MODELAR: A MODular and EvaLuative framework to improve surgical Augmented Reality visualization

G. Hattab[†1], F. Meyer[1], R. D. Albrecht[1], and S. Speidel[1]

[1]Division of Translational Surgical Oncology (TSO), National Center for Tumor Diseases (NCT) Partner site Dresden, Germany

Abstract

The use of Augmented Reality (AR) for the visualization of 3D biomedical image data is possible thanks to a growing number of hardware and software solutions. Considerable efforts are made during surgery, where the visual information of the target structures can either be highlighted or dulled. However, as technical challenges and barriers to development decrease, it's increasingly important to take into account the specific capacities and constraints of the surgeon's perceptual and cognitive systems. To address this legitimate problem, we present a practical framework that evaluates the importance of visual encodings and renderings for surgical AR. By conducting a task-specific user study we observed a set of emerging visualization strategies. The given task is to make the kidney boundary visually salient and make the tumor and calyx distinguishable. After having recruited 23 participants, we found two preferred presets to tackle this task. With both presets, the usage of color, depth, and opacity improved the display of the organ bounds while contrasting the tumor and calyx. 19 participants successfully completed the task using MODELAR. Their preference was to either find a good preset where the organ bounds were visible then adjust the color of target objects or vice versa. MODELAR helped us better identify effective visualization that best fit the task requirements. Our evaluation results and the modular framework MODELAR is freely available and open source at https://github.com/ghattab/MODELAR.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [Computer Graphics]: Multimedia Information Systems—Artificial, augmented, and virtual realities H.5.1 [Computer Graphics]: Multimedia Information Systems—Evaluation/methodology I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture

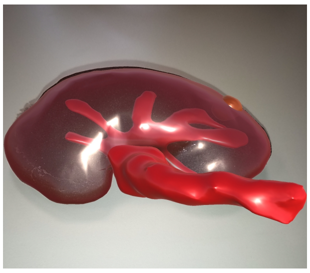
1. Introduction

The visualization of three-dimensional (3D) biomedical image data in the context of Augmented Reality (AR) yields a large number of different problems regarding the perception of the surgeon. Occlusion by surgical instruments, blood, and smoke is often a limiting factor for computational approaches. Moreover, it can be hard for the surgeon to understand the spatial division of the scene or to distinguish the individual objects. These problems need to be taken into account when selecting parameters for different visual channels. The task of selecting these parameters is an important aspect of understanding the data and underlying structures from the surgeons perspective. In this context, we have developed a framework that allows the user to select parameters for different visual channels so that they can be rendered on the fly into an AR visualization. While we defined presets to maximize visual channel differences, all parameters are user-settable. In our case and for evaluation purposes, we used the example of a surgical right kidney phantom (c.f. Fig. 1a) with an example renal calyx and tumor [KWW*15]. We defined a surgically realistic task: visualize the kidney boundary, the calyx, and an example tumor while they are all visually salient. By using our framework, we collected interaction data for 23 participants who solved the task. The framework enabled us to explore the task-specific parameter space, i.e. the parameters that lead to the solution. Indeed, this framework isn't restricted to this knowledge domain and could be extended to select appropriate parameters when an AR visualization is deployed for other clinical and non-clinical applications. In this paper, we present the contributions of the first free to use MODular and EvaLuative medical AR framework (MODELAR) for task-specific design and evaluation. This framework allows for: (a). interchangeable object models (i.e. modularity), (b). quick definition of presets to easily explore the visualization (or vis) parameter space, (c). task-specific evaluation of a vis by measuring the task completion time and all choices a participant made (i.e. data interaction), (d). human-readable output, (e). identifying and quantifying problems and, (f). testing the importance of appropriate encodings, renderings, and content for the augmentation.

† Corresponding author

DOI: 10.2312/evs.20201066

(a) *Silicone kidney phantom*



(b) *Augmented kidney*

Figure 1: *Example AR visualizing the kidney phantom using MODELAR. (a) Kidney phantom used as target for AR. (b) Visible objects are the calyx, the tumor, and the kidney surface.*

2. Related Work

Many approaches have employed AR in a surgical context and point towards the potential of head-mounted displays [SBA*16, YCK*18]. Although only a few use cases integrated handheld devices for surgical AR, the importance of different display types and their quality in the context of minimally invasive surgery was investigated. In general, the Maryland Visual Comfort Scale was developed to measure the perceptual qualities of a visual system, by relying on seven different metrics: Contrast, resolution or detail, brightness, sharpness, color, focus, and contrast-ratio [SSL*11]. Amid the rising demand for surgical AR systems, the number of applications in the literature has increased. This ranged from the specific domain application of AR for laparoscopic surgery, from partial nephrectomy [SPS*14], to surgical AR simulators to generally assess the skills of surgeons [LLSG15], to many other surgical specific tasks that involved the visualization of anatomical airways using AR [DHLD*02]. However, there's still a number of problems pertaining to surgical AR and specifically to augmenting or hiding visual elements from the surgical scene [SEL*03]. Perceptual issues that arise in AR were already explored by defining three different setups: Head-worn, handheld, and projector-camera. Each of these comes with its own sets of pros and cons. However, the handheld setup lead to most of the problems, where brightness and contrast of the object of interest were the hardest to perceive which was due to the low quality of the augmentation [KSF10]. Other results indicated that reflections and the viewing angle negatively influence and disturb the quality of the user's perception. Moreover, the visualizations designed in the AR have been found to play an important role [KSF10]. Indeed, while most papers offer general problems or strategies for finding good visualizations, none of them provide a solution to evaluate and find fitting visualizations for a given task. Our framework was built to address this issue as it provides presets that cover the parameter space of the visual encodings used for an AR visualization.

3. Methods

The evaluation framework MODELAR augments the scene on a mobile device, i.e., a tablet or a smartphone, by rendering 3D objects via the game engine Unity. First, we list and describe each visual channel in the next section. Second, we present the methods used to handle the realistic rendering of 3D objects under the Graphics section. Third, the how our framework is used to evaluate if a selected augmentation is reported under the usability section. Fourth and last, to address a well defined task, the Graphical User Interface (GUI) splits the process of completing the task of augmentation into two phases: (a) search phase, or find a good starting point by trial via the selection of a preset that fits as many requirements (Fig. 2a), (b) refining phase, or change the encoding of target channels to fit task requirements (Fig. 2b).

3.1. Visual Channels

The parameter space is defined to maximize the change of visual appearance of an object across different visual qualities or channels. They are applied onto the model using shaders. For a minimum required functionality and not to overwhelm users, we define and use five different channels: color, opacity, falloff, depth and specularity.

- **Color** c is responsible for the color of the selected object. The user can choose 1 out of 10 colors from the Tableau 10 categorical color palette [SS15].
- **Opacity** o sets the opacity for a selected object. Each pixel of the model is set to the same alpha-value. Depth-peeling is used to correctly blend the different transparent/translucent objects.
- **Falloff** f sets the alpha value of each pixel based on the viewing angle and the normal direction of the visible object fragment. This renders the object visible at a certain distance or falloff.
- **Depth** d makes the pixels that are further away from the camera darker in the vis, and enables depth perception [BBPS17].
- **Specularity** s increases specular lighting of an object.

Each channel was discretized to minimize the number of steps required to perceive visible change. This was done through trial and error while limiting the total and empirical number of discretized steps to 5 or 10, for clear changes or more subtle ones, respectively.

For example, by decreasing the opacity channel of an object, the augmentation disappears for that particular object. The discretization of this channel makes this a step-wise process. This design decision permits users to quickly change the visualization of each object to solve a specific task. By default, all channel encodings are user-settable. The parameter space of the main 3D object, that is to say the kidney, is reported in Table 1 for each channel to maximize differences between each preset.

3.2. Graphics

To render different objects in a scene in a clean and artefact free visualization, various methods are used. The specular highlights are rendered using the Blinn-Phong reflection model as it produces more accurate models thanks to the bidirectional reflectance distribution functions [Bli77]. The user can manipulate the exponent of the specular model in discretized steps. Computationally, the overall strength of the specular highlight is energy conserving thanks to a normalization term [AMHH*18]. Transparency is rendered using depth peeling [Eve01]. It delivers the best results in terms of computationally performance on mobile devices and the quality of the result. We favor the usage of depth peeling as it renders a scene

Table 1: *Parameter space for the visual channels that define each preset for the kidney object. Each row reports a preset. Each column reports a visual channel and the total number of discretized steps. In order of appearance, the visual channels are: color c, opacity o, falloff f, depth d, and specularity s. For example, the color channel has 10 discrete categorical colors as defined in the Tableau10 color palette (HEX values are reported).*

	c (10)	o (10)	f (10)	d (5)	s (5)
1	#1F77B4	10	0	0	5
2	#FF7F0E	10	4	1	3
3	#2CA02C	10	6	2	1
4	#8C564B	10	10	3	0
5	#FF7F0E	6	2	1	4
6	#BCBD22	10	8	0	2

multiple times. With each render-pass, one transparent layer is rendered then blended together with the other layers, which allows for complete control over the amount of correctly rendered transparent layers. This control is a function of the hardware capabilities and desired augmentation quality. Computationally, we justify our choice thanks to many pointers: (a) it's more efficient than vertex sorting, as there's no need to sort all vertices in the 3D objects for each rendered frame, (b) it's applicable even if the rendered meshes are convex and the multiple mesh surfaces are intertwined, (c) it fulfills the requirements of being computationally efficient and reliable for the rendering quality and, (d) it allows for user-settable transparency values without limiting responsiveness.

3.3. Usability

The AR framework Vuforia SDK allows the registration of a 3D model to detect a real world 3D object which corresponds to a given shape, texture, and color used for registration. Registration is object dependent and Vuforia SDK may be given interchangeable models to fit a particular context or knowledge domain. In our case, we use a kidney phantom model [KWW*15], as seen in Figure 1a. Once the model is registered, the object of interest is tracked. Upon successfully doing so, the remaining part is to overlap the detected object of interest with the augmentation so additional visual information may supplement the scene and assist in a specific task, e.g. identify the tumor and potential risk structures such as nearby vasculature. To assess the usability of the resulting framework and GUI, we design an example task and conduct a system usability scale (SUS). We recruit participants to solve the previously formulated task of contrasting kidney tumor and calyx while having the organ bounds visible. Typically, a task is given to participants, who then search and find a solution via our framework or app. The task may be defined for another specific domain since target objects used for registration and tracking in MODELAR are interchangeable. Task completion and elapsed time help quantify task-related metrics, such as difficulty and success. While presets could help identify preferences and help participants explore the parameter space. An example augmentation showing the real-time functionality of registration and tracking is depicted in Fig. 1b.

3.4. Graphical User Interface

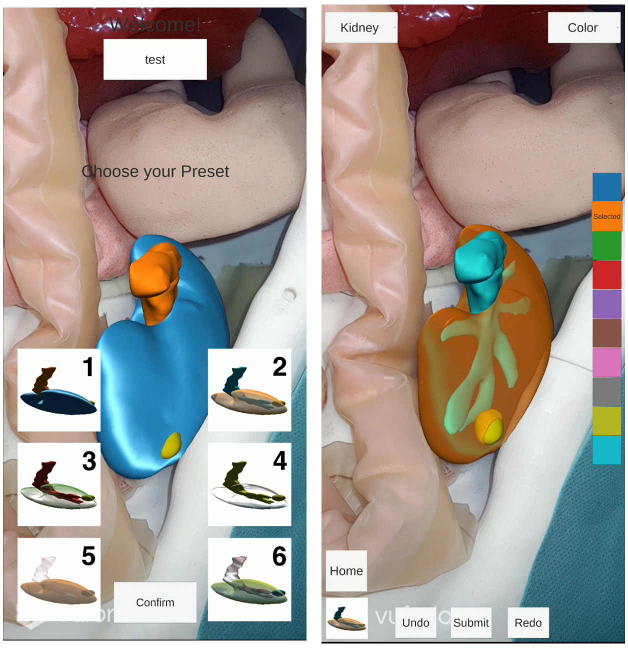
A GUI is provided so participants may complete a given task using MODELAR. Task completion is split into two phases: (a) a search phase to find a fitting visualization for the given requirements and (b), a refining phase to decide on subjective adjustments or to complete missing requirements that aren't provided by one of the presets. The search phase starts with entering the participant ID or name, then a grid of six presets is shown for the participant to select a preset from. Fig. 2a shows the GUI during the first phase. The presets correspond to encoding different value combinations. Presets 1 to 4 rely on a calculated variation of each channel, while presets 5 and 6 are randomly defined. A range of functionality defines the GUI, especially during the refining phase as shown in Fig. 2b. The selected preset is shown in the bottom left corner. The active object is reported in the upper left button while the active channel is reported on the upper right corner, *undo* and *redo* buttons are available to account for mistakes, the *Home* button takes a user back to the search phase to restart and reset the chosen encodings. When satisfied, the solution to the task may be submitted using the Submit button.

3.5. Interaction data

Every interaction is saved in order to be later analyzed. No interaction is possible without entering an identifier or a string of characters. The data is saved in a text formatted file with it as filename. MODELAR saves each selected preset, each channel value for each 3D object, undo and redo actions, as well as the submit action with a timestamp. With such data, many hypotheses could be tested.

4. Results

Twenty three participants were recruited to evaluate the usability of MODELAR in a task specific design. The task was to visualize and contrast the calyx and tumor while clearly distinguishing the kidney boundary. 60% of the recruited participants were male, 40% were female and an average age of 27 years. Over 90% of them had experiences with mobile devices for well over 6 years, while none of them had much experience with AR applications. In terms of presets, presets 2 and 3 were the most chosen. Nineteen participants chose those presets: 9 for preset 2 and, 10 for preset 3. These presets constituted 82% of the participants choice. Participants investigated the presets and made on average 16 preset changes. The minimum preset change was recorded to 2, while the maximum was 53 ($\sigma = 11.55$). For the parameter space and on average, participants chose the following channel values: falloff ($\mu = 4.5$, $\sigma = 1.9$), opacity ($\mu = 9$, $\sigma = 1.8$), depth ($\mu = 1.9$, $\sigma = 1.2$) and, specularity ($\mu = 2.2$, $\sigma = 1.1$). Based on the Kruskal-Wallis test, the overall distribution of the channel usage during the refining phase was identical, yet this phase took twice as long as the search phase $\sim$ 105s. On average, the time a participant took to complete a task was 2.9 min with 36 actions captured by the GUI ($\sigma = 67.9$s). One participant took over 6 min because they didn't understand the task, totalling twice as many actions, i.e. 72. On the contrary, the fastest participant took 1 min with exactly 8 actions. The System Usability Scale evaluated the usability of our app. MODELAR got the SUS score of 70.5 out of 100 with a good adjective rating.



(a) *Search phase.* **(b)** *Refining phase*

Figure 2: *GUI. (a) A preset is selected to fit the given task. The grid displays the presets from left to right and top to bottom. The image is annotated with the preset number. Preset 1 is shown by default after registration. (b) Additional changes are made to fit task requirements that are otherwise unmet by the chosen preset. Preset 2 is shown, with the active object kidney, and the visual color channel is selected with the color orange.*

5. Discussion

MODELAR is a free to use and modular framework that enables user study design and evaluation of augmented reality visualizations given a task. First, thanks to two well defined phases, a fitting visualization is chosen by searching for a suitable preset, then refining it to the task requirements. Indeed, to complete our specific task, presets 2 and 3 provided medium falloff, which seemed to be a good starting point for making the boundaries of the kidney visible. However, the kidney was quite opaque. Yet, lowering the opacity would make the kidney bounds also transparent. For this reason, most participants decided against changing this visual channel and opted for the falloff channel.

Second, different strategies were observed. Some participants selected a preset really fast, without even trying them all out, and made more changes in the second step. Others spent most of their actions selecting a suitable preset at first and only make a few adjustments later on. Such approaches became also visible in the wide range of actions the participants took. While many took much longer to decide between presets, few submitted their results very fast, that is to say after only a few actions and in a very short time.

Third, the search and refining phases permitted participants to explore the parameter space and select a subjectively *good* visualization. Our supplementary video shows an example usage, hence solution to the given task. Future work may extend this framework to assess the perceptual fidelity in surgical AR [SEL*03], or quantify the subjective meaning of appropriate encodings to solving a task.

Fourth, the current version limits the number of steps and makes the overall usability much easier for amateur users. However, this may be seen as limit for expert users. To this end, a future addition such as an expert mode could be introduced where experts use MODELAR without restrictions.

Fifth and last, the relevance of visualization and particularly the design of effective visualizations for surgical AR haven't been much considered in the literature. This is especially important due to the blinding effects of AR on attention [DDC*13]. MODELAR may be used for other legitimate yet typically overlooked questions such as addressing the quality of the surgeon's perception.

6. Conclusion

MODELAR is free to use, easy to deploy, and adapted for task based evaluation design. It permitted us to identify presets that best fit the requirements of a well defined task. The separation of the process into streamlined phases enabled participants to choose a preset then refine it for the task. Since participants preferences converged to similar results in the parameter space, we conclude that MODELAR works and leads to clear results for finding suitable and effective visualizations in the context of surgical AR.

Acknowledgments

The authors are grateful for the funding by a grant from the OP4.1 initiative, which was supported by the Federal Ministry of Economics and Technology (BMWi).

References

[AMHH*18] AKENINE-MÖLLER T., HAINES E., HOFFMAN N., PESCE A., IWANICKI M., HILLAIRE S.: *Real-Time Rendering 4th Edition*. A K Peters/CRC Press, Boca Raton, FL, USA, 2018. 2

[BBPS17] BEHRENDT B., BERG P., PREIM B., SAALFELD S.: Combining pseudo chroma depth enhancement and parameter mapping for vascular surface models. In *VCBM* (2017), pp. 159–168. 2

[Bli77] BLINN J. F.: Models of light reflection for computer synthesized pictures. *SIGGRAPH Comput. Graph. 11*, 2 (July 1977), 192–198. URL: http://doi.acm.org/10.1145/965141.563893, doi:10.1145/965141.563893. 2

[DDC*13] DIXON B. J., DALY M. J., CHAN H., VESCAN A. D., WITTERICK I. J., IRISH J. C.: Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surgical endoscopy 27*, 2 (2013), 454–461. 4

[DHLD*02] DAVIS L., HAMZA-LUP F. G., DALY J., HA Y., FROLICH S., MEYER C., MARTIN G., NORFLEET J., LIN K.-C., IMIELINSKA C., ET AL.: Application of augmented reality to visualizing anatomical airways. In *Helmet-and Head-Mounted Displays VII* (2002), vol. 4711, International Society for Optics and Photonics, pp. 400–405. 2

[Eve01] EVERITT C. W.: Interactive order-independent transparency. 2

[KSF10] KRUIJFF E., SWAN J. E., FEINER S.: Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality* (Oct 2010), pp. 3–12. doi:10.1109/ISMAR.2010.5643530. 2

[KWW*15] KENNGOTT H. G., WÜNSCHER J. J., WAGNER M., PREUKSCHAS A., WEKERLE A. L., NEHER P., SUWELACK S., SPEIDEL S., NICKEL F., OLADOKUN D., MAIER-HEIN L., DILLMANN R., MEINZER H. P., MÜLLER-STICH B. P.: Openhelp (heidelberg laparoscopy phantom): development of an open-source surgical evaluation and training tool. *Surgical Endoscopy 29*, 11 (Nov 2015), 3338–3347. URL: https://doi.org/10.1007/s00464-015-4094-0, doi:10.1007/s00464-015-4094-0. 1, 3

[LLSG15] LAHANAS V., LOUKAS C., SMAILIS N., GEORGIOU E.: A novel augmented reality simulator for skills assessment in minimal invasive surgery. *Surgical endoscopy 29*, 8 (2015), 2224–2234. 2

[SBA*16] SIM K. W., BAKER B., AMIN K., CHAN A., PATEL K., WONG J.: Augmented and virtual reality in surgery–the digital surgical environment: applications, limitations and legal pitfalls. *Annals of Translational Medicine 4*, 23 (2016). URL: http://atm.amegroups.com/article/view/12851. 2

[SEL*03] STOYANOV D., ELHELW M., LO B. P., CHUNG A., BELLO F., YANG G.-Z.: Current issues of photorealistic rendering for virtual and augmented reality in minimally invasive surgery. In *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003.* (2003), IEEE, pp. 350–358. 2, 4

[SPS*14] SCHNEIDER A., PEZOLD S., SAUER A., EBBING J., WYLER S., ROSENTHAL R., CATTIN P. C.: Augmented reality assisted laparoscopic partial nephrectomy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2014), Springer, pp. 357–364. 2

[SS15] SETLUR V., STONE M. C.: A linguistic approach to categorical color assignment for data visualization. *IEEE transactions on visualization and computer graphics 22*, 1 (2015), 698–707. 2

[SSL*11] SEAGULL F. J., SUTTON E., LEE T., GODINEZ C., LEE G., PARK A.: A validated subjective rating of display quality: the maryland visual comfort scale. *Surgical Endoscopy 25*, 2 (Feb 2011), 567–571. URL: https://doi.org/10.1007/s00464-010-1220-x, doi:10.1007/s00464-010-1220-x. 2

[YCK*18] YOON J. W., CHEN R. E., KIM E. J., AKINDURO O. O., KEREZOUDIS P., HAN P. K., SI P., FREEMAN W. D., DIAZ R. J., KOMOTAR R. J., ET AL.: Augmented reality for the surgeon: systematic review. *The International Journal of Medical Robotics and Computer Assisted Surgery 14*, 4 (2018), e1914. 2