Marker-less Head Pose Tracking for Image-guided Cerebral Artery Navigation

Qiuying Wang[1,2,3], Pandeng Zhang[1,2], Dewei Chen[4], Hao Tang[1,2,3], Chang Liu[1,2,*], Jia Liu[1,2,*]

[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, Guangdong, China.
[2] Shenzhen Key Laboratory for Exascale Engineering and Scientific Computing, Shenzhen, China.
[3] University of Chinese Academy of Sciences, Beijing, China
[4] EDAN Instruments Inc., Shenzhen, China
`chang.liu2@siat.ac.cn, jia.liu@siat.ac.cn`

Abstract. Image-guided cerebral artery navigation (CAN) system can provide precise guidance for intracranial artery examination and surgery by aligning 3D medical data with patient's head observed by a depth sensor. Existing CAN systems generally suffer from either susceptibility to location marker offset or weak efficiency. This paper presents a real-time marker-less method to track the patient's head pose based on the MRI data for CAN. Briefly, the 3D facial model is constructed from the patient's MRI data in the pre-operative stage. Then, a 3D local description is proposed to encode the local geometry of the facial model via thin plate spline function. Subsequently, according to the local description of the facial model, the patient's head observed by an RGBD camera is registered with the facial model by maximum weight matching. Eventually, the head pose is accurately tracked in real-time via square-root cubature Kalman filter (SCKF) and iterative closest point algorithm (ICP) during navigation. With each estimated head pose, the patient's vessels in MRI data are visualized onto the RGB image of the patient's head for CAN. The proposed method is evaluated on comprehensive experiments, showing the best core performance metrics than all comparison methods. The average rotational and translational errors of our method are 2.6° and 1.9 mm respectively on the BIWI dataset. The average tracking rate achieves 0.06 s.

Keywords: Cerebral Artery Navigation, Marker-less Head Tracking, Feature Matching, 3D Local Descriptor.

1 Introduction

Image-guided cerebral artery navigation (CAN) system can determine the 6-degree-of-freedom (DOF) pose between the 3D medical data (such as MRI) and the patient's head observed by a depth sensor (such as RGBD camera), and then provide stable and reliable visual guidance for intracranial artery examination and surgery, such as transcranial

Doppler (TCD) examination [1] and vascular surgery [2]. Accordingly, visual CAN has become a crucial research topic in the field of biomedical engineering.

Existing CAN systems can be categorized into marker-based and marker-less methods. Marker-based methods align the 3D medical data with the patient's head by some fiducial markers attached to the patient's head [3], [4]. Such kind of methods can offer sub-millimeter-level pose accuracy, but they are vulnerable to marker perturbation and may lead to discomfort for patients during clinical procedures. In contrast, marker-less methods register the 3D medical data to the patient's head using per-frame global 3D registration and histogram-based local feature descriptors, such as Point Feature Histograms [5], [6], thus suitable for extensive clinical applications. However, the per-frame global 3D registration often suffers from limited real-time performance, and existing histogram-based descriptors provide relatively weak descriptiveness for flat facial regions, which easily leads to feature mismatching. Some robust strategies, such as M-estimate [7] and RANSAC [8], can improve the robustness of feature matching, but they will further reduce the efficiency of CAN.

Recently, substantial novel methods have been proposed to address marker-less head pose tracking. These methods generally fall into two categories: regression-based and optimization-based methods. Regression-based approaches train a regressor to map the face data observed by a sensor directly to the head pose [9], [10]. These methods require large training datasets for tracking accuracy and struggle with cross-sensor generalization. Furthermore, most of these methods can only provide 3-DOF head pose, which does not meet the requirement of CAN. Optimization-based approaches determine the head pose with global optimization algorithm [11], [12]. These methods have the advantage of stability but suffer from relatively low efficiency.

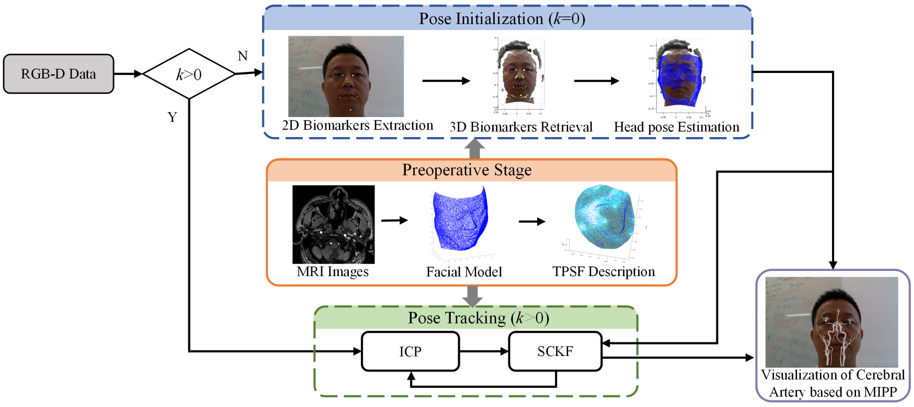


Fig. 1. Workflow of the proposed tracking system.

This paper presents a marker-less system to track the pose of patient's head with the MRI of patient's head and a calibrated RGBD camera in real-time for CAN. The workflow of the proposed system is shown in Fig. 1, where k denotes the index of time instant. In the preoperative stage, the 3D facial model is automatically constructed from patient's MRI, and the local surface around each keypoint of the facial model is described with a thin plate spline function (TPSF). At the beginning of CAN (k=0), the 3D biomarkers are retrieved from the point cloud of the patient's head observed by an RGBD camera. Then, according to the local description of model keypoints, the head pose is estimated by matching the biomarkers with model keypoints via maximum weight matching algorithm of graph. At each subsequent instant (k>0), the head pose is tracked in real time by ICP with the prediction of SCKF as initial value. After that, the SCKF gives the final maximum a posteriori estimate to the head pose and meanwhile predicts the pose of the next instant. With the final head pose, the system visualizes the cerebral arteries in MRI onto the RGB image of the patient's head for CAN by intensity perspective projection (MIPP).

2 Methodology

2.1 Coordinate System and Motion Representation

The proposed tracking system observes the patient's head with an RGBD camera, which consists of a depth camera and an RGB camera. On the depth camera and RGB camera, there exist a depth camera coordinate system ($\mathrm{CCS}_{\mathrm{depth}}$) and an RGB camera coordinate system ($\mathrm{CCS}_{\mathrm{rgb}}$) respectively. The intrinsic parameters of the RGBD camera and the transformation T_d^r from $\mathrm{CCS}_{\mathrm{depth}}$ to $\mathrm{CCS}_{\mathrm{rgb}}$ can be known a priori from calibration. Given the MRI coordinate system (MCS), the pose of the patient's head is defined as the relative transformation T_m^d from MCS to $\mathrm{CCS}_{\mathrm{depth}}$. Once T_m^d is obtained, the transformation T_m^r from MCS to $\mathrm{CCS}_{\mathrm{rgb}}$ is equal to $T_d^r T_m^d$. The cerebral vessels in MRI can be visualized and superimposed on the RGB image of the patient's head with T_m^r via MIPP. T_m^d consists of a rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$. We parameterize $\boldsymbol{R}$ in SO(3) for unambiguity [15]. Let $\boldsymbol{\Omega}$ denote the rotation vector associated with $\boldsymbol{R}$. Then, the head pose can be simply represented by a 6-dimensional pose vector $\boldsymbol{p} =[\boldsymbol{\Omega}^T \boldsymbol{t}^T]^T$.

2.2 Facial Model Construction

In the preoperative stage (see the orange box in Fig. 1), a deformable curve-based edge tracking method is employed to automatically construct a 3D facial model from the patient's MRI data. This method begins with identifying the head region in each MRI slice and then extracts the face edge from each slice by snake-like optimization. Finally, the facial model is constructed by combining all extracted face edges in order. On the facial model, several keypoints $\{\boldsymbol{q}_i, i=1, 2, \ldots, n\}$ are extracted by voxel grid downsampling [8]. This method builds a 3D voxel grid on the facial model and selects the point nearest to each voxel center, yielding uniform keypoint distribution. At each keypoint $\boldsymbol{q}_i$, a local reference frame (LRF) is established with the model points contained in the neighbor sphere of the $\boldsymbol{q}_i$ [8]. Under the LRF, we consider the local surface where $\boldsymbol{q}_i$ is located can be parameterized by a relatively flat and smooth function $z=F_i(x,y)$, in which x, y and z are the three coordinates of a point on the local surface in LRF. Accordingly, the thin plate spline function (TPSF) is quite suitable for expressing $F_i(x,y)$

$$F_i(x,y) = m_0 + m_1 x + m_2 y + \sum_{l=1}^{N} \omega_l U\left(\left\|\begin{bmatrix} x - x_l & y - y_l \end{bmatrix}\right\|\right), x \in [-L, L], y \in [-L, L], \quad (1)$$

where L denotes the supporting radius of the neighbor sphere of $\boldsymbol{q}_i$; N is the number of control points uniformly sampled on the xy-plane of LRF, each of which has the coordinate of (x_l, y_l), $(x_l, y_l) \in [-L, L] \times [-L, L]$; $U(r) = r^2 \ln r$ is the radial basis function, and $\{m_0, m_1, m_2, \omega_1, \ldots, \omega_N\}$ are the N+3 parameters of $F_i(x,y)$. In (1), L and N are hyperparameters, which will be determined in section 3.1. Given that there are m model points within the neighbor sphere of $\boldsymbol{q}_i$, each of which has the coordinate of (u_s, v_s, w_s) in LRF, s=1, 2, …, m, then the parameters of TPS can be determined by

$$\min_{\substack{m_0, m_1, m_2, \\ \omega_1, \cdots, \omega_N}} \sum_{s=1}^{m} \left\| w_s - F_i(u_s, v_s) \right\|^2. \quad (2)$$

2.3 Pose Initialization

At the start of head tracking (k=0), after acquiring the RGB image and 3D point cloud of patient's head, the system first employs a pre-trained ResNet50-Unet network [13] to extract from the RGB image the 2D facial biomarkers, including inner and outer eye corners, the nose root, philtrum, nose tip, mouth corners, and chin tip (see the blue box in Fig. 1), and then obtain the 3D biomarkers by back-projecting the 2D biomarkers onto the 3D point cloud. Supposing that there are w 3D biomarkers obtained $\{\boldsymbol{B}_j, j=1, \ldots, w\}$, the dissimilarity between the neighborhoods of the model keypoint $\boldsymbol{q}_i$ and $\boldsymbol{B}_j$ can be calculated in the LRF established at $\boldsymbol{B}_j$ as

$$\varphi_i(j) = \left\{ \sum_{h=1}^{k_j} \left(z_h - F_i(x_h, y_h)\right)^2 \right\} / k_j, \quad (3)$$

where k_j represents the number of 3D points within the spherical neighborhood of $\boldsymbol{B}_j$ with the support radius L; x_h, y_h and z_h denote the coordinates of a 3D point in the LRF of $\boldsymbol{B}_j$. We construct a bipartite graph whose two disjoint vertex sets consist of keypoints and biomarkers, respectively. The weight of each edge connecting keypoint and biomarker is computed by (3). Then, we find from this graph the keypoint-biomarker matching with minimum weight sum by blossom algorithm [14]. Supposing that each $\boldsymbol{B}_j$ is matched with a model keypoint $\boldsymbol{q}_j^*$, the confidence score that $\boldsymbol{B}_j$ corresponds to $\boldsymbol{q}_j^*$ can be calculated by

$$c_j = \exp\left(-\frac{1}{\sigma^2} \sum_{\substack{f=1, \\ f \neq j}}^{w} \left(\left\| \boldsymbol{q}_j^* - \boldsymbol{q}_f^* \right\| - \left\| \boldsymbol{B}_j - \boldsymbol{B}_f \right\| \right)^2 \right), \quad (4)$$

where σ is set as 0.01 meters. It can be easily seen that the higher the value of c_j, the more likely $\boldsymbol{B}_j$ is matched with $\boldsymbol{q}_j^*$. Accordingly, the head pose will be first computed from all keypoint-biomarker matches via least squares weighted by the confidence scores $\{c_j\}$ in (4), and then refined with facial point cloud and model points by ICP.

2.4 Real-time Head Pose Tracking

At subsequent instant k (k>0), the head pose will be determined by ICP algorithm with the pose predicted by SCKF as initial value. Then, the SCKF will provide the final maximum a posteriori estimate $\boldsymbol{p}_k$ to head pose and predict the pose of the next time instant (see the green box in Fig. 1). In our work, the SCKF is based on uniform velocity dynamic model, which assumes the acceleration of translation and rotation are noises with zero means and constant standard deviations. Let $\boldsymbol{s}_k=[\boldsymbol{\Omega}_k^T\ \boldsymbol{t}_k^T\ \boldsymbol{\omega}_k^T\ \boldsymbol{v}_k^T]^T$ be the state vector at instant k, where $\boldsymbol{\omega}_k$ and $\boldsymbol{v}_k$ denote the angular and the translational velocities respectively, and $\boldsymbol{\Omega}_k$ and $\boldsymbol{t}_k$ are the rotation vector and translation vector respectively. According to kinesics, we have the state model

$$\begin{aligned} \boldsymbol{\Omega}_{k+1} &= \boldsymbol{\Omega}_k + \boldsymbol{J}\left(\boldsymbol{\Omega}_k\right)^{-1} \boldsymbol{\omega}_k, \\ \boldsymbol{t}_{k+1} &= \boldsymbol{t}_k + \boldsymbol{v}_k, \\ \boldsymbol{\omega}_{k+1} &= \boldsymbol{\omega}_k + \boldsymbol{\mu}_k, \\ \boldsymbol{v}_{k+1} &= \boldsymbol{v}_k + \boldsymbol{\delta}_k, \end{aligned} \tag{5}$$

where $\boldsymbol{J}(\boldsymbol{\Omega}_k)$ is the left Jacobian matrix of SO(3) [15], and $\boldsymbol{\mu}_k$ and $\boldsymbol{\delta}_k$ represent the noise for the angular and the translational velocities respectively, and follow the Gaussian noise vectors with zero mean and variances of $W^2\boldsymbol{I}_3$ and $Q^2\boldsymbol{I}_3$ respectively, in which $\boldsymbol{I}_3$ denotes the 3-order identity. The measurement model of the SCKF can then be expressed as follows:

$$\boldsymbol{z}_k = [\boldsymbol{I}_6 \quad \boldsymbol{O}_6]\boldsymbol{s}_k + \boldsymbol{\varepsilon}_k, \tag{6}$$

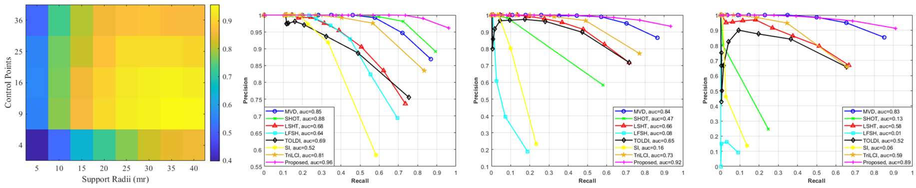
where $\boldsymbol{I}_6$ and $\boldsymbol{O}_6$ denote the 6-order identity and 6-order zero matrix respectively, and $\boldsymbol{\varepsilon}_k$ denotes the observation noise and is considered subject to a Gaussian distribution with zero mean and variance of $V^2\boldsymbol{I}_6$. When $k>0$, $\boldsymbol{z}_k$ is determined by the ICP. The procedures of the SCKF can be found in [16].

3 Experiment Results and Discussion

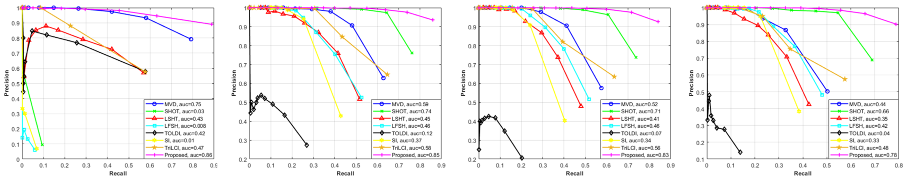
We measure the performance of the proposed method by sufficient synthetic and real trials on a desktop with 4.2GHz Intel Core i7-12700, 12 cores, and 64GB RAM. All algorithms are implemented in MATLAB.

3.1 Similarity Metric Evaluation

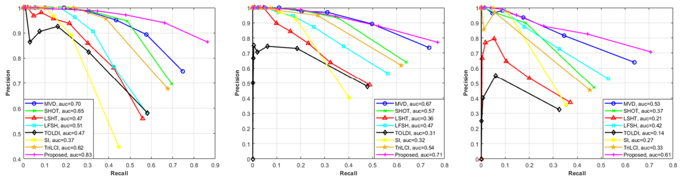
This subsection tests the robustness of the similarity metric proposed in (3) to Gaussian noise, shot noise and mesh decimation. For this purpose, five scenes are synthesized from the angel model in Queen's dataset [17] with a virtual depth sensor at different viewpoints. Therefore, the ground truth (GT) of the pose between each scene and model is known. Several local descriptors are adopted for comparison, including MVD [8], TOLDI [18], SHOT [19], LFSH [20], SI [21], LSHT [22] and TriLCI [23]. The parameters of all comparison methods are set as recommended. The performance of all methods are evaluated by the recall versus precision curve (RPC) and the area under RPC (AUC*pr*). The details of PRC and AUC*pr* can be found in [18]. In our similarity metric, the TPSF has two hyper-parameters - support radius L and the control point number N. The grid search shows that L=35 mesh resolutions (mr) and $N=3^2$ provide the best AUC*pr* (see Fig. 2 (a)).



(a) Hyper-parameter search (b) Noise free (c) Gaussian noise 0.1mr (d) Gaussian noise 0.3mr



(e) Gaussian noise 0.5mr (f) Shot noise 1% (g) Shot noise 1.5% (h) Shot noise 2%



(i) Mesh decimation 1/2 (j) Mesh decimation 1/4 (k) Mesh decimation 1/8

Fig. 2. Performance of our similarity metric and all comparison descriptor methods.

In the Gaussian noise test, we add the Gaussian noise with zero mean and the level gradually increasing from 0 mr to 0.5 mr to each scene. At each noise level, each scene is perturbed 20 times. Fig. 2 (b)-(e) plot the RPC and AUC*pr* of all methods at different noise levels. As seen, at each noise level, the precision of our metric declines gently from nearly 1 to about 0.90 as the recall rises, and almost always keep higher than that of all comparison methods. The AUC*pr* of our metric remains above 0.85 at all the noise levels and is noticeably larger than that of all the others. This trial manifests that our metric is almost not influenced by Gaussian noise and has satisfactory descriptiveness.

To assess the robustness against shot noise, we randomly sample some points from each scene with the percentage (noise level) of 1%, 1.5% and 2% respectively. These sampled points are then displaced by 20mr along their normal directions. At each noise level, each scene is perturbed 20 times. The performance of all methods under shot noise are presented in Fig. 2 (f)-(h). As observed, our metric has the greatest AUC*pr* among all methods, which slightly decreases from 0.85 to 0.78 as the noise level increases. The PRC of our metric is almost invariably higher than that of comparison methods. It is noteworthy that even at the noise level of 2%, the precision of our metric can still maintain above 0.9.

In the decimation test, we downsample each scene to 1/2, 1/4, and 1/8 of the original mesh resolution. Fig. 2 (i)-(k) give the information about the PRC and AUC*pr* of all methods. As the degree of downsampling increases, the AUC*pr* of our metric sharply drops from 0.83 to 0.61 and the precision of our metric becomes increasingly sensitive to recall value. This is expected since the scene points become increasingly less. Nonetheless, our metric achieves the largest AUC*pr* and almost the highest PRC among all methods, which validates the outstanding robustness of our metric to mesh decimation.

3.2 Head Registration

We evaluate the initialization algorithm proposed in section 2.3 on the BIWI Head Pose dataset [24] in comparison with Lu's method [6], Gsaxner's method [5] and Guo's method [8]. Both Lu's method and Gsaxner's method tackle marker-less head estimation for CAN, and Guo's method is based on MVD [8], which is the second best in the trial given in section 3.1. Furthermore, Guo's method introduces 1-Point RANSAC to improve the robustness of feature matching. The BIWI Dataset is the most widely used for head pose estimation and tracking and the details can be found in [25]. It offers the RGBD images and the 3D facial model for each subject, which can be directly used as preoperative facial models required by our method. The registration performance is assessed by relative rotation error (RRE), relative translation error (RTE), and average runtime (see [25] for details). To validate the effectiveness of the proposed biomarker-keypoint matching algorithm, we replace this algorithm with 1-Point RANSAC for comparison. The accuracy and efficiency of all methods are listed in Table 1. Among all methods, our method is the most efficient (2.38s) and achieves the best performance with an RTE of 1.9 mm and RRE of 2.6°. Our similarity metric followed by 1-Point RANSAC is the second best in terms of pose errors and efficiency. It achieves pose accuracy close to that of our method but has greatly larger time-consumption than our method. This strongly verifies the descriptiveness of our metric and the robustness of the proposed initialization method against mismatching.

Table 1. Registration errors and average runtime.

Method	RRE (degree)	RTE (mm)	Time (s)
Lu's method [6]	24.5	144.7	11.59
Gsaxner' method [5]	7.4	85.7	3.19
Guo's method [8]	15.8	8.0	192.78
Our metric+1-point RANSAC	4.2	7.1	167.32
Our initialization method	2.6	1.9	2.38

Table 2. Tracking errors and average runtime.

Method	Pitch (degree)	Yaw (degree)	Roll (degree)	RTE (mm)	Time (s)
POSEidon+ [9]	1.5	2.2	1.6	-	0.03
PGCNN [10]	1.1	1.8	1.4	-	0.02
Robust Model [11]	2.2	2.4	2.1	9.8	0.14
Ours	0.8	1.2	0.4	2.3	0.06

3.3 Head Pose Tracking

This subsection evaluates the performance of the proposed real-time head pose tracking algorithm on the BIWI dataset. Tracking performance is quantified by the average runtime, RTE, and the three Euler angle errors (pitch, yaw and roll) [9]. POSEidon+ [9], PGCNN [10] and Robust Model [11] are adopted for comparison. POSEidon+ and PGCNN are state-of-the-art head pose estimation methods based on convolutional neural networks. Robust Model deals with head pose tracking with particle swarm optimization. Table 2 demonstrates the performance of all methods. As seen, the proposed approach achieves the highest tracking accuracy, with the three Euler angle errors of approximately 0.08°, 0.12° and 0.04° respectively, and RTE of 2.3 mm, compared to all comparison methods. The runtime of our method is about 0.06 seconds/frame. Accordingly, although not the fastest, our method can satisfy the real-time requirement of CAN. In contrast, Robust Model suffers from high time-consumption, and POSEidon+ and PGCNN can only provide 3-DOF head pose, which cannot be used for CAN.

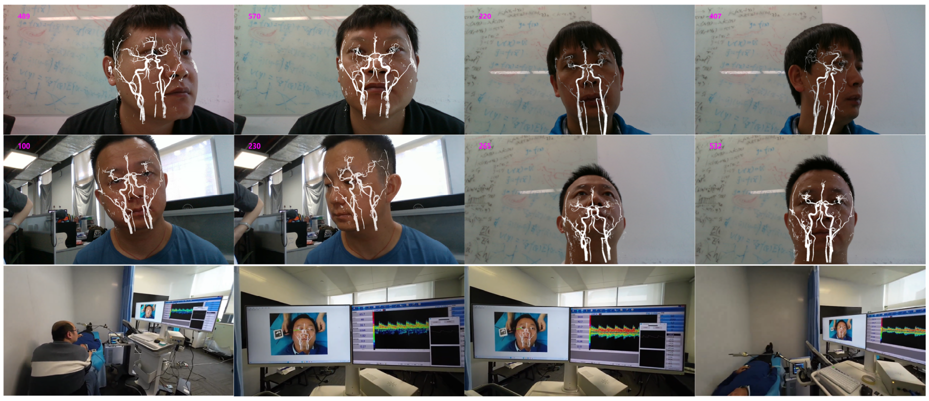


Fig. 3. Real application tests for our system.

3.4 Real Application

This subsection will exploit two real application tests to validate the effectiveness of the proposed system for CAN. The system employs an Intel RealSense D435i camera, which has the resolution of 640×480 and the frequency of 30Hz. In the first application test, several participants sit in front of the camera, and the system tracks their head poses and visualizes their cerebral vessels in real time. Some examples are demonstrated in the first two rows of Fig. 3, where the pink number denotes the frame index. As seen, our system can maintain stable head pose tracking even under significant head movement. In the second application test, our system simultaneously tracks the poses of the participant's head and a marked TCD scanner in real time and visualizes the cerebral vessels and the ultrasound beam for TCD examination navigation. The video provided in the supplementary material records the whole process of TCD navigation. Some example frames of the video are shown in the bottom row in Fig. 3, where the green line represents the ultrasound beam. This test manifests that our system can help the TCD operator localize the vessel quickly.

4 Conclusion

This paper presents a marker-less head pose tracking system for CAN. In the preoperative stage, the system automatically constructs the 3D facial model of patient from the MRI data of patient's head. During the online stage, the system first robustly registers the 3D facial model with patient's head via blossom algorithm, and then continuously tracks the 6-DOF head pose in real-time by ICP and SCKF. The estimated pose enables real-time visualization of cerebral vessels through MIPP for intuitive CAN guidance. Substantial trials validate the effectiveness of our system in both accuracy and efficiency. While promising, the system has several limitations. First, this work assumes rigid facial surfaces during neurosurgery or TCD procedures, without consideration of facial deformation. Second, the impact of varying biomarker number needs further analysis. In our future work, we will focus on these limitations and evaluate the navigation system in collaboration with clinicians.

Acknowledgments. This research is financed by Shenzhen Science and Technology Innovation Commission (Grant No. KJZD20230923114609020) and National Natural Science Foundation of China (Grant No. 62273324).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. A. Neulen, C. Greke, E. Prokesch, J. König, D. Wertheimer, and A. Giese, "Image guidance to improve reliability and data integrity of transcranial Doppler sonography," Clin. Neurol. Neurosurg., vol. 115, no. 8, pp. 1382–1388, Aug. 2013, doi: 10.1016/j.clineuro.2012.12.025.
2. S. Eom, S. Kim, S. Rahimpour, and M. Gorlatova, "AR-Assisted Surgical Guidance System for Ventriculostomy," in 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Mar. 2022, pp. 402–405. doi: 10.1109/VRW55335.2022.00087.
3. M. Wang and Z. Song, "Automatic localization of the center of fiducial markers in 3D CT/MRI images for image-guided neurosurgery," Pattern Recognit. Lett., vol. 30, no. 4, pp. 414–420, Mar. 2009, doi: 10.1016/j.patrec.2008.11.001.
4. A. Neulen, "Image-Guided Transcranial Doppler Ultrasound for Monitoring Posthemorrhagic Vasospasms of Infratentorial Arteries: A Feasibility Study," WORLD Neurosurg., 2020.
5. C. Gsaxner, A. Pepe, J. Wallner, D. Schmalstieg, and J. Egger, "Markerless Image-to-Face Registration for Untethered Augmented Reality in Head and Neck Surgery," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds., Cham: Springer International Publishing, 2019, pp. 236–244. doi: 10.1007/978-3-030-32254-0_27.

6. S. Lu, Y. Luo, D. Xie, W. Wang, W. Zheng, and J. Gong, "Spatial Mapping Method of Craniosynostosis Surgical Robot Based on Point Cloud Registration," in 2021 6th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), Jul. 2021, pp. 1–6. doi: 10.1109/ACIRS52449.2021.9519346.
7. Q.-Y. Zhou, J. Park, and V. Koltun, "Fast Global Registration," in Computer Vision – ECCV 2016, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham: Springer International Publishing, 2016, pp. 766–782. doi: 10.1007/978-3-319-46475-6_47.
8. W. Guo, W. Hu, C. Liu, and T. Lu, "3D object recognition from cluttered and occluded scenes with a compact local feature," Mach. Vis. Appl., vol. 30, no. 4, pp. 763–783, Jun. 2019, doi: 10.1007/s00138-019-01027-7.
9. G. Borghi, M. Fabbri, R. Vezzani, S. Calderara, and R. Cucchiara, "Face-from-Depth for Head Pose Estimation on Depth Images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 3, pp. 596–609, Mar. 2020, doi: 10.1109/TPAMI.2018.2885472.
10. Y. Xu, C. Jung, and Y. Chang, "Head pose estimation using deep neural networks and 3D point clouds," Pattern Recognit., vol. 121, p. 108210, Jan. 2022, doi: 10.1016/j.patcog.2021.108210.
11. G. P. Meyer, S. Gupta, I. Frosio, D. Reddy, and J. Kautz, "Robust Model-Based 3D Head Pose Estimation," in 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile: IEEE, Dec. 2015, pp. 3649–3657. doi: 10.1109/ICCV.2015.416.
12. P. Padeleris, X. Zabulis, and A. A. Argyros, "Head pose estimation on depth data based on Particle Swarm Optimization," in 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Jun. 2012, pp. 42–49. doi: 10.1109/CVPRW.2012.6239236.
13. F. Aboudi, C. Drissi, and T. Kraiem, "A Hybrid Model for Ischemic Stroke Brain Segmentation from MRI Images using CBAM and ResNet50-Unet," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 2, 2024, doi: 10.14569/IJACSA.2024.0150296.
14. J. Edmonds, "Maximum matching and a polyhedron with 0,1-vertices," J. Res. Natl. Bur. Stand. Sect. B Math. Math. Phys., vol. 69B, no. 1 and 2, p. 125, Jan. 1965, doi: 10.6028/jres.069B.013.
15. T. D. Barfoot, State Estimation for Robotics, 1st ed. Cambridge University Press, 2017. doi: 10.1017/9781316671528.
16. H. Tang, C. Liu, Y. Su, Q. Wang, and W. Hu, "Model-based monocular 6-degree-of-freedom pose tracking for asteroid," Front. Space Technol., vol. 5, p. 1337262, Feb. 2024, doi: 10.3389/frspt.2024.1337262.
17. B. Taati and M. Greenspan, "Local shape descriptor selection for object recognition in range data," Comput. Vis. Image Underst., vol. 115, no. 5, pp. 681–694, May 2011, doi: 10.1016/j.cviu.2010.11.021.
18. J. Yang, Q. Zhang, Y. Xiao, and Z. Cao, "TOLDI: An effective and robust approach for 3D local shape description," Pattern Recognit., vol. 65, pp. 175–187, May 2017, doi: 10.1016/j.patcog.2016.11.019.
19. S. Salti, F. Tombari, and L. Di Stefano, "SHOT: Unique signatures of histograms for surface and texture description," Comput. Vis. Image Underst., vol. 125, pp. 251–264, Aug. 2014, doi: 10.1016/j.cviu.2014.04.011.
20. J. Yang, Z. Cao, and Q. Zhang, "A fast and robust local descriptor for 3D point cloud registration," Inf. Sci., vol. 346–347, pp. 163–179, Jun. 2016, doi: 10.1016/j.ins.2016.01.095.
21. A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999, doi: 10.1109/34.765655.

22. C. Hui, R. Wang, X. Wen, J. Zhao, W. Chen, and X. Zhang, “A Novel Recognition Algorithm in 3D Point Clouds based for on Local Spherical Harmonics,” in 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China: IEEE, Aug. 2019, pp. 1041–1046. doi: 10.1109/ICMA.2019.8816499.
23. W. Tao, X. Hua, X. He, J. Liu, and D. Xu, “Automatic multi-view registration of point clouds via a high-quality descriptor and a novel 3D transformation estimation technique,” Vis. Comput., vol. 40, no. 4, pp. 2615–2630, Apr. 2024, doi: 10.1007/s00371-023-02942-7.
24. G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. Van Gool, “Random Forests for Real Time 3D Face Analysis,” Int. J. Comput. Vis., vol. 101, no. 3, pp. 437–458, Feb. 2013, doi: 10.1007/s11263-012-0549-0.
25. H. Yu, F. Li, M. Saleh, B. Busam, and S. Ilic, “CoFiNet: Reliable Coarse-to-fine Correspondences for Robust PointCloud Registration,” in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2021, pp. 23872–23884. Accessed: Oct. 15, 2024. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/hash/c85b2ea9a678e74fdc8bafe5d0707c31-Abstract.html