

Semi-Automatic Initial Registration for the iRay System: A User Study

Tian Xie¹(orcid.org/0000-0003-0204-1124), Mohammad M. Islam¹, Alan B. Lumsden², and Ioannis A. Kakadiaris^{1*}(orcid.org/0000-0002-0591-1079)

¹Computational Biomedicine Lab, Dept. of Computer Science, Univ. of Houston, Houston, USA

²Methodist Research Institute, Houston, USA

ioannisk@uh.edu

Abstract. Simultaneous localization and mapping based augmented reality (AR) is trending in mobile AR applications. With the help of depth sensors, both accuracy and speed have improved. However, the method that performs the initial registration to align virtual objects with real scenery is not well developed, especially for some applications requiring very accurate registration (e.g., systems used in medical applications). For the iRay system, which is a mobile AR system using the Structure Sensor, we propose to use an iterative closest point algorithm to initially register the scanned mesh with the torso surface obtained pre-operatively, and then use SLAM to track pose. In this paper, a semi-automatic initial registration strategy is evaluated by a user study. This strategy is designed to help the user modify the selection of the 3D scanning area, so that the errors introduced by subjective differences can be reduced. The results indicate that the proposed strategy helps the users improve initial registration quality and reduces the average needed time.

Keywords: Medical AR, Mobile AR, 3D Registration, User Study.

1 Introduction

Simultaneous localization and mapping (SLAM) based augmented reality (AR) systems are becoming more and more popular in recent years. SLAM differs from traditional marker-based registration strategies in its ability to track camera pose from the spatial environment, not just markers. This provides AR freedom to explore the space rather than just observing small markers. It also generates a new demand for placing virtual objects directly into the 3D reconstruction.

For most mobile AR applications, the traditional techniques are based primarily on 2D information, such as markers [1]. When SLAM is applied, additional 3D information is available [2-3]. Virtual objects can be placed directly into 3D locations. The

efficiency and accuracy of the placement are limited by the performance of the SLAM algorithm used for mobile platforms.

With the advent of depth sensors, SLAM has become much more accurate and faster in recent years. Some AR glasses (e.g., Hololens [4] and Meta 2 [5]) are using depth-enhanced tracking systems to achieve robust pose estimation in real time. However, the registration methods have limited efficiency. In general, authoring requires an initial registration between virtual objects and a real 3D scene. It is application-specific, but always involves 3D shape/surface recognition. For instance, Hololens provides an interactive 3D authoring method based on surface recognition so that users can place models on the desk and walls with simple gestures. Although user-friendly, this kind of 3D authoring is not accurate and always depends on the user's perception.

When used in the medical domain, the initial registration accuracy for 3D authoring would be very restricted. It is necessary to develop accurate 3D registration methods for SLAM-based mobile AR applications.

In our previous work, we presented a mobile AR system named iRay [6]. It is implemented on the iPad and uses a Structure Sensor to obtain an accurate (1.5 mm) depth map. We have shown that the registration between 3D point clouds is useful for initial registration. However, the accuracy is highly dependent on the similarity of the two objects to be registered because we are using a simple iterative closest point (ICP) algorithm to ensure real time performance.

Currently, in iRay, we are using the MRI data of a male volunteer and his scanned torso surface mesh to perform our initial registration. Performance is not stable due to factors such as clothing, fullness of the stomach, and mostly the scanning area's size and location, which are selected by users with individual differences.

Focused on reducing these unstable inputs, a visual guidance strategy is proposed to help understand which area should be scanned for a better registration. The designed visual guidance includes an adjustable 3D box and semi-transparent overlay of the actual pre-operative torso surface model. Users can adjust the box's depth distance to fit the model onto the real torso. A user study was performed and indicated that this semi-automatic strategy can help the user to avoid most incorrect selections and improve their selection quality and speed.

The rest of the paper is organized as below. Section 2 illustrates the iRay system's constraints and major factors that make authoring unstable. Section 3 describes the user study with statistical results. Finally, Section 4 includes our discussion.

2 System Overview

2.1 The iRay System

The iRay system [6] uses mobile AR devices in a medical domain. Currently, the mobile platform is an Apple iPad and later will be extended to include AR glasses. A Structure Sensor [7] is used as the depth and SLAM information provider. Both 3D cloud points of the scanned area and active camera pose can be obtained by the Structure Sensor SDK.

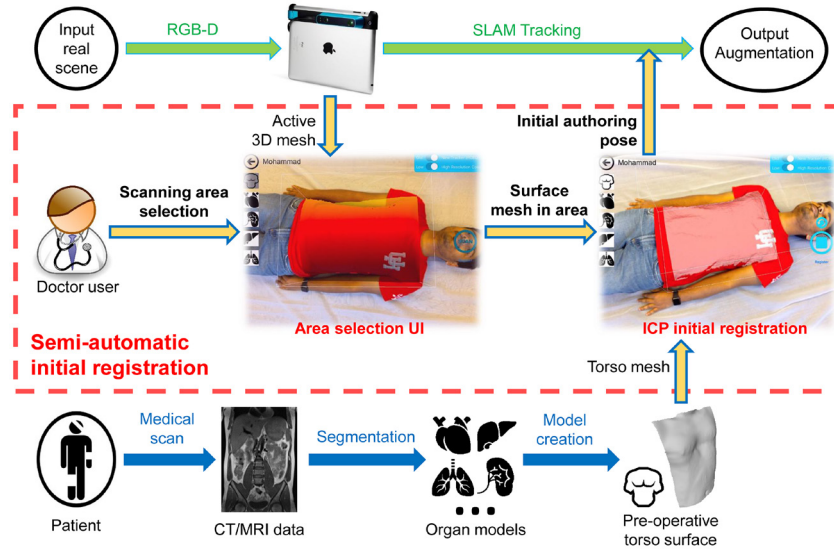


Fig. 1. Initial registration for iRay.

Figure 1 depicts the framework of the iRay system noting the semi-automatic initial registration module. In Figure 1, the blue arrows indicate offline related to patient data preparation. The green arrows indicate online processing running in every frame. The yellow arrows indicate the initial authoring functions performed only once just before the tracking loop starts. Note that the user selection of the scanning area is the only manual input in this system.

2.2 Factors of Instability

Figure 1 depicts the area selection user interface (UI). A preview of the 3D scanning area is drawn in red-to-yellow. This preview is decided by the cube box, which is rendered as a white wireframe in the middle of the display screen. Users can control this box's size and depth (the physical focal distance relative to iPad camera). They can also change the iPad's pose to focus on the region of interest.

This step is manual because automatically searching for or recognizing a torso surface in an unknown 3D environment is challenging. Sometimes the user may not notice small details. The area they selected might be too large, too small, too slanted, or sometimes may not include the whole surface. These incorrect selections are underestimated because they appear normal and tiny. Unfortunately, their influence is sometimes significant and cannot be corrected.

In our system, ICP is applied to register the 3D point clouds because ICP has been proven to be very efficient in depth-related work [8-9]. In the iRay system, the reference point cloud is always set to the scanned torso mesh, and the comparison point cloud is the pre-operative model's vertices. When ICP registration is complete, the estimated 3D transformation can be directly applied to virtual organ models. The ICP fitting

method is point-to-plane, and the iterations end when the transformation matrix is stable. Both point clouds are sampled to 500 points to balance accuracy and speed.

Under this specific setting, there are three factors that cause instability to the registration: 1) orientation of patient's head, 2) size of scanning area, and 3) horizontal and vertical displacement. The details are discussed in the following.

2.2.1 Orientation of Patient's Head

When the active mesh is captured in the wrong orientation (Figure 2), iRay's ICP will just converge to a local minimum. That is because we do not allow the ICP search large rotations, as that is too time consuming.

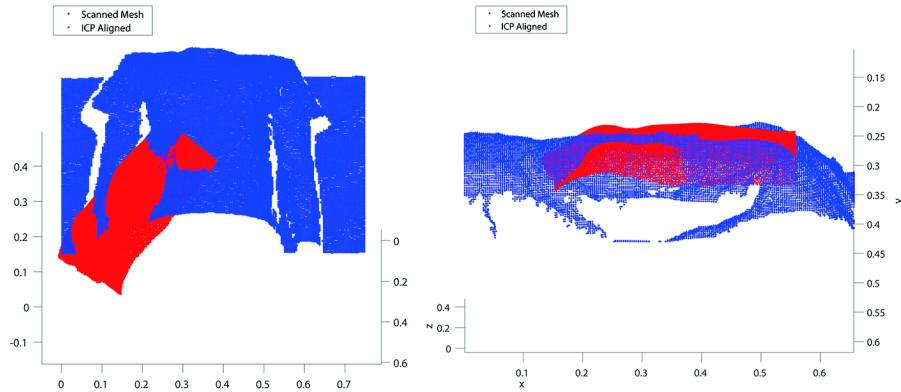


Fig. 2. Depicted are two examples of incorrect orientation registration. The red mesh denotes the pre-operative torso model that needs to register. The blue mesh denotes the reference mesh captured by the depth sensor. Both meshes are presented in full idensity without sampling.

This incorrect orientation occurs consistently if there is only an empty wireframe rendered in the screen. The user always needs guidance to find the correct orientation. Thus, in the strategy we propose in this paper, the comparison torso mesh is overlaid in the middle of the cube box. The user can easily infer the orientation as they can quickly recognize the shoulder, chest or other anatomical features, so the two meshes will be in the same orientation.

2.2.2 Size of Scanning Area

Size of scanning area matters because it affects the quantity of information contained in the reference mesh. If the size is too large, too many points and noise are introduced. If the size is too small, there is not enough data to achieve good registration.

Figure 3 depicts examples of a small scanning size issue. The registration quality cannot be measured as the reference mesh is just a flat patch and we do not know which body part it is. Figure 4 depicts an example of large scanning size. The registration result is adequate but still has a small drift. That is because the reference (blue) point cloud is always sampled to 500 points as discussed earlier; therefore, larger size implies sparser mesh. Too sparse mesh leads to low registration quality.

In some cases, larger scanning size is preferred as it generally contains additional information. For these reasons, in the strategy presented in this paper, the size of the

cube box is unchangeable to prevent the small scanning size issue. Specifically, the new size is a little larger than the actual size of the torso model, so that the reference mesh has additional vertices and the system can be more tolerable to the user's subjective biases.

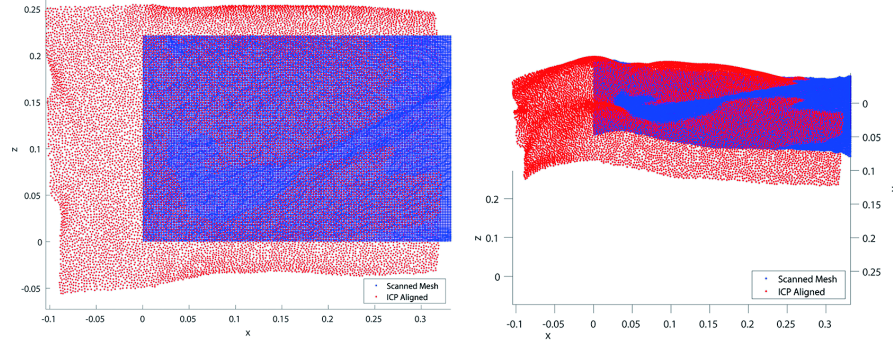


Fig. 3. Registration example when the reference (blue) mesh is too small. Depicted are two views of the same data.

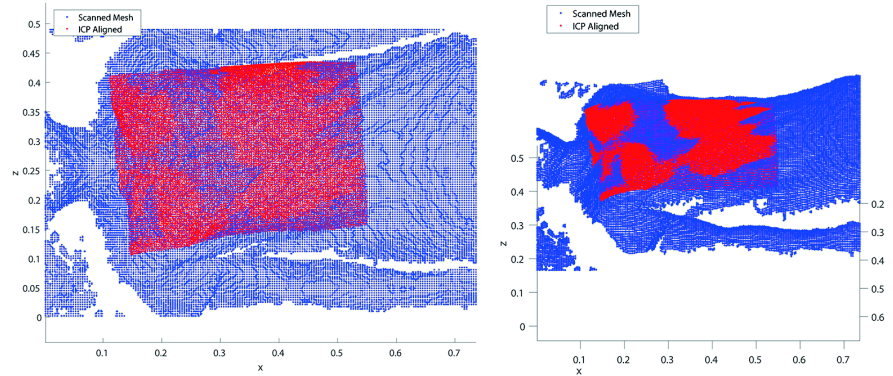


Fig. 4. Registration example when the reference (blue) mesh is too large. Depicted are two views of the same data.

2.2.3 Horizontal and Vertical Displacement

The vertical displacement is caused by the incorrect depth of the scanning box relative to the iPad camera. The horizontal displacement occurs on the perpendicular plane of the camera focal line. Figure 5 depicts two examples of these two kinds of displacement.

Figure 5(L) depicts an example of vertical displacement where the top surface of torso is incomplete because the scanning box is set too low, which means the depth is too far from the camera. This occurs even when a red depth preview has been shown. Some users may not care about the preview due to their individual preferences. Figure 5(R) depicts an example where the left side of torso has not been scanned because the

scanning box is displaced horizontally. This occurs when the user stands too far away from the patient, and s/he forgets to include the whole torso.

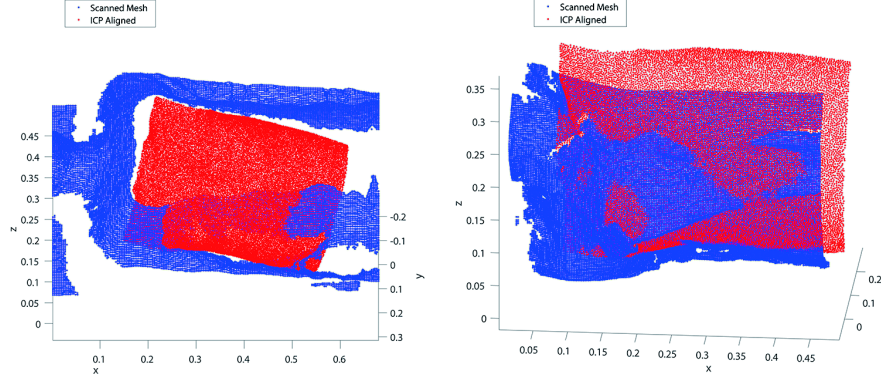


Fig. 5. Depiction of two registration examples with different displacement. (L) Depicted is an example of vertical displacement. The middle part of the torso has not been included in the box space. (R) Depicted is an example of horizontal displacement. The torso mesh is incomplete as one side has not been included in the scanning box.

To overcome the vertical displacement, the depth sensor automatically finds a supporting surface so the box can always stand above the background and will not be too low. However, when the patient is lying on a narrow bed (or similar situation), the system fails to find a correct or valid surface, so the strategy has been changed to a depth-adjustable version.

Another option was to scan the whole depth information within the selected frustum, so the top surface would always be complete. However, in some scenarios, there could be some medical equipment placed on the patient's body, and those data should not be included in the scanning mesh. Due to these constraints, training of the user is the best solution.

3 User Study

3.1 Selection Quality

Due to the above factors of instability, we propose a semi-automatic initial registration strategy to help the user modify the selection of the 3D scanning area. The interface is shown in Figure 6. The major change of the strategy over our prior work is to overlay the actual pre-operative torso surface model in the middle of the scanning box. When presented with this pre-rendered surface, the users can easily understand where to put the scanning box, so they can better avoid the factors of instability.

The box's size is automatically defined by the pre-operative model depicted in green color (Figure 6). This surface model serves as the comparison mesh in the initial registration. The major target of authoring is to align this model onto the patient data so that

when the user fits this green model onto the real torso, it is a good selection. Figure 6(R) depicts an example of good selection. The registration accuracy of this good selection can be inspected in Figure 7.

The distances visualized in Figure 7 are computed using a k-d tree-based nearest neighbor search. The search is performed among all the points without any sampling. The reference cloud contains 17,383 points. They are all generated from a 2D depth map with 1.5 mm precision. The comparison cloud has 14,726 points with a distribution density of about 2-3 mm.

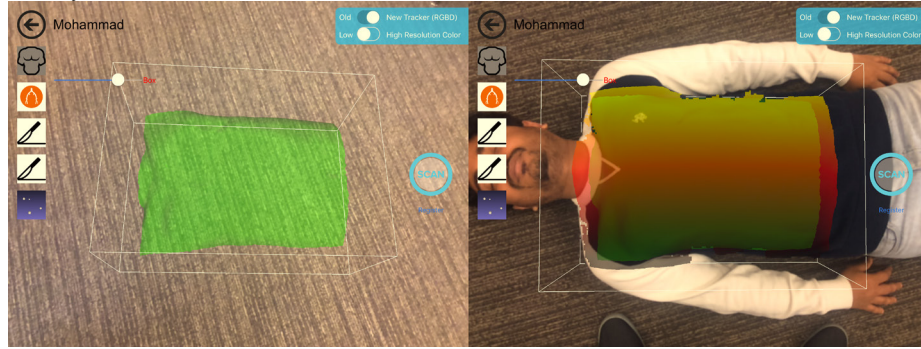


Fig. 6. Examples of new initial registration strategy. (L) Depicted is an example of the new User Interface. (R) Depicted is an example of good selection.

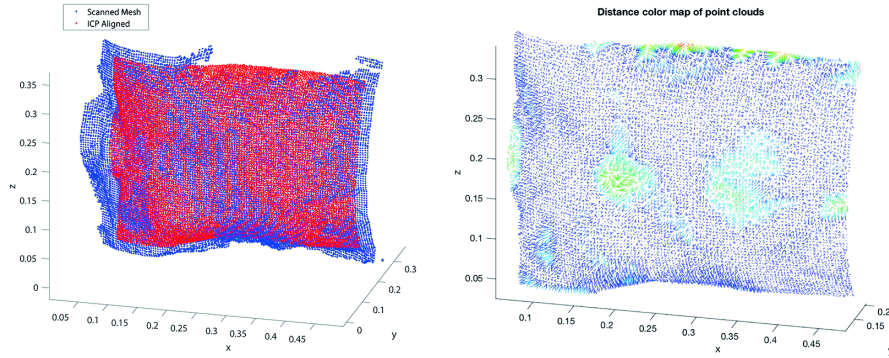


Fig. 7. Depiction of the registration result of the selection depicted in Figure 6. (L) Depiction of the aligned meshes. (R) Color map of the distances (computed by nearest neighbor search) between the point clouds. Red color denotes large distances while blue color small distances.

In this example, the smallest distance is 0.15 mm while the largest distance is 21.14 mm. The red color denotes the largest value, which is located in the top middle of Figure 7(R). In that region, the reference mesh was not complete, so the nearest neighbor is still far away.

This phenomenon of incomplete mesh can be used to estimate the quality of box area selection. If the quality is very low, such as in Figure 5, many points do not have any corresponding reference, and their nearest neighbor will be very far away. For example,

in Figure 5(L), the average distance of two point clouds is 42.86 *mm*, while in a good example such as Figure 6, the average distance is only 3.72 *mm*.

Therefore, we apply this *average distance measure* to describe the quality of box selection in the user study. If the user’s selection does not contain enough useful information, such as some important parts are cut out from the box (as in Figure 5), the average distance will be very high. On the other hand, if the user has selected the correct area, this value will be small.

It should be mentioned that this average distance can only describe the quality of box selection. It cannot represent the authoring accuracy of the AR initial registration. The AR registration accuracy should be measured by FRE and TRE [10].

3.2 Experimental Design

We designed a user study to compare the proposed strategy with the strategy presented in our previous work [6]. This strategy (V2.0) is described above and depicted in Figure 6, rendering the pre-operative model before registration starts. The old strategy (V1.0) can be used to change the box size but has no visual guidance. Table 1 summarizes the comparison of these two strategies. The new strategy does not use the size control and allows for model overlay.

Table 1. Capabilities available for initial registration.

Functions	V1.0	V2.0
Size control	✓	
Depth control	✓	✓
Depth preview	✓	✓
Model overlay		✓

After five minutes of short training and one to two times of practice, the operator is being asked to select the target area five times with the new strategy, and then five times with the old. Every selection starts with the same parameters, such as default box size and depth. The devices used are the same pair of iPad mini and Structure Sensor. The volunteer to be scanned is the same person wearing the same type T-shirt in the same pose (lying down on the same couch). The time that the user takes from the moment s/he starts controlling the iPad to when s/he presses the scan button is recorded. Every scan is recorded as 3D mesh, as is the ICP registration result. The average distance described in Section 3.1 is being computed after the experiments. Users do not receive any measurable feedback about the quality of their selection; they just try to select the whole torso area as they have been told in the training step.

3.3 Results

Eighty trials from eight different operators are evaluated in this study. The operators are between 24-31 years old, with at least a college education. None of them have any knowledge of iRay, nor do they have any medical experience. After the tests, there is a survey to rate themselves and provide feedback.

Time and Quality: Figure 8 depicts the average distance and average operation time of 80 tests (40 tests of each version). Note that comparing to the old strategy, new strategy has reduced the average distance by 11.3% and the average operation time by 28.1%. That means the users are achieving a better selection quality in less time, which indicates that our new strategy is very effective.

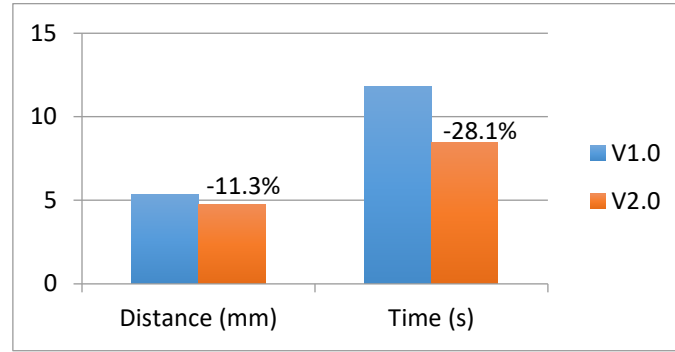


Fig. 8. Depicted are the average distance and operation time of all tests.

Reduction of Error: Although the average selection quality appears good (around 5 *mm* in Figure 8), subjective error is inevitable. As described above, our new strategy is designed to reduce factors of instability. For each selection in the user study, if the final average distance is larger than 6 *mm*, it will be labeled as an error, and the reason will be analyzed.

The final result indicates that there are 17.5% subjective errors among all trials in the test of old version and only 7.5% in the new; all of these errors are caused by displacement, as the factors of orientation and size are all voided after the training. This result demonstrates that our new strategy can efficiently reduce the subjective errors by 57.1%, which also proves its effectivity.

Survey: Users indicated that they agree with the description “new version strategy is easier to use” with an average score of 8.1, and they agree that “New version strategy has better performance (UI design)” with an average score of 8.5.

4 Conclusion

In this paper, we propose a new initial registration strategy to be used with our iRay system. It overlays the actual pre-operative torso surface model in the middle of a scanning box to help the user avoid some factors of instability caused by the user’s subjective decision.

A user study has been performed to evaluate the new strategy. The results show that the new strategy can help the users reduce subjective errors by 57%, improve the selection quality by 11.3% and increase the operation speed by 28.1%, which demonstrates the efficiency of the design.

Note that the testers in this study are young in age and have no medical knowledge. As our application is designed for physicians, we plan to conduct tests with physicians to receive their feedback.

Acknowledgements. This work was funded in part by the Methodist Research Institute and the UH Hugh Roy and Lillie Cranz Cullen Endowment Fund. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of the sponsors.

References

1. Coquillart, S., Brunnett, G., Welch, G.: Virtual Realities: Dagstuhl Seminar 2008. Springer Science & Business Media (2010). doi: 10.1007/978-3-211-99178-7
2. van den Hengel, A., Hill, R., Ward, B., Dick, A.: In situ image-based modeling. In Proc. 8th IEEE International Symposium on Mixed and Augmented Reality, pp. 107-110. Orlando, FL, USA (2009). doi: 10.1109/ISMAR.2009.5336482
3. Langlotz, T., Mooslechner, S., Zollmann, S., Degendorfer, C., Reitmayr, G., Schmalstieg, D.: Sketching up the world: In situ authoring for mobile augmented reality. *Personal and Ubiquitous Computing*, vol. 16, no. 6, pp. 623–630. (2012) doi: 10.1007/s00779-011-0430-0
4. Hololens. <https://www.microsoft.com/microsoft-hololens/en-us>.
5. Meta 2. <http://buy.metavision.com/products/meta2>.
6. Kakadiaris, I.A., Islam, M.M., Xie, T., Nikou, C., Lumsden, A.B.: iRay: Mobile AR Using Structure Sensor. In Proc. 15th IEEE International Symposium on Mixed and Augmented Reality. Merida, Mexico (2016).
7. Structure Sensor. <https://structure.io>.
8. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A.: KinectFusion: Real-time dense surface mapping and tracking. In Proc. 10th IEEE International Symposium on Mixed and Augmented Reality, pp. 127-136. Basel, Switzerland (2011). doi: 10.1109/ISMAR.2011.6092378
9. Newcombe, R.A., Fox, D., Seitz, S.M.: DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 343-352. Boston, MA, USA (2015).
10. Shamir, R.R., Joskowicz, L., Spektor, S., Shoshan, Y.: Localization and registration accuracy in image guided neurosurgery: A clinical study. *International Journal of Computer Assisted Radiology and Surgery*, vol. 4, no. 1, pp. 45-52. (2009) doi: 10.1007/s11548-008-0268-8