center of each structure could be calculated in image space with a distance transformation. This morphological operator gives the most inner point of a binary image and, thus, a valid anchor point [4]. By applying this operator outwards, a silhouette of each object with an adjustable distance can be created. This silhouette would only grow in free space and stop if other structures where hit, which prevents overlapping. This is useful for two things: first, the gaze point can be checked against the obtained silhouettes to determine a corresponding object. With this, the time-consuming step to manually build a larger bounding object can be omitted. Secondly, the margin of the silhouette can also be used to get a label position. This position could be closer to the object and, thus, would address feedback from the interview. However, using this image-based approach has some disadvantages. Small structures, which are surrounded by others, would be hard to focus on, because the out-

ward distance transform would not create a margin. A second problem arises with anchor points. During interaction, visible structures could be partially occluded, which would result in anchor points that abruptly change positions.
The feedback from the interview included a possibility to improve the prototype. The problem with selecting objects which are in between two structures of interest could be addressed with the following two approaches. First, labels are only shown if the user's gaze position intersects it for a short but certain amount of time. Secondly, the speed of the changing eye position could be used as an indicator for the intent of the user.

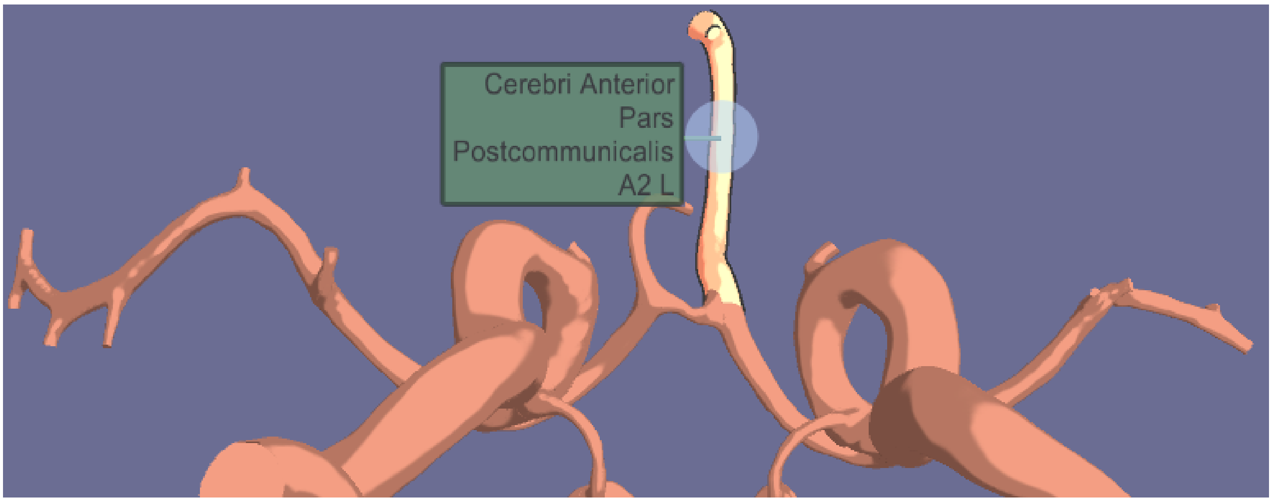


Figure 4: The gaze area is represented by a transparent circle. The selected structure is highlighted by showing its silhouette and a brighter color.

5 Conclusion

We presented a prototype which allows gaze-based interaction to show labels regarding user demands. Our approach runs at interactive frame rates and is frame-coherent. The label position is determined with legibility and clarity in mind, so that an unambiguous structure-to-label relationship is visualized and occlusion and overlapping is prevented. The gaze-based annotation approach would be suitable for environments where sterility is important. Therefore, an evaluation in an interventional setting, e.g., during a surgery, is our next step. Here, it could be tested if the eye tracking would be reliable enough for focusing on a structure during an intervention.

6 Acknowledgements

This work was partly funded by the German Federal Ministry of Education and Research (BMBF) within the research campus STIMULATE under grant number '13GW0095A'. We would like to thank *Dornheim Medical Images* for providing us with the software *Dornheim Segmenter*.

7 References

[1] Oeltze-Jafra, S., Preim, B., Survey of Labeling Techniques in Medical Visualizations, Eurographics Workshop on Visual Computing for Biology and Medicine (VCBM), Austria (2014)
[2] Imhof, E., Positioning names on maps, Cartography and Geographic Information Science, **2(2)** 128-144 (1975)
[3] Preim, B., Ritter, A., Strothotte, T., Pohle, T., Forsey, D.R., Bartram, L., Consistency of Rendered Images and Their Textual Labels, Proc. of CompuGraphics, Portugal (1995)
[4] Ali, K., Hartmann, K., Strothotte, T., Label Layout for Interactive 3D Illustrations, Journal of the WSCG, **13(1)** 1-8 (2005)
[5] Hartmann, K., Götzelmann, T., Ali, K., Strothotte, T., Metrics for Functional and Aesthetic Label Layouts, Smart Graphics (SG), Germany (2005)
[6] Mühler, K., Preim, B., Automatic Textual Annotation for Surgical Planning, Proc. of Vision Modeling and Visualization (VMV), Germany (2009)

[7] Stellmach, S., Dachselt, R., Designing gaze-based user interfaces for steering in virtual environments. Proc. of the Symposium on Eye Tracking Research and Applications (ETRA), USA (2012)
[8] Sibert, L. E., Jacob, R. J. K., Evaluation of eye gaze interaction, Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI), USA (2000)
[9] Atkins, M.S., Tien, G., Khan, R.S., Meneghetti, A., Zheng, B., What do surgeons see: capturing and synchronizing eye gaze for surgery applications, Surg. Innov., **20(3)** 241-248 (2013)
[10] Velloso, E., Turner, J., Alexander, J., Bulling, A., Gellersen, H., An Empirical Investigation of Gaze Selection in Mid-Air Gestural 3D Manipulation, Conference on Human-Computer Interaction, Germany (2015)
[11] Shneiderman, B., The eyes have it: a task by data type taxonomy for information visualizations, IEEE Symposium on Visual Languages, USA (1996)