J. Shanghai Jiaotong Univ. (Sci.), 2014, 19(5): 617-623
DOI: 10.1007/s12204-014-1550-2

Kinect-Based Automatic Spatial Registration Framework for Neurosurgical Navigation

ZHANG Li-xia[1] (张丽霞), *ZHANG Shao-ting*[2] (张少霆), *XIE Hong-zhi*[3] (谢洪智)
ZHUANG Xia-hai[4] (庄吓海), *GU Li-xu*[1*] (顾力栩)
(1. School of Biomedical Engineering, Shanghai Jiaotong University, Shanghai 200030, China; 2. Department of Computer Science, University of North Carolina at Charlotte, NC 28223, USA; 3. Peking Union Medical College Hospital, Beijing 100730, China; 4. Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China)

Abstract: As image-guided navigation plays an important role in neurosurgery, the spatial registration mapping the pre-operative images with the intra-operative patient position becomes crucial for a high accurate surgical output. Conventional landmark-based registration requires expensive and time-consuming logistic support. Surface-based registration is a plausible alternative due to its simplicity and efficacy. In this paper, we propose a comprehensive framework for surface-based registration in neurosurgical navigation, where Kinect is used to automatically acquire patient's facial surface in a real time manner. Coherent point drift (CPD) algorithm is employed to register the facial surface with pre-operative images (e.g., computed tomography (CT) or magnetic resonance imaging (MRI)) using a coarse-to-fine scheme. The spatial registration results of 6 volunteers demonstrate that the proposed framework has potential for clinical use.
Key words: surface-based registration, neurosurgical navigation, coherent point drift (CPD)
CLC number: TP 391.7 **Document code:** A

0 Introduction

Image-guided surgery has been widely used in operating room and benefited due to its time-saving and minimally invasive features, especially in the area of neurosurgery. One of the most important tasks is to align the intra-operative patient's position with the preoperatively acquired diagnostic image series (e.g., computed tomography (CT) or magnetic resonance imaging (MRI)), which shows the patient's internal structures. With a surgical instrument mapped with the preoperative images, the surgeon can avoid damaging delicate structures during a treatment[1].

Landmark-based registration is the most popular approach used in routine clinic[1]. However, artificial landmarks are not convenient, since they are usually stuck to patients' skin before acquiring the image data, and they have to be fixed in position during all the navigation registration procedures. Anatomical landmarks are difficult to be defined, especially for inexperienced surgeons. Another optional registration approach is based on the markerless surface. In many image-guided neurosurgeries, 3D probes (e.g., electromagnetic, optical) or laser scanners are used to acquire the geometric surface points on the patient's face in patient space[2]. However, it is time-consuming to gather surface information with probes contacting patient's face and it may introduce unexpected human operation error. Laser scanners can get the whole surface of patient's face with contact-free operation, and using a high-resolution laser scanner for spatial registration achieves promising results[3-4]. However, high-resolution laser scanners are expensive and the surface collection also involves moving the device freehandedly by a surgeon[5].

During the neurosurgery the patient's head position may be changed, so the alignment mapping patient space with image space is affected. In routine clinic, a reference panel is attached to the patient's head. Therefore, the reference's transformation can be detected and calculated by the optical navigator instantly. Afterwards, the previous spatial registration is updated. However, the reference panel may be occluded by the surgeon, and cannot be detected by the tracking device.

In this paper, our main contribution is to design a comprehensive framework to solve these problems in neurosurgical navigation. We choose Kinect[6], which

Received date: 2014-05-15
Foundation item: the National Natural Science Foundation of China (Nos. 61190120, 61190124 and 61271318) and the Biomedical Engineering Fund of Shanghai Jiaotong University (No. YG2012ZD06)
∗**E-mail:** gulixu@sjtu.edu.cn

is a hands-free device, to obtain the facial surface of a patient in real time automatically, and register it with the pre-operative images. Kinect enables us to conduct registration during neurosurgery. The surface data sets for registration are from different modalities. Therefore, we use a coarse-to-fine scheme for robust and efficient registration. Experimental results of 6 volunteers demonstrate that the proposed framework has potential for clinical use.

1 Method

1.1 Framework

Optical navigator is used as tracking device (Fig. 1). It can track position of optical sensor which is attached to the surgical device during a surgery. A surface-acquiring device is used to sample facial points of patient. In this study, we use Kinect to get the facial surface. Kinect integrates a color camera to provide color image, an infrared (IR) emitter to emit 830 nm IR light beams and an IR camera to read the IR beams reflected back to the sensor. The reflected beams are converted into depth information measuring the distance between the subject and sensor[6]. Each depth frame is presented as a low-resolution image with each gray value related to the distance. Depth image data can be transformed into 3D data stored as (x, y, z). The x, y and z axes make up a right-handed coordinate system (Fig. 1). The quality of the 3D data is influenced by the low resolution of Kinect[6].

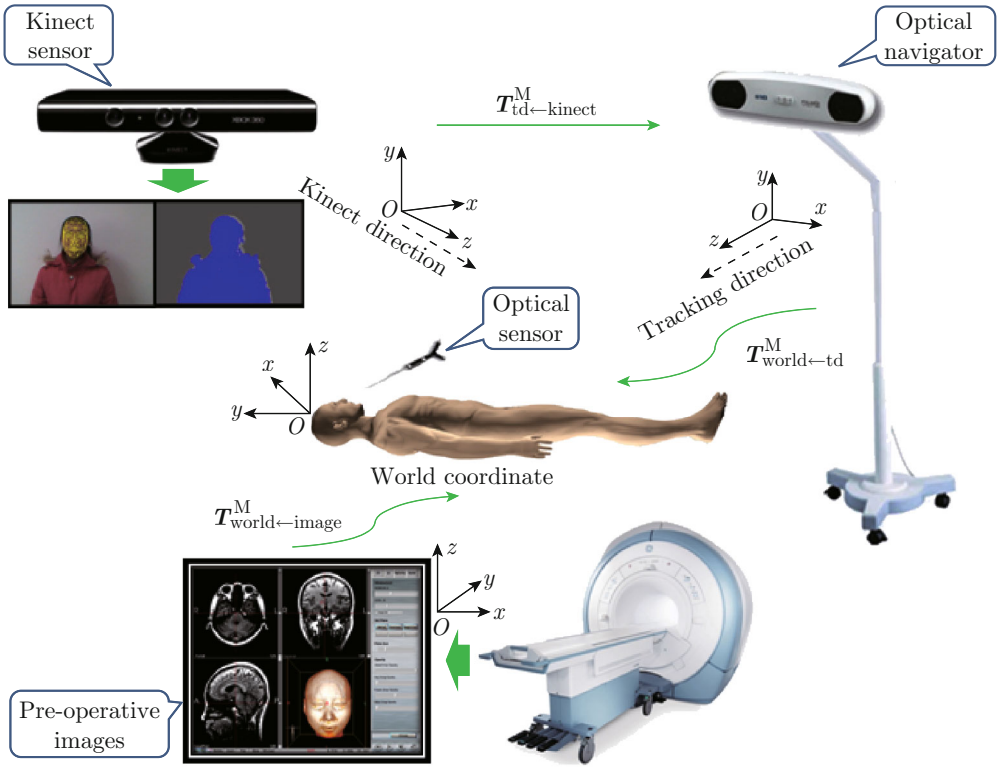


Fig. 1 Neurosurgical navigation system and coordinate transformations of spatial registration

In Fig. 1, coordinate transformations involved in the spatial registration are shown. The calibration is necessary to transform Kinect coordinate to that of the tracking device. A popular approach is to find the correspondence $\boldsymbol{T}^{\mathrm{M}}_{\mathrm{td}\leftarrow\mathrm{kinect}}$ between checkerboard's corner points from two devices, while capturing a moving point over time is an alternative way[7]. Then, $\boldsymbol{T}^{\mathrm{M}}_{\mathrm{world}\leftarrow\mathrm{td}}$ is provided by tracking device mapping its coordinate to world coordinate. $\boldsymbol{T}^{\mathrm{M}}_{\mathrm{world}\leftarrow\mathrm{image}}$, which is defined by medical instrument acquiring pre-operative images, aligns world coordinate with image coordinate. The key point of this study is to achieve registration between Kinect data and image data within world coordinate.

1.2 Data Acquisition and Face Segmentation

Figures 2(a) and 2(c) show our color and depth images obtained from Kinect sensor. In this study, we divide the facial data acquisition into 6 steps:

(1) Segment human body from depth image using Kinect software development kit (SDK) in order to check whether there is a human in front of the sensor. As shown in Fig. 2(b), the human is labeled in blue.

(2) Convert the human body's depth data to 3D data in depth camera coordinate by Kinect SDK.

(3) Track human's face on the color image using a tracker from Kinect SDK (Fig. 2(d)).

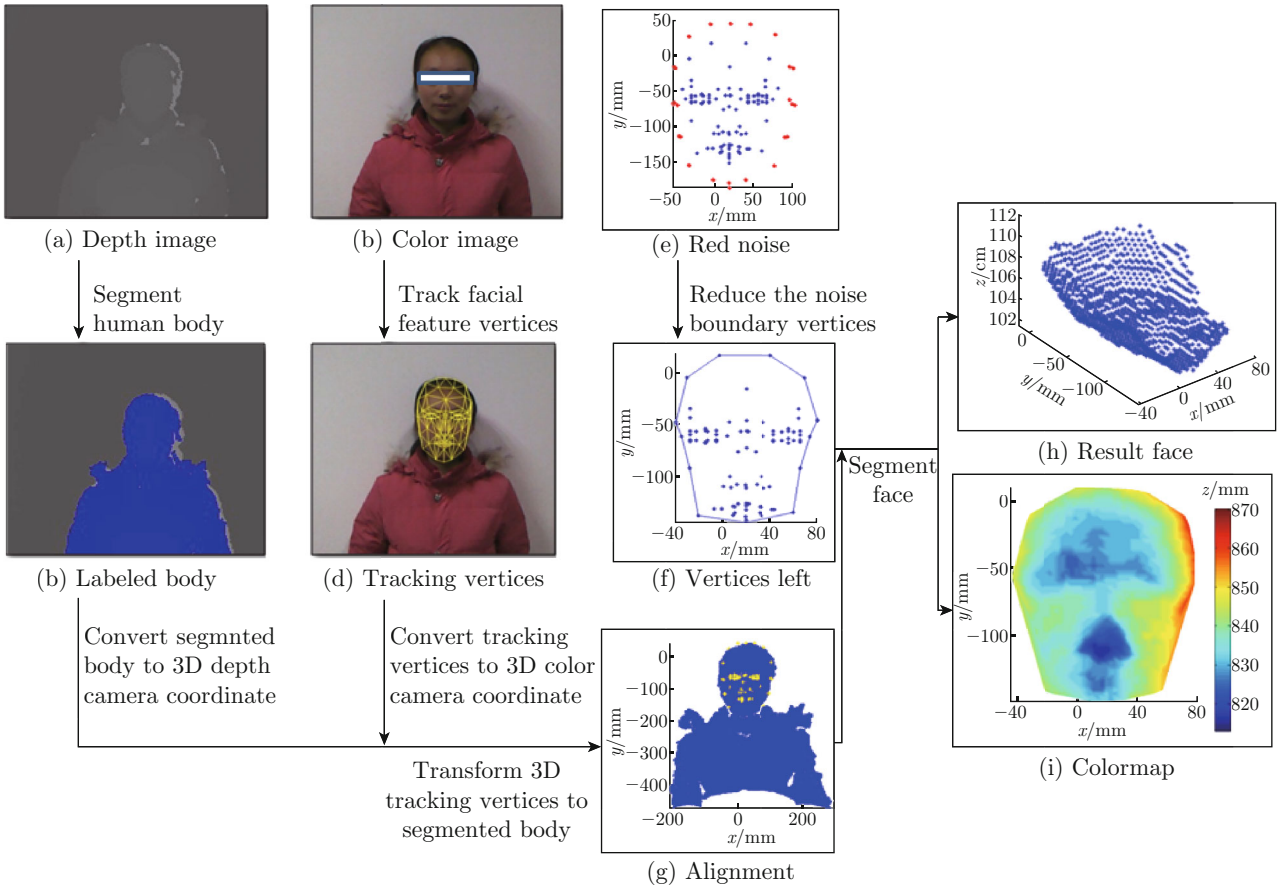


Fig. 2 Kinect imaging and face segmentation

(4) Convert the tracking vertices to 3D data in Kinect color camera coordinate.

(5) Transform the 3D tracking results in Kinect color camera coordinate to depth camera coordinate (Fig. 2(g)). The vertices determine the location of human's face in Kinect depth camera coordinate.

(6) Segment human facial surface from body limited to boundary of the 3D tracking vertices.

However, tracking results may contain noise around boundary vertices, such as human's hairs which are not features of face (Fig. 2(d)). So we reduced the noise (red vertices in Fig. 2(e)). Then, we join adjacent boundary vertices of the rest using linear interpolation in xOy plane and generate a closed curve (Fig. 2(f)). The facial points whose xOy data are inside the curve are finally segmented (Fig. 2(h)). Figure 2(i) shows colormap of the segmented face, where red means distance between Kinect and the human is large and blue denotes the distance is small.

For pre-operative image series, firstly Otsu algorithm[8] is applied to the patient's head MRI/CT. After the threshold, the pixels inside the head surface are labeled with 1 and outside with 0. Then, we search along rays casting frontally through MRI/CT volume, and save the edge points with intensity changed from 0 to 1. By integrating the edge points, patient's facial surface is extracted from MRI/CT.

1.3 Coarse-to-Fine Registration

Direct registration between the pre- and intra-operative data may result in local minimum and inaccuracy, since Kinect data are low-resolution and asymmetry with MRI/CT data. Therefore, we design a coarse-to-fine registration scheme to improve the robustness.

1.3.1 *Coarse-Registration*

We employ principal component analysis (PCA) to define a bounding box with three main oriented axes for two sets of facial points, respectively. PCA does eigendecomposition of data's covariance matrix. Eigenvectors that correspond to the three largest eigenvalues define oriented axes of the points. The origin is at the mean point. Then, we assume that $\boldsymbol{T}_{\text{coarse}}^{\text{M}} = [\boldsymbol{\lambda}_1 \quad \boldsymbol{\lambda}_2 \quad \boldsymbol{\lambda}_3 \quad \boldsymbol{\lambda}_4]^{\text{T}}$, where $\boldsymbol{\lambda}_i\,(i = 1, 2, 3, 4)$ are 4D vectors and $\boldsymbol{E}(\boldsymbol{X})$ represents the diagonal matrix with $\text{diag}(\boldsymbol{X}, \boldsymbol{X}, \boldsymbol{X}, \boldsymbol{X})$. The transformation is defined as:

$$\begin{bmatrix} \boldsymbol{E}(\boldsymbol{\mu}_1) \\ \boldsymbol{E}(\boldsymbol{\mu}_2) \\ \boldsymbol{E}(\boldsymbol{\mu}_3) \\ \boldsymbol{E}(\boldsymbol{\mu}_4) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \\ \boldsymbol{\lambda}_3 \\ \boldsymbol{\lambda}_4 \end{bmatrix} \begin{bmatrix} \boldsymbol{E}(\boldsymbol{\nu}_1) \\ \boldsymbol{E}(\boldsymbol{\nu}_2) \\ \boldsymbol{E}(\boldsymbol{\nu}_3) \\ \boldsymbol{E}(\boldsymbol{\nu}_4) \end{bmatrix}, \tag{1}$$

where $\boldsymbol{\mu}_i$ $(i = 1, 2, 3, 4)$ and $\boldsymbol{\nu}_i$ $(i = 1, 2, 3, 4)$ contain eigenvectors and centroids of two surface data sets, respectively. Therefore, we can obtain the transformation matrix $\boldsymbol{T}^{\mathrm{M}}_{\mathrm{coarse}}$ by solving the equation set. This is coarse registration which can provide surface registration with a fine start to converge quickly.

1.3.2 *Fine-Registration*

Iterative closest point (ICP) is a natural choice because of its simplicity and low computational complexity. In recent years, many methods have been proposed for robust point set registration, which are able to handle noise, outliers, and occlusions[9-12]. However, most of them are not fast enough for our clinical application. The algorithm coherent point drift (CPD) is both robust and efficient, owing to the use of fast Gauss transform and low-rank matrix approximation to reduce the computational complexity[13]. CPD explicitly formulates the alignment of two point sets as a maximum likelihood (ML) estimation problem of Gaussian mixture model (GMM)[13]. If we align point data set S to T, s and t represent one point from S and T respectively, CPD regards the transformation $\phi(s)$ as GMM centroids, and T are points generated by GMM. Let M and N be point numbers of S and T. The GMM probability density function is:

$$p(t) = \sum_{m=1}^{M+1} P(m)p(t|m), \tag{2}$$

where $P(m)$ is a weight coefficient; $p(t|m) = \frac{1}{(2\pi\sigma^2)^{D/2}} \exp^{-\frac{\|t-\phi(s)\|^2}{2\sigma^2}}$, D represents dimension of point sets and σ^2 is covariance parameter. Especially, the uniform distribution $p(t|M+1) = 1/N$ accounts for noise and outliers whose weight is $P(M+1) = \omega$. We set $P(m) = (1-\omega)1/M\,(m = 1, 2, \cdots, M)$. It indicates that $\omega = 1$ means only noise and outliers remained, $\omega = 0$ represents no noise or outliers. The transformation is calculated by maximizing likelihood function using Expectation Maximization (EM) algorithm.

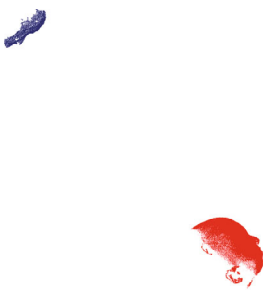
2 Experiments and Results

2.1 Work Platform and Surface Acquisition

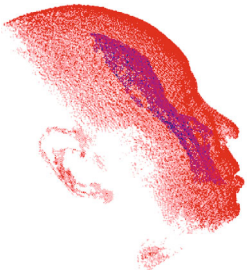
We tested our surface-based spatial registration approach with 6 volunteers. MRI examinations were performed at Renji Hospital (Shanghai, China) using GE 3.0T Signa HDxt machine (GE Medical Systems, USA). For each volunteer, MRI head scan were taken with 512 pixel × 512 pixel × 252 pixel and spatial resolution was 0.429 7 mm × 0.429 7 mm × 1 mm. Kinect resolutions were 640 pixel × 480 pixel for color image and 320 pixel × 240 pixel for depth image. The frames were fed into a PC (32-bit windows, Pentium(R) Dual-Core CPU, 3GB RAM). Surface contains nearly 71 000 points acquired from MRI and 1 600 points from Kinect, where point coordinate unit is millimeter. Two surfaces are not symmetry, where most MRI points do not have corresponding Kinect points.

2.2 Coarse-to-Fine Registration Results

Firstly, coarse registration was performed by aligning two oriented bounding boxes of the two facial point sets. In Fig. 3, the blue surface is from Kinect and the red one is from MRI.



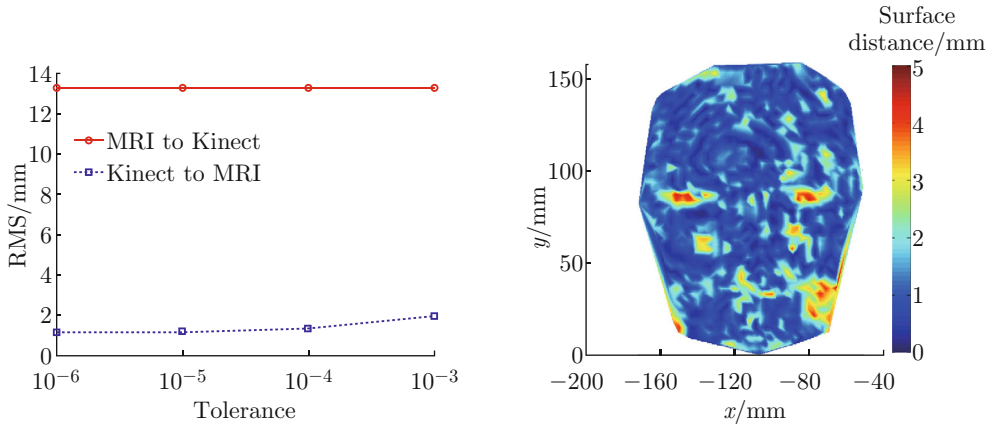
(a) Initial position of two surfaces



(b) Position after coarse registration

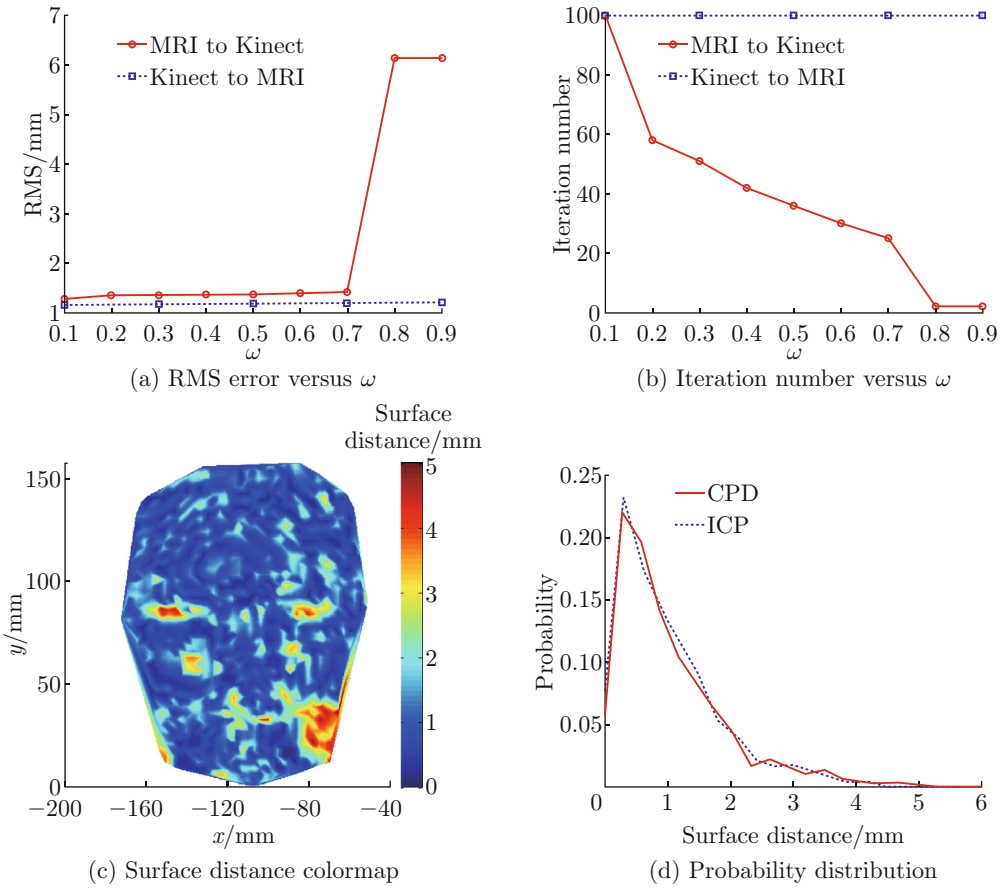
Fig. 3 Performance of coarse registration

For fine registration, ICP usually uses the data set with small number to search for correspondences in the data set with large number[14]. Here we use root mean square (RMS) error to evaluate the registration, which is defined as mean of the distances between the aligned point correspondences from Kinect and MRI[14]. In Fig. 4(a), tolerance stands for stopping criterion, and we choose 100 for maximum number of iterations. It shows that ICP performs better aligning Kinect data to MRI data (blue line). In this paper, we choose tolerance as 10^{-5} for ICP, and Fig. 4(b) shows surface distance colormap aligning data from Kinect to MRI, where red means the registration error is large and blue denotes the error is small. For CPD algorithm, it's necessary to choose an appropriate value for ω, which is standing for the weight of noise and outliers. Figures 5(a) and 5(b) show RMS error and iteration number of registration versus ω respectively, with tolerance as 10^{-5}. We choose 100 for maximum number of iterations. According to Fig. 5(a) CPD performs better when aligning Kinect to MRI (blue line) than MRI to Kinect (red line), here we choose Kinect to MRI. The plot of RMS error versus ω and the plot of iteration number versus ω are almost constants (blue lines (dash lines) in



(a) RMS error versus tolerance (b) Surface distance colormap

Fig. 4 Performance of ICP



(a) RMS error versus ω (b) Iteration number versus ω

(c) Surface distance colormap (d) Probability distribution

Fig. 5 Performance of CPD

Figs. 5(a) and 5(b)). Therefore, we choose 0.5 for ω.

According to the blue lines in Figs. 4(a) and 5(a), the RMS error are similar using CPD and ICP aligning Kinect to MRI. For more comparison, the surface distance colormap using CPD is shown in Fig. 5(c) with ω as 0.5, tolerance as 10^{-5}. Figure 5(d) shows surface distance probability distributions of results in Figs. 4(b) and 5(c). Finally, 6 volunteer experiments were performed to evaluate the accuracy of the proposed coarse-to-fine registration method. Table 1 shows the registration RMS error and the probability of surface distances (< 1 mm) using CPD and ICP. The mean RMS of CPD-based and ICP-based cases are (1.2 ± 0.1) mm and (1.1 ± 0.1) mm respectively. The mean distance (< 1 mm) probability of the CPD-based and the ICP-based cases are $0.525\,0 \pm 0.053\,3$ and $0.528\,4 \pm 0.044\,1$

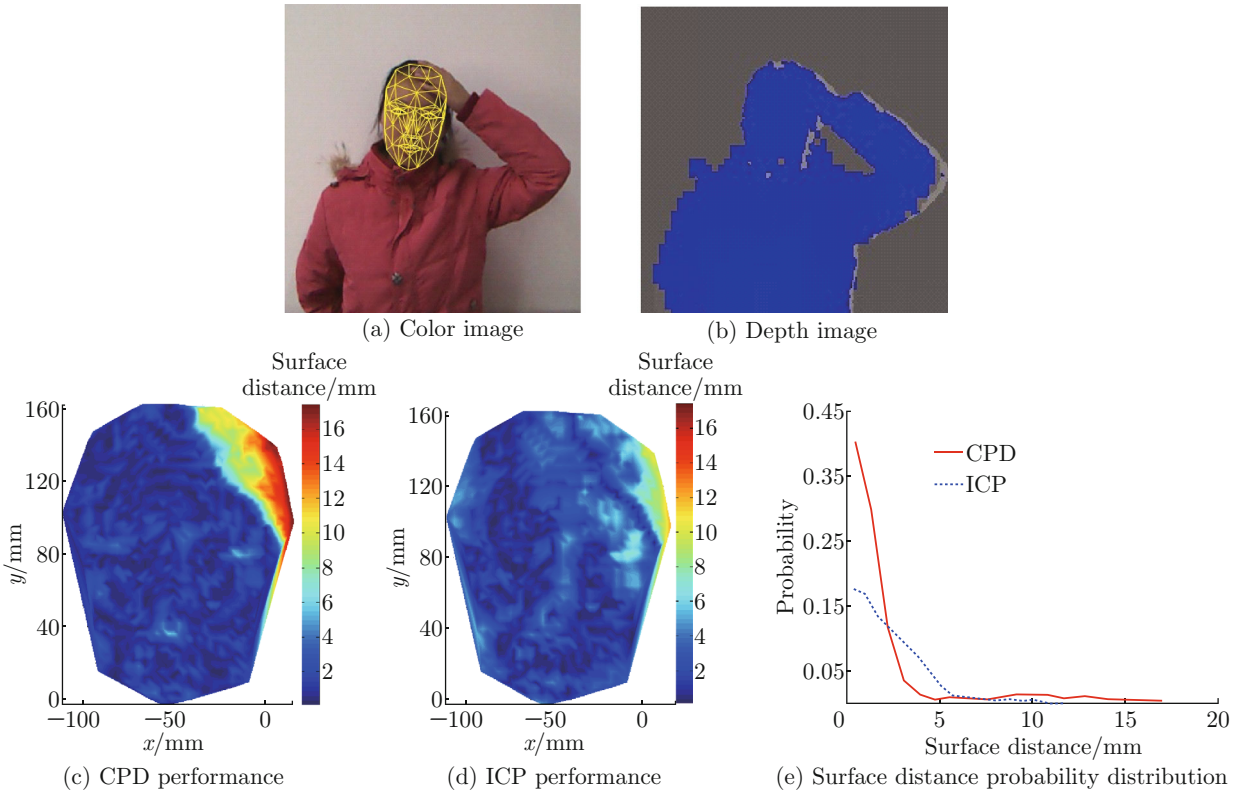
Table 1 RMS error and distance (< 1 mm) probability using CPD and ICP for 6 volunteers

Volunteer	RMS/mm		Probability	
	CPD	ICP	CPD	ICP
1	1.1	1.1	0.572 8	0.547 0
2	1.3	1.2	0.472 7	0.472 7
3	1.0	1.0	0.596 8	0.601 5
4	1.1	1.1	0.534 8	0.524 8
5	1.3	1.2	0.464 7	0.498 1
6	1.1	1.1	0.508 3	0.526 4
Mean ± Std	1.2 ± 0.1	1.1 ± 0.1	0.525 0 ± 0.053 3	0.528 4 ± 0.044 1

respectively. It demonstrates that for our surface registration CPD-based case performs similar to ICP-based one.

2.3 Occlusion Situation

During the neurosurgery the reference panel may be occluded by the surgeon's hand, so the patient's head movement cannot be detected by the tracking device, which may influence the navigation accuracy significantly. Figures 6(a) and 6(b) show the Kinect color and depth images of a volunteer with her forehead occluded by her hand. Kinect can recognize the patient's face with occlusion noise. Then, the robustness of the proposed registration approach based on CPD can handle such a hand occlusion situation, while ICP-based registration is influenced by the hand occlusion (Figs. 6(c) and 6(d), where red means the registration error is large and blue denotes the error is small). Figure 6(e) shows the surface distance probability distribution of RMS error in Figs. 6(c) and 6(d). We can see that the surface distances in the CPD-based case tend to distribute on the small-value portion by comparison. It demonstrates that CPD is more suitable for our surface registration with different situations.



(a) Color image (b) Depth image

(c) CPD performance (d) ICP performance (e) Surface distance probability distribution

Fig. 6 Occlusion situation

3 Discussion

In modern operating room, laser scanner navigation system has been used for neurosurgical surgery. Although it can get patient's facial surface precisely, human interaction is inevitable. Kinect sensor can get patient's facial surface automatically with no human interaction. Meanwhile we provide a novel method to

get the patient's facial points from Kinect color and depth images. During the neurosurgery, Kinect can robustly track the human's face with occlusion noise following head movement, and the registration is updated using CPD-based coarse-to-fine registration method. The proposed coarse-to-fine approach is tested to be more robust and accurate comparing with other traditional algorithms.

After registration has been performed, image space and patient space can be mapped together. However, in clinical applications, the two spaces will always differ from each other slightly[5]. In general, the clinical error is calculated when the surgeon approaches a target during the surgery. Z-touch is now a commercial handheld scanner for surface registration manufactured by BrainLab^{TM} with clinically applied precision of (1.8 ± 0.5) mm[4]. Another known navigation system is Surgical Segment Navigator developed by University of Heidelberg, which can achieve a clinical accuracy of (1.1 ± 0.3) mm[3]. In our study, Kinect-based spatial registration is still in the stage of research and has not yet been applied to any clinical use. However, our registration RMS is (1.2 ± 0.1) mm, which is comparable to the RMS of Z-touch in (1.21 ± 0.34) mm[4].

4 Conclusion

In this paper, we propose a new framework for spatial registration in neurosurgical navigation system based on Kinect. Kinect is employed to reconstruct patient facial surface in patient space automatically, where CPD-based coarse-to-fine registration algorithm is applied for alignment between the Kinect surface and MRI surface. Experiments show that our registration accuracy is promising. Although the quality of 3D data generated by Kinect is low[6], many researches on reconstruction using Kinect have achieved equal accuracy of laser scanners in recent years[15-16]. For future work, we want to integrate Kinect in our navigation system. The system can acquire human facial points automatically, and finally align the patient space with pre-operative image space before and during the neurosurgery.

References

[1] Peters T M. Image-guidance for surgical procedures [J]. *Physics in Medicine and Biology*, 2006, **51**(14): R505-R540.

[2] Maurer Jr C R, Maciunas R J, Michael Fitzpatrick J. Registration of head CT images to physical space using a weighted combination of points and surfaces [J]. *IEEE Transactions on Medical Imaging*, 1998, **17**(5): 753-761.

[3] Marmulla R, Hassfeld S, Lüth T, et al. Laser-scan-based navigation in cranio-maxillofacial surgery [J]. *Journal of Cranio-Maxillofacial Surgery*, 2003, **31**(5): 267-277.

[4] Hoffmann J, Westendorff C, Leitner C, et al. Validation of 3D-laser surface registration for image-guided cranio-maxillofacial surgery [J]. *Journal of Cranio-Maxillofacial Surgery*, 2005, **33**(1): 13-18.

[5] Eggers G, Mühling J, Marmulla R. Image-to-patient registration techniques in head surgery [J]. *International Journal of Oral and Maxillofacial Surgery*, 2006, **35**(12): 1081-1095.

[6] Khoshelham K, Elberink S O. Accuracy and resolution of Kinect depth data for indoor mapping applications [J]. *Sensors*, 2012, **12**(2): 1437-1454.

[7] Berger K, Ruhl K, Schroeder Y, et al. Markerless motion capture using multiple color-depth sensors [C]// *Vision Modeling Visualization*. Bangor, Wales: Eurographics Association, 2011: 317-324.

[8] Otsu N. A threshold selection method from gray-level histograms [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, **9**(1): 62-66.

[9] Li H, Shen T, Huang X. Approximately global optimization for robust alignment of generalized shapes [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(6): 1116-1131.

[10] Jian B, Vemuri B C. Robust point set registration using Gaussian mixture models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(8): 1633-1645.

[11] Zhang S, Zhan Y, Dewan M, et al. Towards robust and effective shape modeling: Sparse shape composition [J]. *Medical Image Analysis*, 2012, **16**(1): 265-277.

[12] Zhang S, Zhan Y, Cui X, et al. 3D anatomical shape atlas construction using mesh quality preserved deformable models [J]. *Computer Vision and Image Understanding*, 2013, **117**(9): 1061-1071.

[13] Myronenko A, Song X. Point set registration: Coherent point drift [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(12): 2262-2275.

[14] Salvi J, Matabosch C, Fofi D, et al. A review of recent range image registration methods with accuracy evaluation [J]. *Image and Vision Computing*, 2007, **25**(5): 578-596.

[15] Hernandez M, Choi J, Medioni G. Laser scan quality 3-D face modeling using a low-cost depth camera [C]// *20th European Signal Processing Conference*. Bucharest, Romania: IEEE, 2012: 1995-1999.

[16] Meyer G P, Do M N. Real-time 3D face modeling with a commodity depth camera [C]// *2013 IEEE International Conference on Multimedia and Expo Workshops*. San Jose, CA: IEEE, 2013: 1-4.