HHS Public Access

Author manuscript

Int J Comput Assist Radiol Surg. Author manuscript; available in PMC 2024 April 02.

Published in final edited form as:

Int J Comput Assist Radiol Surg. 2023 June ; 18(6): 1017–1024. doi:10.1007/s11548-023-02888-0.

Visualization in 2D/3D registration matters for assuring technology-assisted image-guided surgery

Sue Min Cho[1], **Robert B. Grupp**[1], **Catalina Gomez**[1], **Iris Gupta**[1], **Mehran Armand**[1,2], **Greg Osgood**[2], **Russell H. Taylor**[1,2], **Mathias Unberath**[1,2]

[1]Johns Hopkins University, Baltimore, MD, USA

[2]Johns Hopkins School of Medicine, Baltimore, MD, USA

Abstract

Purpose—Image-guided navigation and surgical robotics are the next frontiers of minimally invasive surgery. Assuring safety in high-stakes clinical environments is critical for their deployment. 2D/3D registration is an essential, enabling algorithm for most of these systems, as it provides spatial alignment of preoperative data with intraoperative images. While these algorithms have been studied widely, there is a need for verification methods to enable human stakeholders to assess and either approve or reject registration results to ensure safe operation.

Methods—To address the verification problem from the perspective of human perception, we develop novel visualization paradigms and use a sampling method based on approximate posterior distribution to simulate registration offsets. We then conduct a user study with 22 participants to investigate how different visualization paradigms (Neutral, Attention-Guiding, Correspondence-Suggesting) affect human performance in evaluating the simulated 2D/3D registration results using 12 pelvic fluoroscopy images.

Results—All three visualization paradigms allow users to perform better than random guessing to differentiate between offsets of varying magnitude. The novel paradigms show better performance than the neutral paradigm when using an absolute threshold to differentiate acceptable and unacceptable registrations (highest accuracy: Correspondence-Suggesting (65.1%), highest F1 score: Attention-Guiding (65.7%)), as well as when using a paradigm-specific threshold for the same discrimination (highest accuracy: Attention-Guiding (70.4%), highest F1 score: Corresponding-Suggesting (65.0%)).

Conclusion—This study demonstrates that visualization paradigms do affect the human-based assessment of 2D/3D registration errors. However, further exploration is needed to understand this effect better and develop more effective methods to assure accuracy. This research serves

✉Sue Min Cho scho72@jhu.edu.

Declarations

Conflict of interest All authors have no conflict of interest to declare.

Ethical approval The study was approved by the Homewood Institutional Review Board (HIRB00013292) and conducted according to ethical standards.

Informed consent Informed consents were obtained from all participants.

as a crucial step toward enhanced surgical autonomy and safety assurance in technology-assisted image-guided surgery.

Keywords

Image-guided surgery; Assured autonomy; 2D/3D registration; Perception; Visualization

Introduction

Minimally invasive surgeries (MISs) have become a reliable alternative to traditional open surgeries, as they promise less pain, better postoperative outcomes, and faster recovery. Image-guided navigation and surgical robotics systems are becoming more widespread in their utilization and have been invaluable intraoperative adjuncts during these surgeries [1]. Although generally considered safe and effective, errors during surgery due to human shortcomings, either because of missing information, poor decision-making, lack of dexterity, fatigue, or lack of attention, are not uncommon. To mitigate the impact on patient safety and on undesirable surgical outcomes, substantial advancements in approaches that increase precision and minimize costs have catapulted image-based navigation solutions into the center of attention for MIS research. In addition, the integration of surgical systems and the availability of new generations of robotic systems like the da Vinci have made the adoption of MIS approaches in complex operations possible. Some advantageous features of such a robotic system include a high-fidelity video display, a stereo endoscope, sensors to track the position of each surgical instrument, and a ready-made stream of data for registration and tracking.

Current surgical robots are fully teleoperated, and the introduction of “supervised” autonomy, where subtasks are delegated to the robot and performed autonomously under close supervision by a human surgeon, is much-desired [2]. Such supervised autonomy has the potential to reduce the learning curve, increase overall accuracy, and actualize remote telesurgery.

However, the dream of semi-autonomous or autonomous MIS can only be realized if the operating surgeon is assured that the instantaneous, intraoperative sensory information received is reliable and accurate. Challenges remain in the field of spatial registration of the preoperative data to the intraoperative scene [3]. Errors may result, among other things, from poor exposure of the anatomy, a mismatch between pre- and intra-operative anatomy due to surgical manipulation, or because of the presence of noisy measurements. Image registration plays an essential role in augmenting intraoperative images and is widely used for computer-assisted interventions. 2D/3D registration is a technique that helps find the most optimal geometric transformation between a 3D model of the scene into the same coordinate frame of one or more intraoperative images [4, 5]. In spine and orthopedic minimally invasive operations, such as total hip arthroplasty or osteotomies of the pelvis or femur, 2D/3D registration has been applied to help orient the current position of instruments relative to the planned trajectory, nearby vulnerable structures, and the ultimate target [6]. In image-guided endoscopy, 2D/3D registration provides augmented reality, which enables the display of anatomical structures that are hidden from the direct view by currently

exposed tissues or in interventional radiology. Such visualizations can be beneficial in procedures that involve laparoscopy of the abdomen, arthroscopy of musculoskeletal joints, colonoscopy of the colon, or bronchoscopy of the lung registration.

In this study, we aim to explore the assurance of 2D/3D registration between postoperative CT scans and intraoperative X-rays so that any inaccurate results can be corrected by rerunning the algorithm with altered initialization or hyper-parameters before the semi-autonomous execution in an MIS intervention. We address the assuredness of this confirmation step from a human-in-the-loop perspective, where human stakeholders have the opportunity to evaluate the result of a 2D/3D registration algorithm and approve or reject the result based on perceptual cues provided by a visualization paradigm. We present two novel visualization paradigms for displaying 2D/3D registration results. To sample plausible registration results for user assessment, we make use of a probabilistic formulation of the registration process. We, therefore, assume a traditional single-view, single-object, intensity-based 2D/3D registration approach. We combine this approach with an objective function to construct an approximate posterior distribution over registration results that are approximately equivalent in terms of computational image similarity, i. e., poses that are perceptually similar to the true pose and might have been accepted by image-based 2D/3D registration [7]. With this offset simulation, we investigate the effectiveness of the visualization paradigms in conveying alignment precision to human users.

Related work

The focus of this study is not on which technique of 2D/3D registration but on assuring and verifying the process. The verification of registration, in general, has been explored in two ways, uncertainty, and perception.

Uncertainty in data can seriously affect its analysis and subsequent decision-making. For example, in image-guided surgery, errors may be introduced during data acquisition (through an imaging modality or tracking system), during transformation (by registration or segmentation), or during rendering [8]. However, uncertainty might be ignored because of the inherent difficulty in expressing, computing, or visualizing it. Henceforth, new methods of visualizing registration uncertainty are needed. There have been studies that aim to visualize registration error by estimation of uncertainty [9]. Simpson et al. proposed a method of visualizing registration uncertainty by determining the variation introduced along a linear path [10]. Their visualization method resulted in a statistically significant reduction in attempts required to localize a target and in targets that the pool of subjects failed to localize. Risholm et al. presented data in the case of non-rigid registration [11]. The authors assessed the uncertainty of the estimated posterior distribution by visualizing the interquartile range (IQRs) computed from the marginal posteriors in the form of scalar maps and ellipsoids. The uncertainty information was also emphasized using marginal volumes, marginal visitation count volumes, and marginal confidence bounds.

Numerous visualization methods have been studied to improve spatial perception, including the extraction of relevance-based data only according to pixels and voxels [12]. Volume rendering results were enhanced by improving depth perception with an energy function,

and conjugate gradient method [13] and introducing volumetric halos [14]. A new method called chromatic shadowing was based on a shadow transfer function to handle the over-darkening problem to better allow the perception of depth and surface [15]. There also have been color-based methods to improve depth perception [12, 16]. However, to date, there have not been any studies on the effectiveness of 2D/3D registration visualization methods on the perception of alignment. Perhaps the most similar problem and study design of this study is [3], where the authors explore the 3D registration alignment perception problem in AR/MR environment.

Methods

Study design

In total, 22 participants (14 males, 8 females) from the Johns Hopkins University were recruited through convenience sampling for the user study. Participants' ages ranged from 18 to 39 ($M = 25.86$, $SD = 4.80$). Our target participant sample was from the general population because the target users who will be assuring the safety of these computer-assisted intervention systems will likely be company representatives who provide technology support.

A user interface platform based on Next.js was developed for the data collection. We followed a within-subjects design in which participants were exposed to all visualization paradigms. We randomized the order of the paradigms, the X-ray images, and the registration offsets to prevent learning effect. The user study began with the instructions for the task and informed agreement for participation. For each visualization paradigm, an explanation of the paradigm and example cases of the different degrees of offsets were given. Then, for 12 cases, an X-ray image and its registration overlay were shown, and the user was asked to assess the alignment with a slider and provide their confidence level of assessment. After all three paradigm sections, the user was asked to fill out a poststudy survey for their demographics and preference ranking of the paradigms. This user study was approved by the local IRB.

Visualization paradigms

For all three visualization paradigms (Fig. 1), accessible colors were used.

Paradigm 1: Neutral—This is the baseline paradigm that is most commonly used in 2D/3D registration methods of rigid anatomy. The edge map is developed from the digitally reconstructed radiograph (DRR) of the simulated registration offset.

Paradigm 2: Attention-guiding—This paradigm is generated by computing the mutual information (MI) of local regions between the original X-ray and the DRR. The MI values are then min-max normalized, and those greater than 0.0 and less than 0.3 are circled as the areas of uncertainty that a human should look closely at. The search window size and the circle size are parameters that were set as 100 pixels and 250 pixels, respectively, for our visualization generation. The circles guide the user's attention and serve as an indirect cue of the misalignment.

Paradigm 3: Correspondence-suggesting—This paradigm is generated by searching through the local regions of the DRR to find the best local match in the original X-ray and visualizing the correspondence vectors as arrows. The window size and the extended search size are parameters that were set as 20 pixels to generate our visualizations. The arrows provide a direct cue of possible misalignment by revealing the magnitude and direction of the offset.

Simulation and sampling registration offsets

To generate image offsets of varying strengths that are plausible considering the 2D/3D registration problem, we rely on a technique initially proposed for uncertainty quantification in 2D/3D registration [7]. Concisely, this technique produces poses that, given an image similarity function, could have possibly been accepted as a solution to the 2D/3D registration problem, making them more relevant to this study that images with arbitrary offsets.

Specifically, we leverage the single-view, pelvis-only, portion of the intraoperative 2D/3D registration strategy from [6]. Observed image data include a 3D CT volume of the pelvis (V) and a 2D fluoroscopy image of the pelvis (I). The registration problem of finding the pelvis volume pose, $\theta \in \mathbb{R}^6$, with respect to the projective imaging geometry is defined by:

$$\min_{\theta \in SE(3)} \mathcal{S}(\mathcal{P}(\theta; V),\ I) + \mathcal{R}(\theta), \tag{1}$$

where $\mathcal{P}$ indicates a projection operator creating DRRs, $\mathcal{S}$ indicates a similarity measure between DRRs and fluoroscopy, and $\mathcal{R}$ is a regularization term. The $\mathcal{P}$ operator encapsulates the intrinsic parameters of the imaging device which are considered known and kept fixed during the registration process.

Let θ^* be the solution to (1). We assume that registration strategies are more likely to report pose estimates nearer to θ^*, rather than further away, with the distribution of estimates depending on the shape of objective function. Candidate registration estimates are modeled as offsets, $\Delta\theta$, from the true solution, so that $\theta = \theta^* + \Delta\theta$. Proceeding in a Bayesian fashion, the image similarity component of (1) is used to construct the likelihood distribution, $p(I|\Delta\theta)$ and the regularization term is used to model the prior, $p(\Delta\theta)$, distribution. Simulated registration solutions may be sampled from the posterior distribution, which is proportional to their product: $p(\Delta\theta|I) \propto p(I|\Delta\theta)p(\Delta\theta)$.

A Boltzmann distribution is used to define the likelihood distribution with the image similarity component of (1) used as the energy function:

$$p(I|\Delta\theta) \propto \exp\left\{-\frac{1}{2}\mathcal{S}\left(\mathcal{P}\left(\theta^* + \Delta\theta;\ V\right),\ I\right)\right\}. \tag{2}$$

Standard sampling strategies, e. g., Metropolis Hastings, are not computationally practical when using this exact likelihood. Thus, we begin by approximating the similarity component using a quadratic model: $Q(\Delta\theta) = \Delta\theta^T H \Delta\theta + \mathcal{S}(\mathcal{P}(\theta^*;\ V),\ I)$, where H is a 6×6 positive-definite matrix. Q is obtained by evaluating $\mathcal{S}$ on a 6D regularly spaced grid about θ^* and performing a least squares fit of H. Each rotation parameter ranged from −1° to 1° in 0.5° offsets and each translation parameter ranged from −1 mm to 1 mm in 0.5 mm offsets, for a total of 15,625 grid points. Using Q as the likelihood energy term yields:

$$p(I|\Delta\theta) \propto \exp\left\{-\frac{1}{2}\left[\Delta\theta^T H \Delta\theta + \mathcal{S}(\mathcal{P}(\theta^*;\ V),\ I)\right]\right\}$$
$$\propto \exp\left\{-\frac{1}{2}\left[\Delta\theta^T H \Delta\theta\right]\right\}.$$

The prior distribution is modeled as a multivariate normal distribution with diagonal covariance matrix, $\Sigma_{\text{prior}} = \text{diag}(\sigma_1^2, \ldots, \sigma_6^2)$. The posterior distribution is thus:

$$p(\Delta\theta|I) \propto \exp\left\{-\frac{1}{2}\left[\Delta\theta^T H \Delta\theta\right]\right\}\exp\left\{-\frac{1}{2}\left[\Delta\theta^T \Sigma_{\text{prior}}^{-1} \Delta\theta\right]\right\} \quad (3)$$

$$\propto \exp\left\{-\frac{1}{2}\left[\Delta\theta^T \left(H + \Sigma_{\text{prior}}^{-1}\right)\Delta\theta\right]\right\} \quad (4)$$

Since both H and $\Sigma_{\text{prior}}^{-1}$ are positive definite matrices, their sum is also positive definite, making (4) equivalent to a multivariate normal distribution with zero-mean and covariance matrix $\left(H + \Sigma_{\text{prior}}^{-1}\right)^{-1}$. Therefore, this approximate posterior enables an efficient sampling process of candidate registration results.

Twelve pelvis fluoroscopy images from 5 unique specimens and 3 approximate viewing directions (i. e., anterior–posterior, oblique, partial) were selected from a dataset [1] containing reference ground truth pelvis poses and anatomical landmarks [6]. For each image, 1,000 simulated registration offsets were sampled using the approximate posterior distribution. To draw enough samples with the smaller offset values, the standard deviation values of the prior distribution were all set to either 1 mm/1°. The mean Target Registration Error (mTRE) was computed using the ground truth 3D landmarks and the 3D transformed points from the offset, and used to sample the offsets to be used in the study. For each X-ray image, 4 offsets (1 from each offset group bin, 0: 0–2 mm, 1: 2–4 mm, 2: 4–6 mm, 3: 6–8 mm), were sampled.

[1] https://doi.org/10.7281/T1/IFSXNV. This dataset was annotated by a single operator experienced with the interpretation of fluoroscopy in the context of 2D/3D registration for orthopedic applications using a custom image analysis pipeline.

Experimental variables

We considered one independent variable, the visualization paradigms, with three levels. One dependent continuous variable was collected with the question, “How would you assess this registration result?” to measure the user’s perceived assessment of the registration. Two other continuous dependent variables were the perceived confidence measured with a 5-point Likert scale and the preference ranking of the paradigms.

Hypotheses

Our general assumption is that visualization paradigms have an effect on the alignment assessment of human users. We formulated the following hypotheses:

H1. For each visualization paradigm, users can perceive and detect registration errors with better than random guessing performance.

H2. Users can differentiate registration errors more accurately when using visualization paradigms that encode more cues of spatial misalignment.

Results

We used Python and RStudio for the data analysis. The statistical tests included one fixed effect and participants as a random effect. For all the statistical tests reported, $p < .05$ is considered as a significant effect.

Can users perceive different offset groups?

We first validated whether participants perceived differences across all the offset levels. A one-way repeated-measures analysis of variance (ANOVA) was conducted to examine the effect of the offset group on participants’ alignment assessments. We observed a significant main effect of the offset group on the alignment assessments ($F(3, 63) = 34.86$, $p < .001$, $\eta_p^2 = .19$). Post hoc pairwise comparisons using Tukey’s HSD test revealed that the assessments were significantly higher (all p-values $p < .001$) in the group with offsets between 0 and 2 mm ($M = 0.66$, $SD = 0.28$) than in the group with offsets between 2 and 4 mm ($M = 0.50$, $SD = 0.25$), 4 and 6 mm ($M = 0.41$, $SD = 0.27$), and 6 and 8 mm ($M = 0.31$, $SD = 0.24$). Likewise, the assessments when the offsets were between 2 and 4 mm were significantly higher than those when the offsets were between 4 and 6 mm ($p = .003$) and 6 and 8 mm ($p < .001$). Moreover, the assessments were significantly higher ($p = .001$) in the group with offsets between 4 and 6 mm than in the group with offsets between 6 and 8 mm. Figure 2 (left) presents these differences.

In [3], it was shown that conventional mixed reality visualization paradigms do not effectively allow users to accurately assess 3D alignment errors within 4 mm, and in image-guided surgical navigation, mTRE under 2 mm is considered an acceptable registration. For these reasons and with the validation that users can differentiate between the four offset groups, we focus on differentiating between offsets smaller than 4 mm for the following analyses.

Do visualization paradigms affect assessment performance?

With the group with offsets between 0 and 2 mm and the group with offsets between 2 mm and 4 mm binarized as adequate and inadequate registration quality, respectively, and the assessment values provided by the users ranging from 0 to 1, we construct a receiver operator characteristic (ROC) for each of the paradigms (Fig. 2). It can be seen that all three paradigms allow the users to perform better than random guessing. The overall performance of the visualization paradigms for the task of registration assessment, can be seen with the area under the curve (AUC), where the Attention-Guiding paradigm ranks highest with 0.71, followed by 0.66 for the Correspondence-Suggesting and 0.64 for the Neutral paradigm.

To further interpret the assessment performance of the paradigms, we perform analyses considering both an absolute threshold of 0.50 for assigning adequacy as well as an adaptive threshold with each of the paradigm's optimal assignment threshold. The paradigm-specific thresholds were calculated by finding the indices of the maximum difference between the true positive rate (TPR) and the false positive rate (FPR), and these points are drawn on the ROC curve (Fig. 2). The optimal thresholds for the three paradigms were calculated to be 0.80, 0.72, 0.62 for paradigms 1 through 3, respectively. Then, the prediction labels are computed by binarizing the user's assessment value to accepting and rejecting the registration results based on the given thresholds.

Table 1 shows the assessment performance of each of the paradigms by the two threshold methods. For the absolute threshold method, Correspondence-Suggesting has the highest accuracy of 65.1% and Attention-Guiding has the highest F1 score of 65.7%. For the adaptive threshold method, Attention-Guiding has the highest accuracy of 70.4% and Corresponding-Suggesting has the highest F1 score of 65.0%.

A one-way repeated-measures ANOVA was conducted to examine the effect of the visualization paradigm on assessment performance measured as the fraction of correct assessment outcomes (true positives and true negatives) with respect to the number of registrations evaluated in the two more precise offsets using both the absolute and adaptive thresholds. The tests did not reveal a significant difference in the assessment accuracy across the visualization paradigms for the absolute threshold ($F(2, 42) = 1.06$, $p = .356$) nor the adaptive thresholds ($F(2, 42) = 0.62$, $p = .541$). Using the absolute threshold, participants' average accuracy for the Neutral visualization was 56.5% ($SD = 21.3$), 61.4% ($SD = 16.7$) for the Attention-Guiding, and 65.0% ($SD = 18.6$) for the Correspondence-Suggesting paradigms. As expected, the average accuracy values were overall larger when using the adaptive thresholds (Neutral: $M = 65.3$, $SD = 16.7$, Attention-Guiding: $M = 70.2$, $SD = 15.0$, and Correspondence-Suggesting: $M = 67.3$, $SD = 16.8$).

What are the users' self-reported ratings?

Perceived assessment confidence—A one-way repeated-measures ANOVA was conducted to examine the effect of the visualization paradigm on participants' subjective confidence on their assessments (from 1 = not confident to 5 = very confident). The test revealed a significant main effect of the visualization paradigm on subjective confidence ratings ($F(2, 42) = 4.26$, $p = .021$, $\eta_p^2 = .02$). Post hoc pairwise comparisons using Tukey's

HSD test showed that on average, confidence ratings are significantly larger in the Correspondence-Suggesting visualization ($M = 3.59$, $SD = 0.91$) than in the Attention-Guiding visualization ($M = 3.26$, $SD = 0.93$), p .013. No significant differences were found with respect to Neutral visualizations ($M = 3.44$, $SD = 0.91$).

Perceived preference—Participants ranked their preference of the three visualization paradigms (from 1 = most preferred to 3 = least preferred), and we used a Friedman's test to compare these ratings. Friedman's test showed a statistical significant difference in the mean ranks among the three visualization paradigms ($\chi^2(2) = 6.42$, $p = .040$). The post hoc Wilcoxon tests with a Bonferroni correction showed that the mean ranks of the Attention-Guiding ($M = 2.26$, $SD = 0.56$) and Correspondence-Suggesting paradigms ($M = 1.53$, $SD = 0.84$) were significantly different ($p = .047$). There were no significant differences between the Neutral visualization ($M = 2.21$, $SD = 0.86$) and other paradigms.

Discussion and conclusion

In this study, we present two new visualization paradigms for displaying 2D/3D registration results and conduct a comparison study with an existing visualization method for assessing registration errors. With the user interface displaying simulated registration results, users evaluate the registration error by assessing how well the registration aligned the 2D with the 3D data.

Our findings show that visualization paradigms do matter in the human-based assessment of 2D/3D registration. With all three visualization paradigms, not only are users able to differentiate mTREs ranging in the groups of 0–2 mm, 2–4 mm, 4–6 mm, and 6–8 mm, but can distinguish offsets smaller than 4 mm. It is also shown that the two novel paradigms, Attention-Guiding and Correspondence-Suggesting visualization, allow users to perform better in the registration assessment task, with higher accuracy and F1 scores than the Neutral paradigm in both absolute and adaptive thresholding methods. In addition, our tests revealed that participants have significantly higher perceived confidence ratings and preference using the Correspondence-Suggesting paradigm than the Attention-Guiding paradigm. Interestingly, the optimal threshold was lower for the Correspondence-Suggesting paradigm (0.62) than the Attention-Guiding paradigm (0.72), suggesting that the Correspondence-Suggesting paradigm allows users to evaluate the registration errors more conservatively, while making their decision more confidently.

While the results hold on average, the overall accuracy needs to be improved, and there is still considerable variance in the alignment perception for individual decisions. It can be observed in Fig. 2 (Left) that although participants' ability to distinguish between different types of errors on average, there exists both strong accepts and rejects of registrations, regardless of the error category. Furthermore, despite the improvement of accuracy and F1 scores of the novel paradigms compared to the standard Neutral paradigm Table 1, the performance is not sufficient for the reliable detection of spatial misalignment. The factors that may cause this variance, as well as how to reduce this variance and increase accuracy, should be further investigated.

Human factors can be explored to gain a better understanding of how users process the information provided to them to perceive and assess alignment. For instance, the analysis of their gaze pattern, including fixation and entropy, can be correlated to the (in)accuracy of their registration assessment. Other human factors may include the time it takes for the user to evaluate the result, the time spent on the training session before the assessments, the number of transitions from overlay to X-ray image, and the users' risk adversity level.

We use mTRE as the metric to quantify and differentiate the registration offsets, but the offsets can be deconstructed into the translational and rotational error components. Moreover, the mean target reprojection error, in addition to the mTRE, may also play a role in assessing how perceptible even large 3D registration offsets actually are when visualized in 2D projections. It is also worth considering that while our registration error is minimal (i.e., below 1–2 mm mTRE), the ground truth of the dataset used in this study is not perfect. Further analyses of these error components and the human factors mentioned above, can provide insight into how registration results mislead users and how they mislead image similarity metrics. With these insights, more effective methods can be developed and tested to communicate misalignment errors to human inspectors for accurate verification of 2D/3D registration.

This study serves as a preliminary exploration of this emerging research field in human-in-the-loop safety assurance for technology-assisted image-guided surgery. Our findings reveal that the quality assurance of 2D/3D registration is indeed impacted by visualization paradigms. However, a more comprehensive understanding of this impact is necessary to identify opportunities for enhancement. Gaining a deeper understanding of these effects is essential for developing more effective methods to ensure accuracy and safety in surgical procedures. Ultimately, these insights will be critical for advancing assured surgical autonomy and expanding its applicability across diverse surgical scenarios.

References

1. Kochanski RB, Lombardi JM, Laratta JL, Lehman RA, O'Toole JE (2019) Image-guided navigation and robotics in spine surgery. Neurosurgery 84(6):1179–1189 [PubMed: 30615160]
2. Fiorini P, Goldberg KY, Liu Y, Taylor RH (2022) Concepts and trends in autonomy for robot-assisted surgery. Proceed IEEE 110(7):993–1011
3. Gu W, Martin-Gomez A, Cho SM, Osgood G, Bracke B, Josewski C, Knopf J, Unberath M (2022) The impact of visualization paradigms on the detectability of spatial misalignment in mixed reality surgical guidance. IJCARS 17(5):921–927
4. Markelj P, Tomaževič D, Likar B, Pernuš F (2012) A review of 3d/2d registration methods for image-guided interventions. Med Image Anal 16(3):642–661 [PubMed: 20452269]
5. Unberath M, Gao C, Hu Y, Judish M, Taylor RH, Armand M, Grupp R (2021) The impact of machine learning on 2d/3d registration for image-guided interventions: a systematic review and perspective. Front Robotics AI 8:716007
6. Grupp RB, Unberath M, Gao C, Hegeman RA, Murphy RJ, Alexander CP, Otake Y, McArthur BA, Armand M, Taylor RH (2020) Automatic annotation of hip anatomy in fluoroscopy for robust and efficient 2d/3d registration. IJCARS 15(5):759–769
7. Grupp RB (2020) Computer-assisted fluoroscopic navigation for orthopaedic surgery. PhD thesis, Johns Hopkins University
8. Weiskopf D (2022) Uncertainty visualization: concepts, methods, and applications in biological data visualization. Front Bioinform 10.3389/fbinf.2022.793819

9. Gillmann C, Saur D, Wischgoll T, Scheuermann G (2021) Uncertainty-aware visualization in medical imaging-a survey. Comp Graph Forum 40:665–689
10. Simpson AL, Ma B, Chen E, Ellis RE, Stewart AJ (2006) Using registration uncertainty visualization in a user study of a simple surgical task. In: MICCAI, pp. 397–404 Springer
11. Risholm P, Pieper S, Samset E, Wells WM (2010) Summarizing and visualizing uncertainty in non-rigid registration. In: MICCAI, pp. 554–561. Springer
12. Wang J, Kreiser M, Wang L, Navab N, Fallavollita P (2014) Augmented depth perception visualization in 2d/3d image fusion. Comp Med Imag Graph 38(8):744–752
13. Zheng L, Wu Y, Ma K-L (2012) Perceptually-based depth-ordering enhancement for direct volume rendering. IEEE Trans Visualiz Comp Graph 19(3):446–459
14. Bruckner S, Gröller E (2007) Enhancing depth-perception with flexible volumetric halos. IEEE Trans Visualiz Comp Grap 13(6):1344–1351
15. Šoltészová V, Patel D, Viola I (2011) Chromatic shadows for improved perception. In: Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering, pp. 105–116
16. Chuang J, Weiskopf D, Moller T (2009) Hue-preserving color blending. IEEE Trans Visualiz Comp Graph 15(6):1275–1282

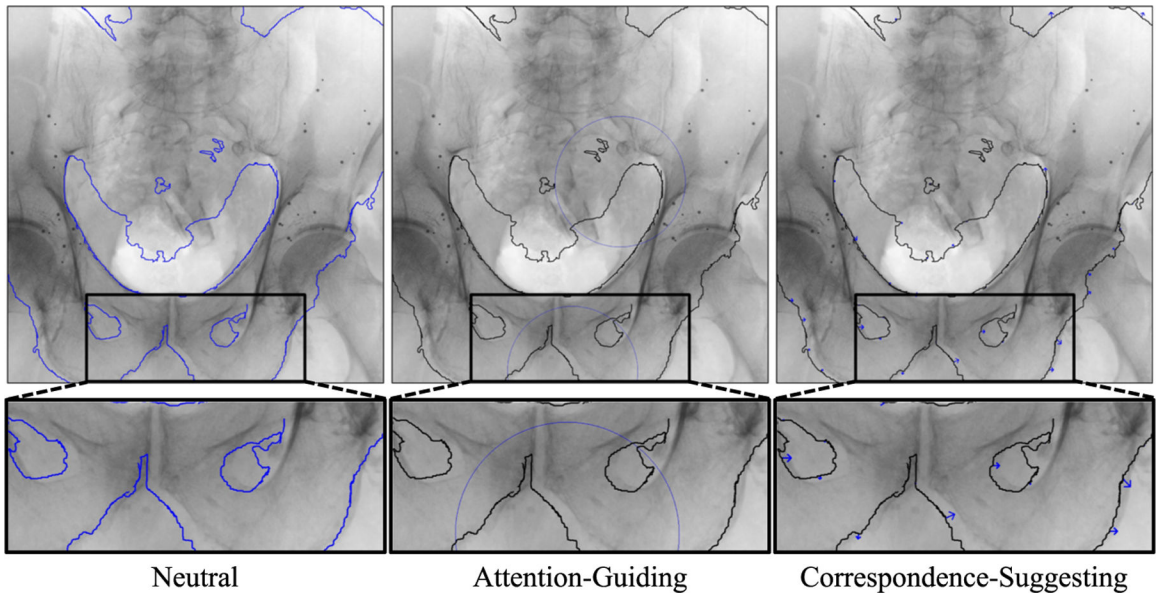

Fig. 1.
Example images of the visualization paradigms with same offset of mTRE = 6.43 mm

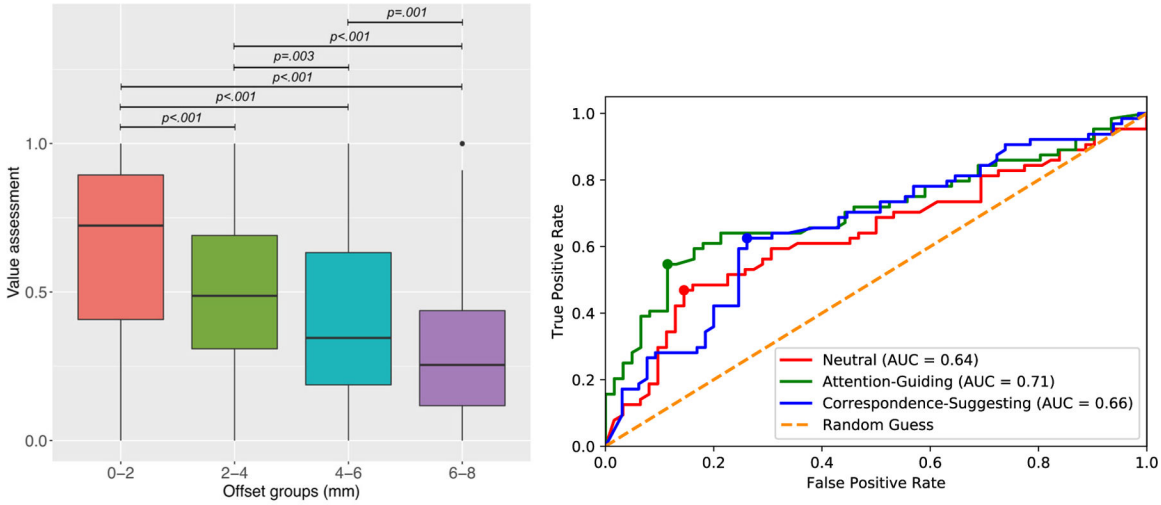

Fig. 2.

Left: Perceived assessment values of the registration under different offsets. Right: ROC plot of the three paradigms. The points refer to the optimal paradigm-specific thresholds

Table 1

Assessment performance metrics of paradigms by threshold method

Threshold	Paradigm (t: threshold)	TPR (%)	TNR (%)	ACC (%)	F1score (%)
Absolute	1 (t = 0.50)	71.9	40.3	56.4	62.6
	2 (t = 0.50)	71.9	50.8	61.6	**65.7**
	3 (t = 0.50)	64.1	66.2	**65.1**	64.6
Adaptive	1 (t = 0.80)	45.3	85.5	65.1	56.9
	2 (t = 0.72)	53.1	88.5	**70.4**	64.8
	3 (t = 0.62)	60.9	73.9	67.4	**65.0**

Bold font indicates the highest accuracy and F1 scores for each method