**ORIGINAL ARTICLE**

# A robust automated markerless registration framework for neurosurgery navigation

Long Jiang[1]
Shaoting Zhang[2]
Jie Yang[1]
Xiahai Zhuang[3]
Lixia Zhang[1]
Lixu Gu[1]*

[1]*School of Biomedical Engineering, Shanghai Jiao Tong University, People's Republic of China*

[2]*Department of Computer Science, University of North Carolina at Charlotte, NC, USA*

[3]*Shanghai Advanced Research Institute, Chinese Academy of Sciences, People's Republic of China*

*Correspondence to: L. Gu, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, People's Republic of China.
E-mail: gulixu@sjtu.edu.cn

## Abstract

**Background** The registration of a pre-operative image with the intra-operative patient is a crucial aspect for the success of navigation in neurosurgery.

**Methods** First, the intra-operative face is reconstructed, using a structured light technique, while the pre-operative face is segmented from head CT/MRI images. In order to perform neurosurgery navigation, a markerless surface registration method is designed by aligning the intra-operative face to the pre-operative face. We propose an efficient and robust registration approach based on the scale invariant feature transform (SIFT), and compare it with iterative closest point (ICP) and coherent point drift (CPD) through a new evaluation standard.

**Results** Our registration method was validated by studies of 10 volunteers and one synthetic model. The average symmetrical surface distances (ASDs) for ICP, CPD and our registration method were $2.24 \pm 0.53$, $2.18 \pm 0.41$ and $2.30 \pm 0.69$ mm, respectively. The average running times of ICP, CPD and our registration method were 343.46, 3847.56 and 0.58 s, respectively.

**Conclusion** Our system can quickly reconstruct the intra-operative face, and then efficiently and accurately align it to the pre-operative image, meeting the registration requirements in neurosurgery navigation. It avoids a tedious set-up process for surgeons. Copyright © 2014 John Wiley & Sons, Ltd.

**Keywords** neurosurgery navigation; 3D surface reconstruction; surface segmentation; SIFT; surface distance

## Introduction

Since surgical navigation enables minimally invasive procedures, it gradually becomes a normal procedure in the field of neurosurgery with the advance of volumetric CT and MRI. In this procedure, a prerequisite is the registration of the pre-operative planning and intra-operative data. It can be illustrated as an alignment of a pre-operatively obtained diagnostic image series to a co-ordinate system that is specific to the intra-operative patient's neuroanatomy (1). Effective registration is crucial for the quality of the medical procedure. There are three factors to evaluate the registration methods for navigation procedures: accuracy, efficiency and robustness (2).

A widely used method of registering the pre-operative image to intra-operative patient is to identify the common landmarks from the image and patient domains. Two types of landmark are available, artificial and anatomical. Artificial landmarks have been used as the gold standard for validation of other registration methods, but they require additional and non-trivial surgical procedures (3). Since the number of landmarks appearing in both modalities is usually small, and there are variations in the identified locations of the landmarks on the patient, as well as some difficulty in identifying exactly the same locations within the patient's three-dimensional (3D) image, the approach using anatomical landmarks, e.g. the nasal tip and the inner or outer corner of the eyes, may not be accurate (4). In addition, marker-based registration is time-consuming, with too much user interaction, which is a major defect discouraging the popularity of surgical navigation.

Therefore, markerless registration has become increasingly popular. Maurer *et al.* (5) introduced a markerless registration approach that uses a probe to sample points on the surface of the patient, and determines the best match of this point-cloud to the surface extracted from the patient's 3D image. It does not need to find the corresponding markers between the images and patient, and is adequate for many neurosurgical purposes. Unfortunately, such manual selection is time-consuming and is subjective to intra- and inter-observer variations. Recently, optical techniques have been investigated for intra-operative 3D reconstruction in the medical field. A large amount of literature has been published in the field of range imaging in medicine (6,7). In 2007, Konietschke *et al.* (8) used the DLR 3D-Modeller (3DMo) to perform markerless, contact-free registration. Nicolau *et al.* (9) proposed a system to reconstruct the surface of a patient's abdomen, and then perform the surface registration between reconstructed skin and the segmented skin from pre-operative images with ICP, to guide percutaneous punctures in interventional radiology. Reuben *et al.* (10) used a fast 3D surface scanner to acquire scans of patients' faces and aligned them with their pre-operative MR or CT head images. Schaller *et al.* (11) and Placht *et al.* (12) proposed time-of-flight (ToF)-based systems for automatic patient set-up in radiation therapy. Olesen *et al.* (13) presented a system for head motion tracking in 3D brain imaging. However, these systems require sophisticated operation, much set-up time or special equipment. In addition, there are some systems that have been used in the neurosurgical operating room (OR) (14–17) and have proved to be promising clinical applications. However, as they increase the burden on the surgeons and need much interaction, the adoption of image-guided surgery systems is adversely affected, especially in developing countries such as China. Bauer *et al.* (18) proposed a multi-modal surface registration scheme that enables direct alignment of the intra-fractional structured light data (Microsoft Kinect) to a reference shape extracted from the pre-fractional tomographic planning data. However, Kinect has limited accuracy. In our proposed framework, we chose structured light, since this technology has good precision and does not require any special equipment, i.e. only a camera and a projector are needed.

In this study, the pre-operative face image is segmented from the brain MRI/CT, while the intra-operative face image is reconstructed with structured light projection, using phase-shifting interferometry (19). Instead of using markers, we propose a new robust method for registration, and compare it with two different markerless registration methods. Based on the ASD, we also propose a new evaluation metric to assess the quality of the registration.

# Materials and methods

We propose a framework to map pre-operative images with the intra-operative patient position, where the optical navigator is used as a tracking device (Figure 1). It can track the position of a probe that is attached to a surgical device during surgery. Specifically, the histogram-based threshold method is employed for computing the threshold in each slice of the pre-operative images, and then the segmented facial surface is extracted. Structured light projection is used to calculate the reconstructed facial surface from the intra-operative patient. In order to align the segmented face and the reconstructed face, we propose an efficient and robust method, which aligns several corresponding SIFT points. In the following, we mainly introduce the 3D reconstruction and 3D–3D surface registration.

## 3D surface reconstruction with structured light

*Absolute phase map construction*
Phase measuring profilometry (PMP) is used in our projection calibration and surface reconstruction. First, three sinusoidal fringe patterns are generated and projected to the object sequentially with a projector. In addition, the experiments indicate that when the grey value is in the range 100–200, the grey-scale curve is approximately linear. Therefore, the vertical sinusoidal fringe patterns are defined as:

$$\mathrm{I}(x) = 50 \times \left(1 + \cos\left(\frac{2\pi x}{\rho} + \delta\right)\right) + 100 \qquad (1)$$

where $x$ is the horizontal coordinate of the pattern, $\mathrm{l}(x)$ is the intensity that only depends on the $x$, $\rho$ is the period of the fringe pattern in pixels, and $\delta$ is the phase shift, which is set to $-2\pi/3$, 0 and $2\pi/3$, respectively.
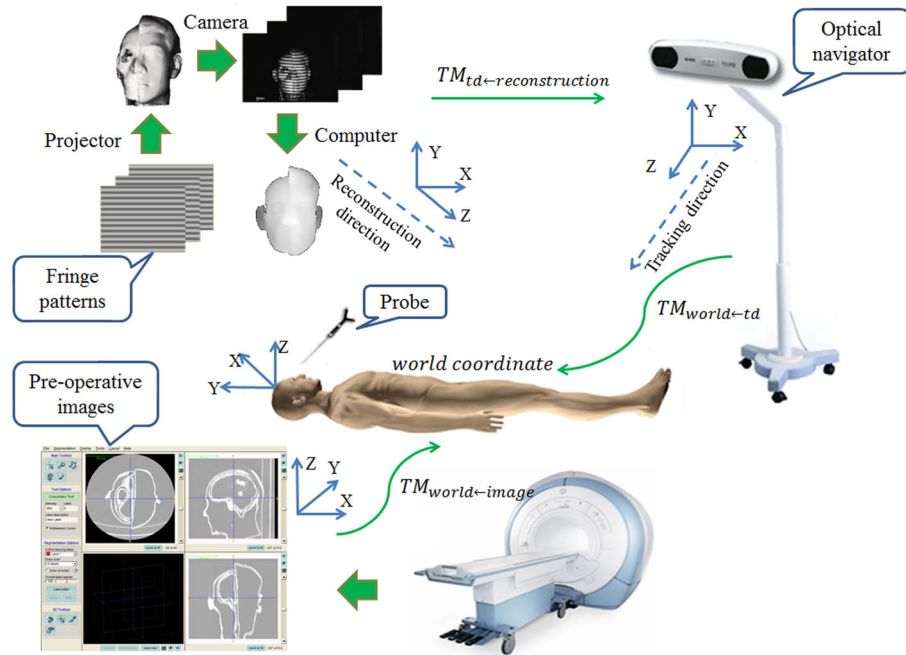
**Figure 1.** Neurosurgical navigation system and the coordinate transformations involved in spatial registration. $TM_{td \leftarrow reconstruction}$ is the transformation matrix from the reconstruction module coordinate to the tracking device coordinate. $TM_{world \leftarrow td}$ is the transformation matrix relating the tracking device coordinate to the world coordinate. $TM_{world \leftarrow image}$ is the transformation matrix aligning preoperative image coordinate with world coordinate

Then, these patterns are captured with a camera and denoised by Gaussian smoothing to compensate for the error caused by the camera. The intensity of the $i$th image is represented as:

$$I_i(x,y) = A(x,y) + B(x,y) \cos (\emptyset(x,y) + \delta_i) \qquad (2)$$

where $A(x, y)$ is the average intensity, $B(x, y)$ is the intensity modulation, ($i = 1,2,3$) is the phase shift and $\emptyset(x, y)$ is the phase to be determined.

Finally, by solving the above equation, the phase value $\emptyset(x, y)$ at each pixel is calculated as:

$$\emptyset(x,y) = \tan^{-1}\left( \sqrt{3}\left( \frac{I_1 - I_3}{2I_2 - I_{1-I_3}} \right) \right) \qquad (3)$$

The phase value $\emptyset(x,y)$ is modulo $2\pi$ at each pixel whose value is in the range 0 to $2\pi$. To increase phase measurement accuracy, fringe patterns with multiple fringes are usually used. In such case, the phase-shifting algorithm provides only a relative phase map, and the absolute phase map is obtained using a phase-unwrapping algorithm.

The above patterns are projected with a video projector and then captured by a camera. To obtain the absolute phase map, we use an additional corner pattern to convert the relative phase map to its corresponding absolute phase map. The corner's absolute phase value is assumed to be zero. We use (20) to detect the corner, as it is robust even in a curved surface. A multilevel quality-guided

phase unwrapping algorithm is used to unwrap the phase level by level (21). This algorithm not only improves the previous scan-line phase-unwrapping algorithm significantly, but also reduces the processing speed slightly.

*System parameters calibration*

The parameter calibration has two stages: camera calibration and estimation of the projector's projection centre position. The calibration can be carried out before surgery and be done only once. Our camera is calibrated using the MATLAB toolbox (22). The projector is regarded as the inverse of the camera, and it is represented accurately with a pinhole camera model. Early work in projector calibration includes the method proposed by Hu *et al.* (23) and Legarda-Saenz *et al.* (24). However, in our system, the position of the projection centre is the only projector parameter that is required by the surface reconstruction algorithm. The calibration method is to move a flat board under the structured light from a position far from the camera to a close position (25).

## Surface reconstruction

After calibration of the system and establishment of an absolute phase map, a dense 3D point cloud that represents the object's surface is achieved by the point cloud construction algorithm. Tao *et al.* provided a new method
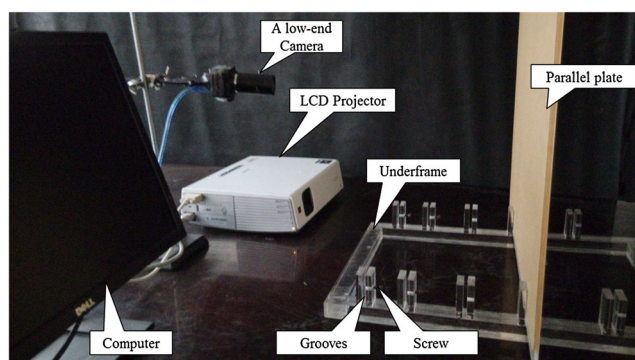
**Figure 2.** The hardware configuration of our 3D surface reconstruction module. It consists of a camera, a projector, a computer and a calibrated model. The calibrated model was used to evaluate the performance of surface reconstruction and calculate the projector centre

(25) to calculate 3D points. In their method, the object point is calculated by the intersection of two lines emanating from the camera centre and projector centre. Different from their method, we do not directly calculate the 3D point from the intersection of two lines, but from the midpoint of a line perpendicular to these two lines. Because these two lines are non-coplanar lines, they actually do not intersect. After the data acquisition, Gaussian smoothing is required to suppress random noise.

## 3D–3D surface registration

In neurosurgery, the brain's motion is constrained by the skull (2). In our case, we also need to ensure that the person's facial expression does not change significantly between the pre-operative and intra-operative stages. Thus, our proposed system is based on a rigid anatomy assumption. We perform a rigid registration using three methods: ICP, CPD and our proposed method based on SIFT.

### ICP and CPD
The ICP algorithm is a popular method for point sets registration (26), which is achieved by iteratively minimizing the energy with the rigid transformation (R, T) between two point sets. However, it is required that the initial positions of the two point sets are close enough by ICP, which is not always possible. Myronenko *et al.* (27,28) introduced a new probabilistic method for non-rigid and rigid registration of two point sets, i.e. CPD, which is a probabilistic approach, is shown to be accurate and robust in the presence of outliers and missing points, and is effective for the estimation of complex, non-linear, non-rigid transformations. In our work, these constructed faces usually lack some parts (e.g. ears); meanwhile, MRI/CT provide a complete face. CPD is able to handle registration of these two datasets without any modification.

### Proposed method
Since the reconstructed face and the segmented face contain tens of thousands of points, the speed is low to align these two point sets directly. In order to improve the efficient of the registration, we chose to use a two-dimensional (2D) depth map rather than a point set to represent the 3D face. The depth map stores 3D surface information in a 2D format because the value of each pixel stands for depth. The framework can be described as in Figure 2. SIFT has been widely used for extracting
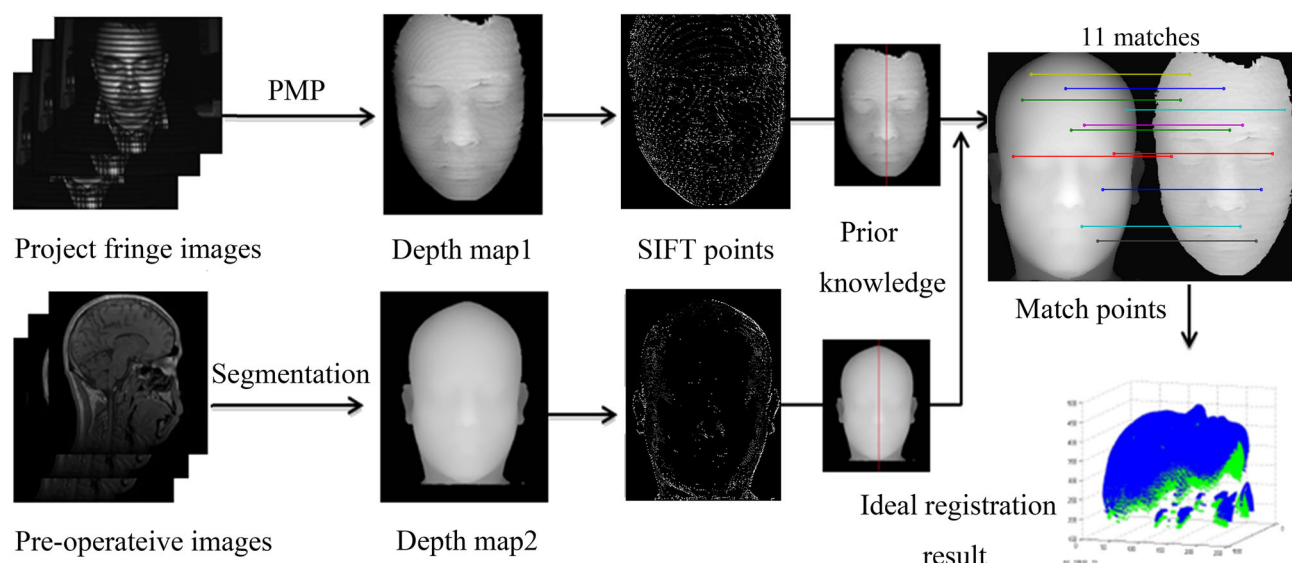


**Figure 3.** The framework of our registration method. Depth map 1 comes from the reconstructed face and depth map 2 is obtained from the segmented face. Curvature maps are generated from two depth maps, on which SIFT points are detected. The point matching is constrained by *a priori* knowledge. After matching these points, alignment is accomplished using Procrustes analysis

**Table 1. Performances obtained with 3D surface reconstruction**

| Ground truth distance (mm) | Estimated distance (mm) | RMS (mm) |
|---|---|---|
| 49.1 | 48.99 | 1.28 |
| 98.1 | 97.86 | 1.17 |
| 144.1 | 143.45 | 1.08 |
| 162.0 | 161.80 | 0.97 |
| 197.8 | 197.68 | 0.90 |
| 236.9 | 236.77 | 0.84 |

All calculations were repeated three times. Mean error was < 0.7 mm and the RMS was in the range 0.8–1.3 (mean 1.04 ± 0.17) mm.

features from images and has proved to be effective in medical image alignment (29). Toews and Wells (30) applied SIFT to 3D data, but it was time and space consuming. In our work, we apply SIFT to 2D images to obtain better computational efficiency. However, as the depth map of a face tends to be oversimplified, without enough features for precise location, simply applying SIFT to a depth map does not work well. Instead, we chose to calculate Gaussian curvature from the depth map, reshaping it into a curvature map, where the value of each pixel maps to its surface Gaussian curvature. Curvature is specific to anatomy, tolerant to a range of rotations and its features meet the requirement of SIFT. SIFT points in pre- and intramaps are matched according to their descriptor vectors. Furthermore, to improve robustness, we employ the random sample consensus (RANSAC) (31) to prune outliers during point matching. We also use prior knowledge (e.g. the facial medial axis can be located by detecting the highest vertical line on the depth map; points from different side of the line should not be paired) to constrain the pair matching, which comes from the depth map, and can further exclude outliers generated from symmetrical characteristics of faces. With the matching results, the global transform aligning two coordinate systems is computed using Procrustes analysis (32), which is described in more detail as follows.

The first step is point detection from a curvature map with SIFT. Using SIFT, potential interest points are identified by scanning the image over locations and scales. It is computed by constructing a Gaussian pyramid and searching local peaks (i.e. key-points) in a sequence of difference-of-Gaussian (DoG) images.

$$\max\nolimits_{x,y,\sigma}|f(x,y,k\sigma)-f(x,y,\sigma)| \qquad (4)$$

where $f(x, y, k\sigma)$ represents the convolution of the curvature map with a Gaussian kernel of variance $\sigma^2$, $k$ is a multiplicative scale-sampling rate, and 'max' denotes the local maximum. Besides, a descriptor vector for each key-point is built.

The second step involves points matching with extended constraints. For point $P_i \in S1$, its nearst neighbour $P_j$ in $S2$ can be defined as the key point with the minimum Euclidean distance for the invariant descriptor vector. To discard points that do not have any good match to the database, we should compare the distance of the closest neighbour to that of the second-closest neighbour. Empirically, 0.8 is the most appropriate value for threshold, which eliminates 90% of the false matches while discarding <5% of the correct matches (29). Therefore, in our case, we reject all matches in which the distance ratio is >0.8, i.e. when $P_j$ reliably matches $P_i$:

$$\frac{|D_i - D_j|}{|D_i - D_k|} \le 0.8 (P_k \in S2, k \ne j) \qquad (5)$$

where $D_i$ denotes the descriptor of point $P_i$.

Even though we employ RANSAC to improve the robustness, there are still some remaining outliers generated because of the facial symmetrical characteristics. Therefore, we employ *a priori* knowledge for correction (Figure 3). This can be achieved using integral projection (33).

Finally, once these points are matched, the global transform matrix between two point sets can be computed straightforwardly using Procrustes analysis (32). For two key-point set $S_1 = (x_1, ...., x_m)$ and $S_2 = (y_1, ...., y_m)$, we can calculate transformation (translation matrix T and rotation matrix R) between $S_1$ and $S_2$ and using Procrustes analysis:

$$\min\nolimits_{T,R} \sum_{i=1}^{n} \|(Rx_i + T) - y_i\| \qquad (6)$$

The pre-operative face can be segmented by a histogram-based threshold. Using the above approaches, we reconstruct the intra-operative face to achieve the requirements of registration.
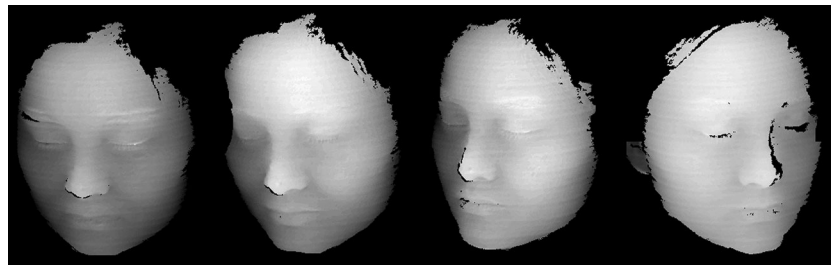


**Figure 4. The results of the reconstructed face in which the volunteer's face and the camera are in different positions**
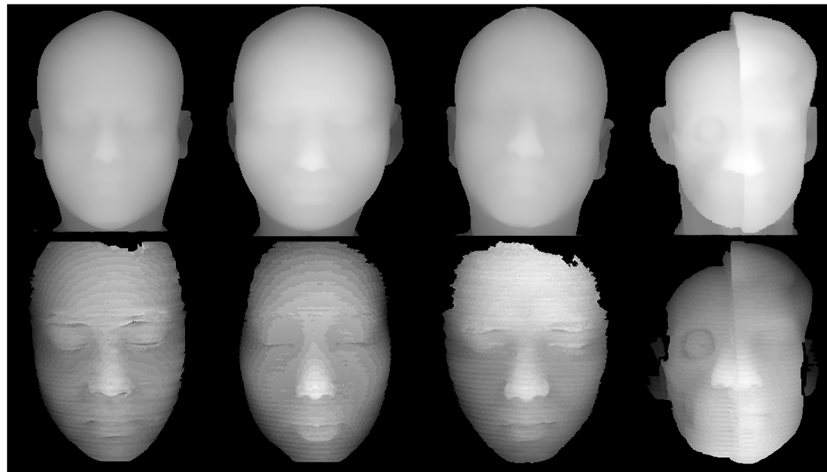
**Figure 5.** The first three images of the upper row are MRI-segmented faces of volunteers, and the fourth image of the upper row is the CT-segmented face of the model. The bottom images are the corresponding reconstructed faces

*A new evaluation metric for registration*

A new surface distance based method is proposed to evaluate the quality of registration. First, we employ the average symmetrical surface distance (ASD) (34):

$$ASD(A,B) = \frac{1}{\sum s_B + \sum s_A} \left( \sum_{s_A, \in s(A)} d\big((S_A, S(B))\big) \right.$$
$$\left. + \sum_{s_B, \in s(B)} d\big((S_B, S(A))\big) \right) \quad (7)$$

where $S(A)$ and $S(B)$ denote the set of surface voxels of the segmented face and the reconstructed face, respectively. $S_A$, $S_B$ is an arbitrary point of surface $S(A)$ and $S(B)$, respectively. The $d$ operator is used to calculate the shortest Euclidean distance between a point and the another surface. As the quality of the reconstruction surface is not ideal (it is not very smooth and lacks some marginal regions), the ASD of all points does not correctly reflect the quality of registration (the ASD of the missing part may be very large). Ultimately, we used points in regions of interest (nose, lips and eyes) to calculate ASD instead of all points, where these regions can represent the main features of the face.

Then, we propose to obtain the histogram distribution $H$ of the $d\{[S_B, S(A)]\}$:

$$H(d(i)) = \sum_{x_{i-1}}^{x_i} d((S_B, S(A))) | d((S_B, S(A))) \in (x_i, x_{i-1}) \quad (8)$$

where $H$ reflects the accuracy of registration. The greater proportion of low value in $H$ means the closer of two point sets, which indicates a better registration result. Therefore, the more points distribute within the low area means a better registration result. We evaluate the accuracy of our registration results with this method, and record registration time.

**Table 2. Registration performances with ICP, CPD and our method**

|  | ICP | CPD | Our method |
| --- | --- | --- | --- |
| Number of points in S1 | 32465.5 | 32465.5 | 32465.5 |
| Number of points in S2 | 5683.1 | 5683.1 | 56831.3 |
| Computation time (s) | 343.46 | 3847.56 | 0.58 |
| ASD (mm) | 2.24 ± 0.53 | 2.18 ± 0.41 | 2.30 ± 0.69 |
| Computation time ($p$) | 6.94e-12 | 8.30e-11 | |
| ASD ($p$) | 0.99 | 0.85 | |

We performed these three methods on 10 volunteers and a model, then demonstrated the mean results of these. S1 and S2 represent the segmented face and the reconstructed face, respectively. In S2, we uniformly down-sampled 10 times to increase the speed of ICP and CPD. However, the number of points had almost no influence on our method. We merely calculated the ASD of points in regions of interest. We performed *t*-tests comparing our method and ICP, and our method and CPD, and obtained the corresponding *p* values: for computation time, *p* < 0.01 for both; for ASD, *p* > 0.05 for both. ICP and CPD possessed slightly better accuracy than our proposed method, but our proposed method significantly improved the computational efficiency.

## Experimental results

Experimental results on 10 volunteers and a synthetic model are reported to evaluate the performance of our system. In our study, we acquired 10 MRI images from volunteers and one CT image from the synthetic model at Shanghai RENJI hospital. The MRI images were acquired using a GE Signa Hdxt3.0 T, using the Ax 3D-MPRAGE series and a voxel size ≤1 mm. The CT image was acquired with a GE LightSpeed VCT, using helical mode scan options and a slice thickness of 1.0 mm. Our 3D surface reconstruction module consists of a low-end Comos colour camera (the price is < $60) and a Sony XPL-CX80 LCD projector (Figure 2). All experiments were conducted on a core-I5 PC 2.67 GHz with 3 Gb of RAM. The reconstruction module is implemented in C++, and the segmentation and registration modules are achieved with Matlab.
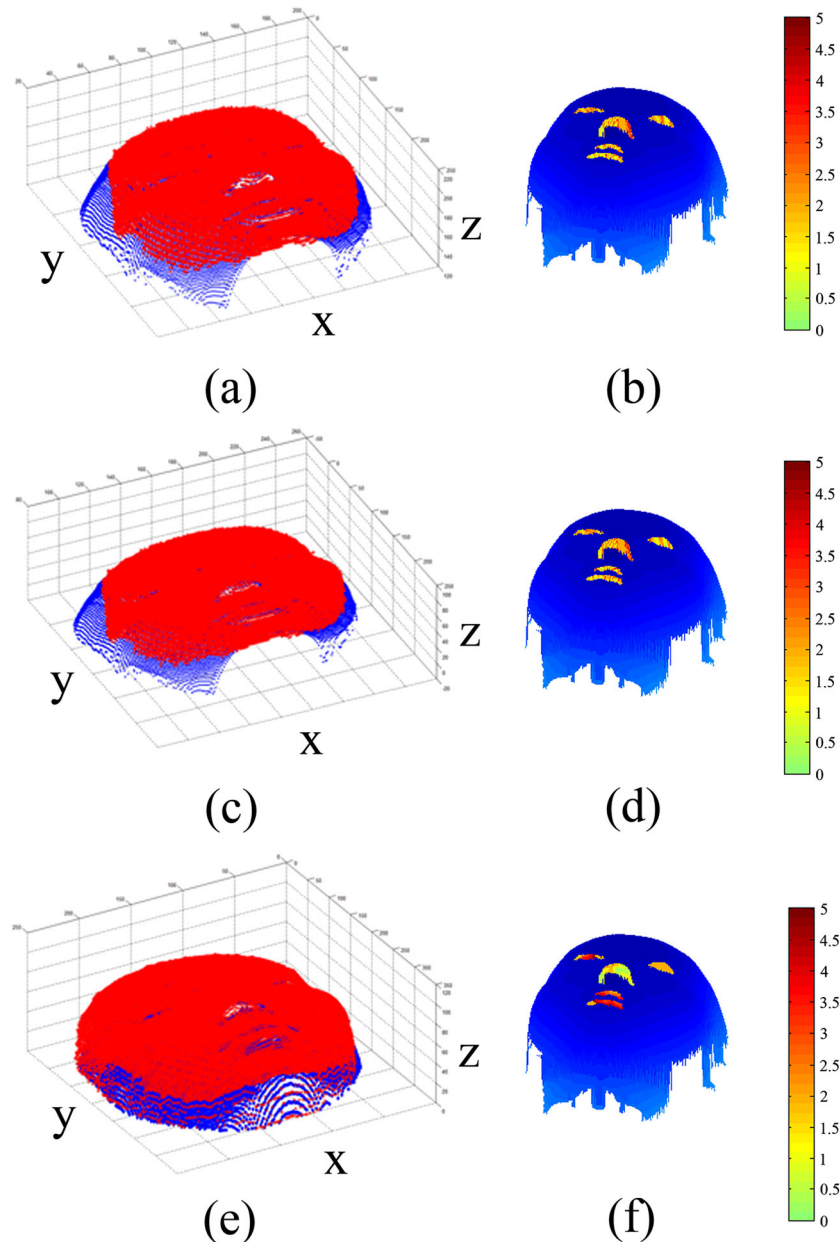


Figure 6. The registration results of one volunteer: (top to bottom) results of ICP (a, b), CPD (c, d) and our method (e, f). (Left column) Position of two point sets after registration: blue points, segmented face; red points, reconstructed face. Both the segmented face and the reconstructed face have been converted to the same coordinate space with NDI tracker before registration. (Right column) Errors in the regions of interest (nose, lips and eyes); errors of other regions are set to <0; errors are displayed on the segmented face

## Evaluation of surface reconstruction

Our 3D surface reconstruction was able to reconstruct objects with structured light. In the experiment, as the background of the object was a black curtain, only the object was illuminated by the light. To assess the accuracy of reconstruction, we used a flat board placed under the structured light and moved it in parallel from a known distance several times. For each position of the board, a surface reconstruction of the flat board was performed. Therefore, a cluster of points was obtained for each board position, and its location with respect to the camera frame could be estimated by averaging the position of each point. We designed a calibration model such that a flat board (opaque) could be inserted into a frame (Figure 2). In the frame, there were several parallel grooves where the flat board was inserted and fixed by screws. This calibration model was also used to calculate the projector centre. Since the translational distance between each board position was known, with accuracy <0.1 mm, it was possible to evaluate the reconstruction accuracy by computing the distance between each possible pair of reconstructed boards. In our case, we averaged the position of $200 \times 200$ pixels located in the centre of the image to estimate distance. We repeated this process three times, and the reconstruction results of the parallel board are shown in Table 1. The mean error was <0.7 mm and the RMS was in the range 0.8–1.3 (mean $1.04 \pm 0.17$) mm. In addition, the surface reconstruction took <5 s.

In addition, we added an experiment to evaluate different positions of the volunteer's face and camera and, as shown in Figure 4, the reconstruction module worked well in this case.

## Evaluation of registration

The intra-operative image was acquired with our 3D surface reconstruction module. We aimed to align the reconstructed face and the segmented face. The depth maps of our reconstructed and segmented faces are shown in Figure 5. The registration process was achieved via ICP, CPD and our method. We also used our proposed metric to evaluate the quality of these three registration approaches.

In our work, there were tens of thousands of points in the segmented and reconstructed faces. Therefore, it was too slow to use ICP and CPD on the complete data. Since the segmented face had smaller noise, which could represent the true face more accurately, we chose to divide the point set of the reconstructed face surface into 10 parts ($P_1$, $P_{11}$, $P_{21}$ … ; … ; $P_{10}$, $P_{20}$, $P_{30}$ … ). Then we performed the registration between these 10 parts and corresponding segmented

face. Note that our proposed method was very efficient, so we did not need to under-sample the points, as in ICP and CPD.

In order to evaluate the performances of registration with ICP, CPD and our method, we calculated the average ASD accuracy and running time of these three methods (Table 2) and demonstrated the results after registration (Figure 6); the average accuracies and computational times for each approach are listed in Table 2. Although ICP and CPD possessed slightly better accuracy than our proposed method (the difference was no more than 0.15 mm), our proposed method significantly improved the computational efficiency (<0.6 s). Meanwhile, the variances of accuracy of all three methods were <0.7 mm. The registration results of one volunteer are shown in Figure 6; we found that all of them could be accurately aligned. Note that we just focused on the regions of interest (nose, lips and eyes) and the errors of other regions were set to be <0. The error was denoted as the shortest Euclidean distance between points of the segmented face and the reconstructed face after alignment. Compared with ICP and CPD, our proposed method had more points with errors >2 mm (Figure 6f); this indicated a relatively poor registration result. In addition, with our evaluation method, we got an average histogram distribution to judge the quality of each registration approach (Figure 7), where we just focused on the regions of interest (a total of 961 points). Note that >86% points distributed within the range 0–2.5 mm using ICP and CPD; meanwhile, ~83% points distributed within the same range using our
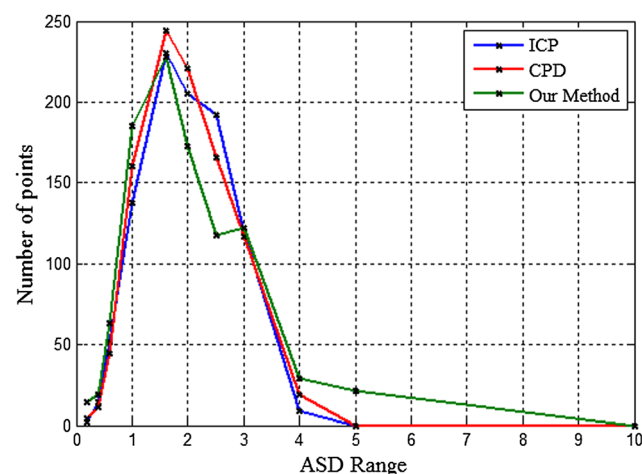


**Figure 7.** Average histogram distribution, which represents the registration result in the regions of interest, in which the horizontal axis represents the value of ASD and the vertical axis represents the number of points located within the range of the horizontal axis. There are more points distributed within the low area using ICP and CPD, which means a better registration result

proposed method. Therefore, our proposed method could achieve comparable accuracy.

## Discussion

The present study demonstrated the feasibility of markerless registration in neurosurgery navigation surgery with our framework. Specifically, we accurately registered a CT/MRI model of the patient in the camera frame with structured light. However, several limitations have to be considered and analysed in our framework, due to the characteristics of our system.

### Data

In our system, there are two types of data: the reconstructed face and the segmented face. The reconstruction module consists of a low-end camera and a LCD video-projector. Therefore, our system is easy to build with some commonly obtainable appliances. Unfortunately, these hardware settings might bring problems on the reconstructed face. In the reconstruction module, the quality of the captured images influences the accuracy of reconstruction to a certain degree. Moreover, our camera and projector are susceptible to environmental factors, such as light and illumination.
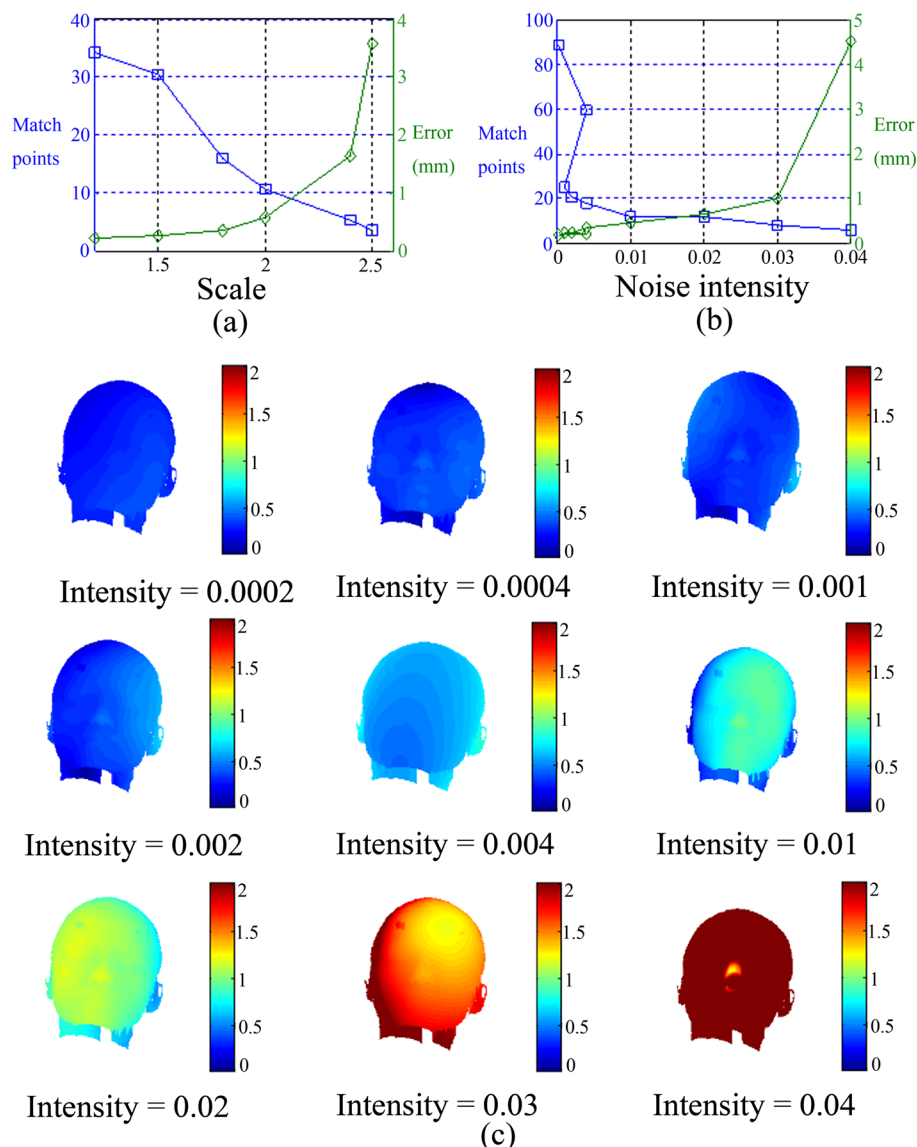


**Figure 8.** (a) Average number of matched SIFT points and registration error (average point-to-point distance among the whole surface compared with manually selected landmarks) when the image scale changes in the intra-operative map. (b) Average number of matched SIFT points and registration error when stripe noise is added. The value of noise intensity is defined as: maximal (*noise*)/maximal(*depth*). (c) Visual comparisons between manually selected points and using our auto algorithm when noise is added. The colour maps to registration errors

Therefore, the quality of our captured images may not always be ideal for reconstruction. Our experiments showed that the RMS error of the reconstructed plane is about ~1 mm, which is certainly not perfect. In addition, on the reconstructed face, we can also observe some stripe noise because of the ambient light. On the other hand, we need to segment the surfaces of pre-operative images (CT/MRI) and we just extract the face for consistency. Although it is not hard to segment the surface in CT, the segmented surface in MRI may be slightly beneath the skin, because of the noise. To alleviate this problem, points in the MRI are clustered using Quickshift (35) for better edge detection, and the histogram-based threshold method is employed for computing the threshold in each slice. Despite these methods, accurate segmentation of the MRI surface remains a challenge.

## Methods

In our framework, several processes introduce errors. First, in order to obtain the reconstructed face, we have to perform coordinate transformations several times, e.g. camera to world coordinate, projector to world coordinate, and image to world coordinate. Each transformation may influence the accuracy as the truncation of data. Second, we calibrated the camera using the Matlab toolbox provided by Bouguet (22), with a printed checkerboard; errors in this process reached around 0.3 pixels. In addition, we noticed that ICP and CPD possessed slightly better accuracy than our proposed method in the experiments. However, ICP and CPD are not as efficient as our method. Even when we have down-sampled the data, these two methods are still too slow for clinical purposes (>5 min). Our method significantly improves the computational efficiency (i.e. nearly real-time),

which makes it practical for clinical problems. On the other hand, its accuracy (the mean accuracy is ~2 mm) is still comparable to state-of-the-art methods, such as those of Konietschke *et al.* (8) and Nicolau *et al.* (9), using ICP-based registration (the mean accuracy is ~2 mm). The registration task of our method takes <1 s. Although the systems that have been used in neurosurgical operating rooms (the mean accuracy ~2 mm) have comparable accuracy with ours (14–17), our goal is different. We proposed a novel touchless system with less interaction, where the system could automatically detect the patient's facial surface and perform registration independent of users. At present, our results may not yet be applicable for clinical practice. Combination with other registration methods, such as a marker-based method, might be one solution. For example, it is possible to consider the deformation constraint (36,37) and the sparsity constraint (38) to improve accuracy and robustness and the ability to handle small movements. Nonetheless, our surgeons still agree that it has good potential. The reason is that the final product could accurately and physically integrate our projector inside the tracker system, and use a tracker camera instead of our additional one. Compared with handing a laser-scanner system, it can avoid unnecessary coordinate conversions to reduce accuracy loss. In addition, we will apply infrared light instead of visible light. Furthermore, we performed tests on the synthetic data to evaluate the robustness of our method (Figure 8). When the scale change is no larger than 2, on average >10 point matches are found, ensuring high registration accuracy (error no more than 0.6 mm). When the noise intensity is <0.03, the registration error is slight (no more than 1.0 mm) and grows slowly, whereas the error suddenly climbs to 4.0 mm when noise intensity reaches 0.04. Our registration method proved to maintain reasonable robustness,
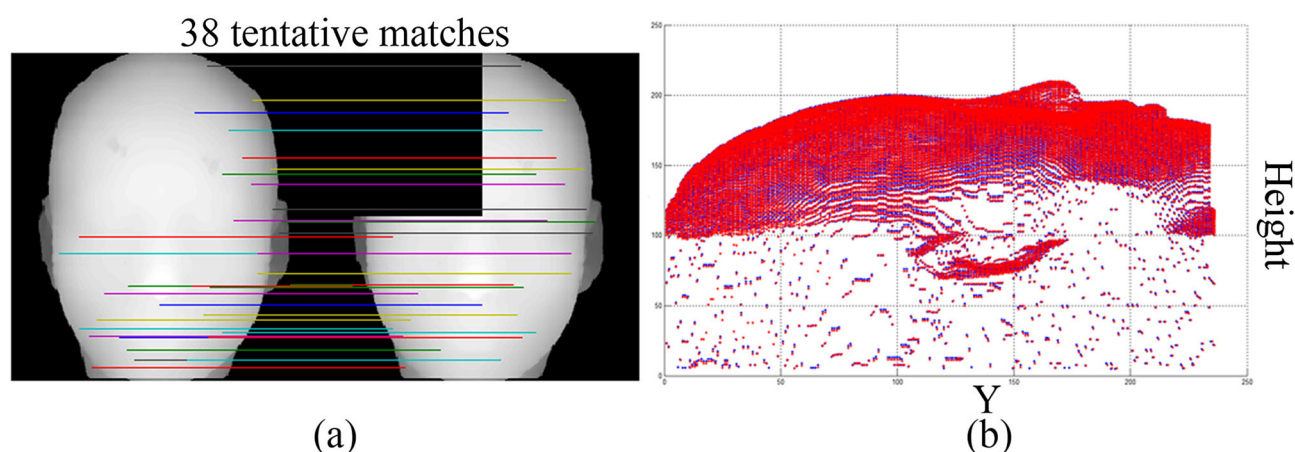


**Figure 9.** The registration result of one MRI dataset where one of the dataset lacks the top left face: (left) match points extracted from Gaussian curvature map by SIFT, which are shown in the depth map; (right) registration result of these two dataset, where the blue points represent the full face, while the red points represent the incomplete face

yet the stripe noise and scale change in reconstructed intra-operative depth map notably reduced the registration accuracy. In addition, it is robust even when partial data is missing, since key points on missing part have no corresponding points, and registration will not utilize these points (Figure 9). The main errors in our current reconstruction system result from the camera calibration and coordinate transformation. If these system errors are fixed, the registration accuracy would be further improved, enabling us to achieve a markerless framework with greater accuracy yet far lower cost than in previous work.

Compared with the marker-based registration, our system does not require complicated navigation preparations and hence significantly saves the effort of physicians. Therefore, our system is portable and will able to be used *in vivo* in the future.

## Conclusions

This paper proposes a robust and automated markerless registration framework using SIFT-based method. In neurosurgery navigation, it can complete registration automatically and free surgeons from a tedious pre-operative manual operation. We apply our method to register the CT/MRI model of the patient in the camera frame, and compare it with several other methods. The experimental results demonstrate that our method is the most appropriate for this application. In the future, we plan to accurately and physically integrate our projector inside a tracker system and directly use a tracker camera.

## Acknowledgements

## Conflict of interest

The authors have stated explicitly that there are no conflict of interest in connection with this article.

## Funding

## References

1. Miga MI, Sinha TK, Cash DM, *et al*. Cortical surface registration for image-guided neurosurgery using laser-range scanning. *IEEE Trans Med Imaging* 2003; **22**(8): 973–985.
2. Yaniv Z, Cleary K. Image-guided procedures: a review. *Comput Aid Intervent Med Robot* 2006; **3**: 1–64.
3. West J, Fitzpatrick JM, Wang MY, *et al*. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J Comput Assist Tomogr* 1997; **21**(4): 554–568.
4. Peters TM. Image-guided surgery and therapy: current status and future directions. *Int Soc Optics Photon* 2001; **4319**: 1–12.
5. Maurer CR, Jr, Maciunas RJ, Fitzpatrick JM. Registration of head CT images to physical space using a weighted combination of points and surfaces. *IEEE Trans Med Imaging* 1998; **17**(5): 753–761.
6. Maier-Hein L, Mountney P, Bartoli A, *et al*. Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery. *Med Image Anal* 2013; **17**(8): 974–996.
7. Bauer S, Seitel A, Hofmann H, *et al*. Real-time range imaging in health care: a survey. In Time-of-Flight and Depth Imaging: Sensors, Algorithms, and Applications. Springer: Berlin, Heidelberg, 2013; 228–254.
8. Konietschke R, Busam A, Bodenmüller T, *et al*. Accuracy identification of markerless registration with the DLR handheld 3D-modeller in medical applications. In Proceedings of College and University Retiree Associations of Canada (CURAC), 2007; 6.
9. Nicolau SA, Brenot J, Goffin L, *et al*. A structured light system to guide percutaneous punctures in interventional radiology. *Int Soc Optics Photon*, 2008; 700016–700016–11.
10. Shamir RR, Freiman M, Joskowicz L, *et al*. Surface-based facial scan registration in neuronavigation procedures: a clinical study. *J Neurosurg* 2009; **111**(6): 1201–1206.
11. Schaller C, Adelt A, Penne J, *et al*. Time-of-flight sensor for patient positioning. *Int Soc Optics Photon* 2009; 726110–726110–8.
12. Placht S, Stancanello J, Schaller C, *et al*. Fast time-of-flight camera-based surface registration for radiotherapy patient positioning. *Med Phys* 2012; **39**(1): 4–17.
13. Olesen OV, Paulsen RR, Hojgaard L, *et al*. Motion tracking for medical imaging: a nonvisible structured light tracking approach. *IEEE Trans Med Imaging* 2012; **31**(1): 79–87.
14. Raabe A, Krishnan R, Wolff R, *et al*. Laser surface scanning for patient registration in intracranial image-guided surgery. *Neurosurgery* 2002; **50**(4): 797–803.
15. Schlaier J, Warnat J, Brawanski A. Registration accuracy and practicability of laser-directed surface matching. *Comput Aid Surg* 2002; **7**(5): 284–290.
16. Eljamel MS, Hofer M. From letterbox to keyhole approach for resecting intracranial lesions. *Stereotact Funct Neurosurg* 2004; **81**(1–4): 30–36.
17. Brodie J, Eljamel S. Evaluation of a neurosurgical robotic system to make accurate burr holes. *Int J Med Robotics Comput Assist Surg* 2011; **7**(1): 101–106.
18. Bauer S, Wasza J, Haase S, *et al*. Multi-modal surface registration for markerless initial patient set-up in radiation therapy using Microsoft's Kinect sensor. In IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011; 1175–1181.
19. Schreiber H, Bruning JH. Phase shifting interferometry. In Optical Shop Testing, 3rd edn. Wiley Online Library: Hoboken, 2006; 547–666.
20. Sun W, Yang X, Xiao S, *et al*. Robust checkerboard recognition for efficient nonplanar geometry registration in projector-camera systems. Proceedings of the 5th ACM/IEEE International Workshop on Projector Camera Systems, 2008; 2.
21. Zhang S, Li X, Yau ST. Multilevel quality-guided phase unwrapping algorithm for real-time three-dimensional shape reconstruction. *Appl Optics* 2007; **46**(1): 50–57.
22. Bouguet JY. Camera calibration toolbox for Matlab, 2004; http://www.vision.caltech.edu/bouguetj/calib_doc/

23. Hu Q, Huang PS, Fu Q, *et al.* Calibration of a three-dimensional shape measurement system. *Optic Eng* 2003; **42**(2): 487–493.

24. Legarda-Sa R, Bothe T, Ju WP. Accurate procedure for the calibration of a structured light system. *Optic Eng* 2004; **43**(2): 464–471.

25. Peng T, Gupta SK. Model and algorithms for point cloud construction using digital projection patterns. *J Comput Inf Sci Eng* 2007; **7**(4): 372–381.

26. Besl PJ, McKay ND. A Method for registration of 3D shapes. *Int Soc Optics Photon* 1992; **14**(2): 586–606.

27. Myronenko A, Song X, Carreira-Perpinán MA. Non-rigid point set registration: coherent point drift. *Adv Neur Inf Process Syst* 2006; **19**: 1009–1016.

28. Myronenko A, Song X. Point set registration: coherent point drift. *IEEE Trans Pattern Anal Mach Intell* 2010; **32**(12): 2262–2275.

29. Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vision* 2004; **60**(2): 91–110.

30. Toews M, Wells WM, III. Efficient and robust model-to-image alignment using 3D scale-invariant features. *Med Image Anal* 2013; **17**(3): 271–282.

31. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 1981; **24**(6): 381–395.

32. Goodall C. Procrustes methods in the statistical analysis of shape. *J R Statist Soc B Method* 1991; **53**(2): 285–339.

33. Zhou ZH, Geng X. Projection functions for eye detection. *Pattern Recog* 2004; **37**(5): 1049–1056.

34. Heimann T, Van Ginneken B, Styner MA, *et al.* Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans Med Imaging* 2009; **28**(8): 1251–1265.

35. Vedaldi A, Soatto S. Quick shift and kernel methods for mode seeking. In Computer Vision – ECCV 2008. Springer: Berlin, Heidelberg, **2008**; 705–718.

36. Li H, Huang X, He L. Object matching using a locally affine invariant and linear programming techniques. *IEEE Trans Pattern Anal Mach Intell* 2013; **35**(2): 411–424.

37. Li H, Huang X, Huang J, *et al.* Feature matching with affine-function transformation models. *IEEE Trans Pattern Anal Mach Intell* 2014; PP(99): 1–1.

38. Zhang S, Zhan Y, Dewan M, *et al.* Towards robust and effective shape modeling: sparse shape composition. *Med Image Anal* 2012; **16**(1): 265–277.