IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 26, 2018, accepted August 28, 2018, date of publication September 10, 2018, date of current version October 12, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2869433

Robust Stereo-Match Algorithm for Infrared Markers in Image-Guided Optical Tracking System

QINYONG LIN[1], RONGQIAN YANG[2,3], ZHESI ZHANG[2], KEN CAI[4,5], ZHIGANG WANG[6], MEIPING HUANG[7], JINHUA HUANG[8], YINWEI ZHAN[9], AND JIAN ZHUANG[10]

[1]School of Medicine, South China University of Technology, Guangzhou 510006, China
[2]Department of Biomedical Engineering, South China University of Technology, Guangzhou 510006, China
[3]School of Medicine, Yale University, New Haven, CT 06520, USA
[4]School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China
[5]College of Automation, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China
[6]Guangzhou Aimooe Technology Co., Ltd., Guangzhou 510006, China
[7]Guangdong Provincial Key Laboratory of Coronary Heart Disease Prevention, Department of Catheterization Lab, Guangdong Cardiovascular Institute, Guangdong General Hospital, Guangdong Academy of Medical Science, Guangzhou 510080, China
[8]Department of Minimally Invasive Interventional Radiology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, China
[9]School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China
[10]Department of Cardiac Surgery, Guangdong Cardiovascular Institute, Guangdong General Hospital, Guangdong Academy of Medical Science, Guangzhou 510080, China

Corresponding author: Rongqian Yang (bmeyrq@foxmail.com)

This work was supported in part by the China Postdoctoral Science Foundation under Grant 2017M612671, Grant 2018T110880, and Grant 2017M620375, in part by the National Natural Scientific Foundation of China under Grant 81671788 and Grant U1401255, in part by the Guangdong Provincial Science and Technology Program under Grant 2016A020220006, Grant 2017B020210008, Grant 2017B010110015, and Grant 2017A040405054, in part by the Fundamental Research Funds for the Central Universities under Grant 2017ZD082, in part by the Guangzhou Science and Technology Program under Grant 201704020228, and in part by the Chinese Scholarship Fund under Grant 201806155010.

ABSTRACT The stereo-match to find the corresponding points in the left and right images from an optical tracking system is a crucial step in 3-D point reconstruction. Unfortunately, the presence of ghost markers remains an issue which is caused by indistinct correspondence when several markers are coplanar with two optical centers. Such ghost markers can lead to erroneous tracking of medical instruments and patients. Some methods have been proposed for this issue, but they are rarely applied in tracking systems. To address the issue, we propose a robust stereo-match method for combining the maximum silhouette reprojection consistency with the epipolar constraint into an objective function. For minimization of the objective function, we reconstruct a virtual sphere through optimal triangulation and construct the 3-D distance errors between the inverse projective rays of the marker contour and the reprojection contour on the virtual sphere surface. A contour-detecting method is proposed to obtain the accurate contours of marker blobs in images. Simulative and real experiments are performed to evaluate the performance of the proposed stereo-match method. We compare the proposed method with other methods performed on several real data sets which are acquired from a human phantom. The experimental results demonstrate that the proposed method is more robust and efficient in searching the correct correspondences of the markers.

INDEX TERMS Ghost markers, image-guided therapy, optical tracking system, silhouette reprojection error, stereo match.

I. INTRODUCTION

Optical tracking is a remote sensing technology commonly used for the conversion of the image observations of infrared (IR) markers captured by multiple cameras into three-dimensional (3D) positions. The position information of IR markers is usually used for tracking the position and orientation of medical instruments and patients in clinical image-guided therapy [1]–[7]. Tracking medical instruments

and patients in positioning lesions with high precision is a common issue in a range of image-guided therapies including orthopedic surgery [8], neurosurgery [9], [10], radiotherapy [11], and even soft tissue surgery [12], [13]. One important purpose of tracking medical instruments and patients is to register the intraoperative physical anatomy to preoperative images for the visualization of the movement of a medical instrument on the preoperative images. The optical tracking technology assists physicists or surgeons improving therapy, thereby reducing invasiveness, treatment time, and complications [11], [14].

Optical tracking systems (OTSs) have been used in clinical image-guided therapies for more than two decades. An OTS typically comprises two cameras and two rings of infrared light sources mounted closely around the camera lenses [15]. Spherical markers covered with reflective material are usually affixed on medical instruments and skins of patients. Each spherical marker should be placed in the field of view of OTS, and the correspondences of markers in the two images should be sought correctly for the unambiguous establishment of the position of each marker in real time. This approach ensures that the movement of the instruments and the change of patient position are tracked in real-time.

However, even with the use of costly professional IR marker-based tracking systems, such as Polaris (Northern Digital, Inc., ON, Canada), which is widely used in clinical image-guided therapy, ghost markers appear in some instances when some markers become coplanar with two optical centers (the Passive Polaris Spectra User Guide in Section 4.12 provides statements of phantom markers). Ghost markers are attributed to the false correspondences of marker sets in two images, which is called false stereo-match. In this situation, the presence of ghost markers may cause erroneous tracking of medical instruments and patients during treatment. Studies have reported the influence of ghost markers on therapy [16], [18].

Several researchers have developed many methods to address the stereo-match problem [19]–[28]. The detection of markers in two-dimensional (2D) images is not complicated because the projection images of markers appear as bright white spots with dark backgrounds. As these bright white spots are homogeneous and smooth, determining a spot that belongs to a particular marker from an image is impossible through the pixels of the spot. Thus, stereo-match methods of adaptive support-weight [20], adaptive normalized cross-correlation method [21], SIFT [23], and SURF [22] are unsuitable for IR maker-based OTSs. An epipolar constraint is generally used for searching stereo correspondence [24]–[26], but it is incapable of determining correct correspondences for the markers coplanar with two optical centers, and subsequent ghost markers are generated after the reconstruction of 3D points. Some researchers use 3D metrics to discard ghost markers by determining the marker combination that matches a predefined pattern [19], [24], [27], [28]. Robert and Manuel Loaiza used a projective invariant to distinguish a pattern formed by four colinear markers and another pattern formed by five coplanar markers [24], [28]. Stroian *et al*. [27] proposed a strategy in which a 3D geometric volume is generated for each marker defined by the two optical centers and a marker. Their study provided a guideline for the placement of markers outside the geometric volume of each other. Furthermore, using a CT scan image as a pattern, Yan *et al*. [19] were able to detect reference markers through their proposed method based on pattern matching. They were able to eliminate ghost markers by matching the CT scan pattern with the 3D positions of markers reported by the Polaris tracking system. Although these methods eliminate ghost markers by certain geometric constraints, most of them are incompatible with many tracking systems, and they often require manual intervention.

Between two marker sets in different images, if the correspondence can be precisely searched, it could reduce the number of false corresponding point pairs which appeared in the stereo-match step and subsequently reduce the number of ghost markers in the 3D reconstruction step. Thus, we propose a robust stereo-match algorithm for searching the correct correspondences between two marker sets. This algorithm combines the geometric information of marker in a 2D image with that in a 3D space. In the aforementioned methods, each marker is represented as the centroid of pixels that belong to a certain marker, whereas the other pieces of information included in the pixels of each marker and the volume size of the marker in 3D space are discarded. In this study, these discarded pieces of information are used for searching correct stereo-match correspondences. An objective function is then established by minimizing the silhouette reprojection error between the reprojection image and observed image for each instance of combination correspondence between markers, which we call the Maximum Silhouette Reprojection Consistency (MSRC) method. This method enables to search the correct combination correspondences with minimum error via a nonlinear optimization method.

Some preliminaries on OTSs are provided in Section II. The details of the MSRC method using the geometric information of a marker in 2D image and 3D space are presented in Section III. In Section IV, the proposed method is validated with simulative and real data. Finally, conclusions are provided in Section V.

II. PRELIMINARIES

A. PINHOLE CAMERA MODEL

A pinhole camera model is introduced to describe the imaging process of the OTS. $\boldsymbol{P} = [x, y, z]^{\mathrm{T}}$ is the coordinate of a 3D point $\boldsymbol{P}$ in the world coordinate frame, and $\boldsymbol{p} = [u, v]^{\mathrm{T}}$ is the image projection of point $\boldsymbol{P}$. The relationship between the 3D point $\boldsymbol{P}$ and its image projection $\boldsymbol{p}$ is expressed as

$$\beta\widetilde{\boldsymbol{p}} = \boldsymbol{A}\left[\boldsymbol{R}\middle|\boldsymbol{t}\right]\widetilde{\boldsymbol{P}}, \tag{1}$$

where β is a nonzero scale factor, and $\boldsymbol{R}$ and $\boldsymbol{t}$ are the rotation matrix and translation vector related world coordinate frame to camera coordinate frame which are called the extrinsic parameters, $\widetilde{\boldsymbol{p}}$ and $\widetilde{\boldsymbol{P}}$ are the homogeneous coordinates of

image point $\boldsymbol{p}$ and 3D point $\boldsymbol{P}$, and $\boldsymbol{A}$ is the intrinsic parameter matrix; its form is given by

$$\boldsymbol{A} = \begin{bmatrix} f_u & 0 & u_o \\ 0 & f_v & v_o \\ 0 & 0 & 1 \end{bmatrix}, \tag{2}$$

where f_u and f_v are the effective focal length in the directions of the image axes u and v, and (u_o, v_o) are the coordinates of the principal point.

The geometric deformation of an object image in the camera image plane is called distortion, which is introduced by the camera lens. The distortion of the camera lens is caused by several physical and technological reasons. The geometirc distortion of the camera lens is primarily affected by the radial distortion. The distortion effect can be described as

$$\begin{cases} \delta_u = (u^d - u_o)(k_1 r_c^2 + k_2 r_c^4) \\ \delta_v = (v^d - v_o)(k_1 r_c^2 + k_2 r_c^4) \\ v = u^d + \delta_u \\ u = v^d + \delta_v, \end{cases} \tag{3}$$

where r_c is the distance from the principal point to undistorted image point $\boldsymbol{p}$. The parameters k_1 and k_2 are the distortion coefficients of the camera. These unknown parameters involved in the pin-hole camera model can be acquired from the results of the camera calibration.

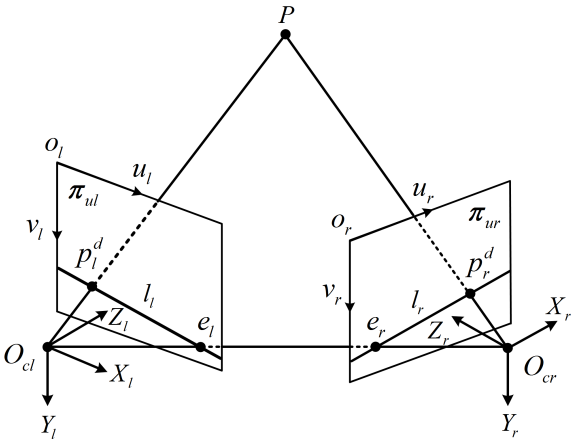

FIGURE 1. Depict of the stereo vision model and the epipolar geometry.

B. MODEL OF STEREO VISION AND EPIPOLAR GEOMETRY

OTS is designed on the basis of the principle of stereo vision (binocular vision), the model of which is shown in Fig. 1. The coordinate frames of left and right cameras are denoted as $O_{cl} - X_l Y_l Z_l$ and $O_{cr} - X_r Y_r Z_r$, respectively. The corresponding image coordinate frames of left and right cameras are denoted as $o_l - u_l v_l$ and $o_r - u_r v_r$, respectively. In this paper, the coordinate frame of left camera is considered the world coordinate frame. Assume that $\boldsymbol{p}_l = [u_l, v_l]^{\mathrm{T}}$ and $\boldsymbol{p}_r = [u_r, v_r]^{\mathrm{T}}$ are the projection coordinates of 3D point $\boldsymbol{P} = [x, y, z]^{\mathrm{T}}$ in the image planes of left and right cameras, respectively. Let $\widetilde{\boldsymbol{P}} = [x, y, z, 1]^{\mathrm{T}}$, $\widetilde{\boldsymbol{p}}_l = [u_l, v_l, 1]^{\mathrm{T}}$, and $\widetilde{\boldsymbol{p}}_r = [u_r, v_r, 1]^{\mathrm{T}}$ be the homogeneous coordinates of $\boldsymbol{P}$, $\boldsymbol{p}_l$, and $\boldsymbol{p}_r$, respectively. Based on the camera model, the mapping relationship between 3D point $\boldsymbol{P}$ and 2D image points $\boldsymbol{p}_l, \boldsymbol{p}_r$ can be expressed as

$$\begin{cases} \beta_l \widetilde{\boldsymbol{p}}_l = \boldsymbol{A}_l \left[\boldsymbol{I} \middle| 0 \right] \widetilde{\boldsymbol{P}} \\ \beta_r \widetilde{\boldsymbol{p}}_r = \boldsymbol{A}_r \left[\boldsymbol{R}' \middle| \boldsymbol{t}' \right] \widetilde{\boldsymbol{P}}, \end{cases} \tag{4}$$

where $\boldsymbol{I}$ is the 3×3 identity matrix, $\boldsymbol{R}'$ and $\boldsymbol{t}'$ are the rotation matrix and translation vector that transform the point coordinates in the left camera frame to the right camera frame.

Points $\boldsymbol{p}_l$ and $\boldsymbol{p}_r$ in the left and right image are related by the so-called epipolar geometry such that they should lie on a pair of corresponding epipolar lines $\boldsymbol{l}_l$ and $\boldsymbol{l}_r$ in the two images. The corresponding pair of points $\boldsymbol{p}_l \leftrightarrow \boldsymbol{p}_r$ will satisfy the following relationship

$$\widetilde{\boldsymbol{p}}_r^{\mathrm{T}} \boldsymbol{F} \widetilde{\boldsymbol{p}}_l = 0, \tag{5}$$

where $\boldsymbol{F}$ is the fundamental matrix that it can be expressed as a 3× 3 matrix

$$\boldsymbol{F} = \boldsymbol{A}_r^{-\mathrm{T}} [\boldsymbol{t}]_\times \boldsymbol{R}' \boldsymbol{A}_l^{-1}, \tag{6}$$

where $[*]_\times$ is defined as a mapping from a 3D vector to a 3×3 antisymmetric matrix

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_\times = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}. \tag{7}$$

The fundamental matrix $\boldsymbol{F}$ can be calculated by the intrinsic and extrinsic parameters of the OTS [29].

When the calibration result of the stereo vision system is obtained and the correspondence between the left and right camera images are established, the coordinate of 3D space point $\boldsymbol{P}$ can be solved by the triangulation based on Eqs. (1)–(4). The objective of the triangulation method is to solve the intersection of the two lines that connect two optical centers to two corresponding image points in a 3D space. Generally, the measured coordinates in the two camera image planes are inaccurate because of various types of noise, such as model error, image noise, and lens distortion [29]. Given two image coordinates with measurement error, the rays emitted from the two optical centers to the corresponding image points cannot be guaranteed to intersect precisely in the 3D space. Thus, the coordinate of point $\boldsymbol{P}$ in 3D space is needed for estimation via certain methods.

III. STEREO-MATCH METHOD

In this section, we describe the MSRC method for OTS. This method constructs silhouette reprojection error and epipolar constraint into an objective function to determine the correct correspondences among combination markers via an optimal method. The implementation of the method mainly comprises three steps. The centroids of the reflective markers are detected in the left and right images in the first step. The second step is the extraction of cases wherein more than

two markers are present in the same epipolar line by applying the epipolar constraint. The third step is the implementation of the stereo-match algorithm using the silhouette reprojection error combined with the epipolar constraint. The stereo-match algorithm in this step is, to the best of our knowledge, the innovative part of this paper.

A. MARKER DETECTION METHOD

Marker detection, which involves the identification, segmentation, and calculation of the centroids of markers in images, is a step that follows image acquisition from two cameras. This step is crucial in acquiring an accurate measurement of the markers in images. Owing to the reflective coating in the surfaces of the markers, the markers reflect near infrared light emitted by the illuminators of the OTS back to the cameras. Thus, the projection images of the markers are presented as nearly round blobs with high intensity. However, most of the other image area is almost black, indicating a relatively high contrast to the markers. The commonly used method to segment the bright blobs from the background is the implementation of a binary thresholding. However, the contour of the marker is not well defined by a threshold. That is, a blurry region appears between the background and the center region of the marker. When the distance between the marker and camera changes, the brightness of the marker varies at different distances between the camera and marker, and the blurry region of the marker's contour changes accordingly. Furthermore, the contour suffers from image digitization error and image noise. These factors introduce uncertainties during the determination of the marker's contour, thereby heavily affecting the detection accuracy of the marker centroid and the accuracy of the OTS. In addition, some objects with a glossy surface may indicate some high-intensity reflections, called noise markers, which can be mistaken as markers. Therefore, a marker detection method must be designed to be capable of eliminating these high-intensity reflections.

Here, we propose a method to detect the correct markers. The centroid of the marker is calculated first in a binary image, and then the contour points of the marker are detected along rays emitted from the calculated centroid in a confined region. Assume that n pixels within a region R_i belong to a marker blob in a binary image and that the coordinates of pixel j inside the region R_i are denoted by $\boldsymbol{m}_j = [u_j, v_j]^{\mathrm{T}}$. The estimated centroid $\hat{\boldsymbol{c}}_i = [u_i^{\hat{c}}, v_i^{\hat{c}}]$ of the ith marker blob that corresponds to region R_i is then calculated by $\hat{\boldsymbol{c}}_i = \frac{\sum_{j=1}^{n} \boldsymbol{m}_j}{n}$. The height H_i and width W_i of the ith marker blob are calculated as

$$\begin{cases} W_i = \max(u_j) - \min(u_j) \\ H_i = \max(v_j) - \min(v_j) \end{cases} \quad j = 1, 2, \cdots, n. \tag{8}$$

Let $J_{rat} = W_i/H_i$, t_1, and t_2 be the pixel aspect ratio of the marker image, threshold value, and the threshold value of the pixel aspect ratio, respectively. The area of the marker image A_i is also computed, and two thresholds of the area A_{t1} and A_{t2} are used to eliminate the noise blobs. The noise blobs are identified by the following criteria using these parameters:

1) The width or height is less than t_1.
2) $\mathrm{Max}(J_{rat}, J_{rat}^{-1})$ is greater than t_2.
3) The area of the marker image A_i satisfies the relationship $A_i < A_{t1}$ or $A_i > A_{t2}$, where $A_{t1} < A_{t2}$.

A rectangular region of height $2H_i$ and width $2W_i$ centered on the calculated centroid of the marker blob is defined for each marker blob, thereby confining the region for detecting the contour of the marker. The contour points are detected along the rays emitted from the calculated centroid such that the rays are calculated at a same interval of angle $\Delta\theta$. The rays are presented as discrete points via the following equations

$$\begin{cases} x_k = u_i^{\hat{c}} + [L + k\Delta l]\cos\theta \\ y_k = v_i^{\hat{c}} + [L + k\Delta l]\sin\theta, \end{cases} \quad k = 1, 2, \cdots, M \tag{9}$$

where $L = \min(W_i, H_i)/4$, Δl is the interval space of any two discrete points on the ray, which is set to 1 pixel in this paper, and $M \leq \max(W_i, H_i)/2\Delta l$. The gradient magnitude along the radial direction is calculated by using the grayscale value of the points (x_k, y_k) and the position next to the maximal gradient magnitude is selected as the contour point. Then, we fit the selected contour points to an ellipse to determine the center of the marker blob. The contour of the fitting ellipse is selected as the contour of the marker blob. The workflow of marker detection is shown in Fig. 2.

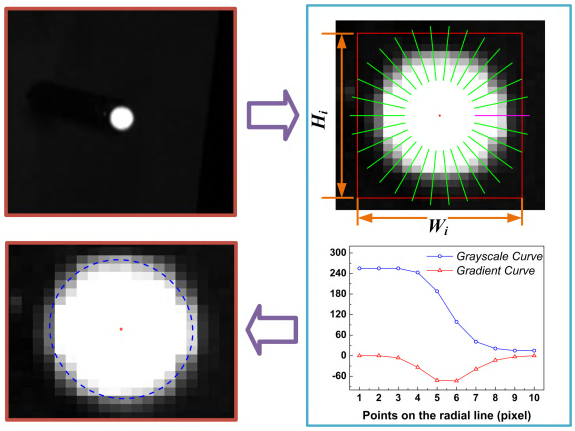

FIGURE 2. Workflow of detecting centroid and edge for a marker blob.

B. SORTING OUT THE CASES OF MORE THAN TWO IMAGE MARKERS ON THE SAME EPIPOLAR LINE

In the stereo-match problem, the epipolar constraint is commonly used in determining correspondences in two images. Any point pair that satisfies the epipolar constraint should lie on the corresponding epipolar line. The epipolar constraint provides a reduced searching space from a 2D image to an epipolar line. In enforcing this constraint, all the cases where more than two markers lie on the same epipolar line can be sorted out, and the cases with only one marker on the corresponding epipolar line can be directly applied to reconstruct

the 3D points by optimal triangulation [30], thereby reducing implementation time.

Assume that the OTS has been calibrated with high accuracy. Thus, the fundamental matrix $\boldsymbol{F}$ is known. Given a measured point $\boldsymbol{p}_l$ in the left image, the corresponding epipolar line $\lambda = \boldsymbol{p}_l\boldsymbol{F}$ can be obtained in the right image. The corresponding point $\boldsymbol{p}_r$ of $\boldsymbol{p}_l$ is usually unable to lie on the line λ exactly because of the error involved in measuring the marker centroids. In reality, the distance between the corresponding point $\boldsymbol{p}_r$ and the line λ is extremely short. We let $\bar{\boldsymbol{p}}_r$ be the orthogonal projections on the line λ. Thus, the perpendicular distance from the point $\boldsymbol{p}_r$ to the epipolar line λ is equal to the distance between the point $\boldsymbol{p}_r$ and $\bar{\boldsymbol{p}}_r$, denoted as $d(\boldsymbol{p}_r, \lambda) = d(\boldsymbol{p}_r, \bar{\boldsymbol{p}}_r)$. A distance threshold d_t is set to find the corresponding point on the line λ. Any point $\hat{\boldsymbol{p}}_r$ on the right image that satisfies the relationship $d(\hat{\boldsymbol{p}}_r, \lambda) \leq d_t$ is considered to lie on the epipolar line λ.

Nevertheless, the marker centroids detected in images are not the real marker centroids because of digitization errors and image noise. Thus, the relationship described by Eq. (5) is not equal to zero and yields a small value. The error introduced by the detection of marker centroids will lead to a large error in the following step of 3D point reconstruction. Involving this error is necessary for obtaining an improved estimation. The coordinates of the detected image points are subject to Gaussian noise such that a deviation of Gaussian distribution exists between the detected image points and their correct coordinates in an image. Given a measured corresponding point pair $\boldsymbol{p}_l \leftrightarrow \boldsymbol{p}_r$ with noise, the correct coordinates of the point pair $\boldsymbol{p}'_l \leftrightarrow \boldsymbol{p}'_r$ should be close to the point pair $\boldsymbol{p}_l \leftrightarrow \boldsymbol{p}_r$ and satisfy the Eq. (5). We will consider the following error into the cost function and seek the solution to find the correct corresponding point pair $\boldsymbol{p}'_l \leftrightarrow \boldsymbol{p}'_r$

$$\boldsymbol{E}^d = d(\boldsymbol{p}_l, \boldsymbol{p}'_l)^2 + d(\boldsymbol{p}_r, \boldsymbol{p}'_r)^2, \tag{10}$$

where $d(*, *)$ represents the Euclidean distance between two points, and this function is subject to the epipolar constraint $\widetilde{\boldsymbol{p}}'^{\mathrm{T}}_r \boldsymbol{F} \widetilde{\boldsymbol{p}}'_l = 0$, $\widetilde{\boldsymbol{p}}'_l$ and $\widetilde{\boldsymbol{p}}'_r$ are the homogeneous coordinates of $\boldsymbol{p}'_l$ and $\boldsymbol{p}'_r$.

The 3D points are easily reconstructed via an optimal triangulation method proposed by Hartley and Strum [30] when only one 2D point is presented on the corresponding epipolar line. However, when several markers are coplanar with two optical centers in a 3D space, several points are on the same epipolar line. Distinguishing the correct correspondences of point pairs from these image points on an epipolar line without any other constraint is impossible. Some constraints, such as the geometrical constraint of distances between points used in [19] and ordering constraint [31], are used for finding the correct correspondences. However, these constraints are inconsistent because the stray markers are independent and can move in 3D space. Thus, the false correspondences of points will cause ghost markers to appear in the space.

To clearly explain the indistinct of correspondence when multiple markers are located in a same plane with the two optical centers, we take the case of two markers coplanar with two optical centers to describe the phenomenon of ghost markers. As shown in Fig. 3, the centroids of the projection images of the two markers are denoted as $\boldsymbol{p}_{l1}$, $\boldsymbol{p}_{l2}$ and $\boldsymbol{p}_{r1}$, $\boldsymbol{p}_{r2}$ in the two image planes. The four rays that connect the two markers with two optical centers intersect each other at four 3D points $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, $\boldsymbol{P}'_1$, and $\boldsymbol{P}'_2$, as shown in Fig. 3. If the correct corresponding pairs are $\boldsymbol{p}_{l1} \leftrightarrow \boldsymbol{p}_{r1}$ and $\boldsymbol{p}_{l2} \leftrightarrow \boldsymbol{p}_{r2}$ in the image planes, the corresponding 3D points are related to $\boldsymbol{P}'_1$ and $\boldsymbol{P}'_2$, whereas the other two 3D points $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are ghost markers which are called vertical ghost markers. Similarly, when the correct corresponding pairs are $\boldsymbol{p}_{l1} \leftrightarrow \boldsymbol{p}_{r2}$ and $\boldsymbol{p}_{l2} \leftrightarrow \boldsymbol{p}_{r1}$ in the image planes, the corresponding 3D points are related to $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, whereas the other two 3D points $\boldsymbol{P}'_1$ and $\boldsymbol{P}'_2$ are ghost markers which are called horizontal ghost markers. Clearly the ordering constraint, which is commonly used for stereo matching, are unable to confirm whether the correspondences are corrected because the reflective markers are stray distributed in the 3D space. Thus, the method is insufficient in identifying the corrected correspondences from the image points on a same epipolar line when two or more markers are coplanar.

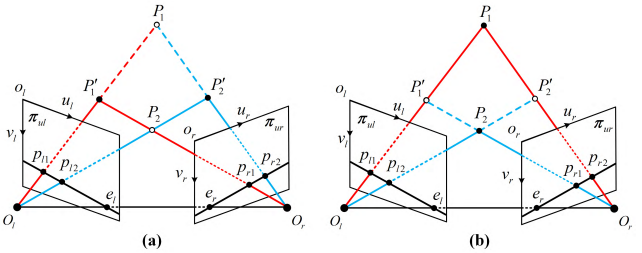

FIGURE 3. False correspondences cause by two markers coplanar with two optical centers that lead to ghost markers. (a) When the correct corresponding pairs are $p_{l1} \leftrightarrow p_{r1}$ and $p_{l2} \leftrightarrow p_{r2}$ in the left and right image planes, the real markers are P'_1 and P'_2, while P_1 and P_2 are ghost markers. (b) If the correct corresponding pairs are $p_{l1} \leftrightarrow p_{r2}$ and $p_{l2} \leftrightarrow p_{r1}$ in the left and right image planes, the real markers are P'_1 and P_2, whereas the ghost markers are P'_1 and P'_2.

C. OBJECTIVE FUNCTION WITH SILHOUETTE REPROJECTION ERROR

As described in the previous subsection, the sparse stereo-match method combined with other constraints, such as epipolar and ordering constraints, cannot address the issue of multiple markers coplanar with two optical centers. Nevertheless, the case of multiple markers located in the same plane with two optical centers is usually encountered in clinical applications wherein OTS is used [16]–[18]. To the best of our knowledge, the existing stereo-match methods are incapable of addressing this case in real time. Thus, false stereo matching frequently occurs and leads to the appearance of ghost markers, which render the OTS to send the false 3D coordinates of reflective markers to the navigation system. Consequently, the navigation system stops working or the treatment fails. Therefore, to satisfy the request of OTS in clinical application, we propose a novel optimization strategy

that addresses the case wherein multiple markers are located in the same plane with the two optical centers, which combines the epipolar constraint and silhouette reprojection error into a cost function.

Suppose that two sets of N points lay on the same epipolar line are detected on the two images, respectively. $G = N!$ groups enumerate all possible pairing of points that each group contains N point pairs. For each point pair $\boldsymbol{p}_{lij} \leftrightarrow \boldsymbol{p}_{rij}$, where $i \in \{1, 2, \cdots, N\}$ and $j \in \{1, 2, \cdots, G\}$ represent the ith point pair in the jth group, assume that the corresponding marker blobs are denoted as $\boldsymbol{I}_{lij} : \boldsymbol{D}_{lij} \subset \mathbb{R}^2$ and $\boldsymbol{I}_{rij} : \boldsymbol{D}_{rij} \subset \mathbb{R}^2$, of which the centroids are $\boldsymbol{p}_{lij}$ and $\boldsymbol{p}_{rij}$. Given the intrinsic and extrinsic parameters of the left and right cameras, the 3D point $\boldsymbol{X}_{ij}$ that corresponds to $\boldsymbol{p}_{lij}$ and $\boldsymbol{p}_{rij}$ can be reconstructed via triangulation. Then, a virtual sphere $\boldsymbol{V}_{ij}$ is constructed and centered on the 3D point $\boldsymbol{X}_{ij}$ with a radius r. The surface of the sphere $\boldsymbol{V}_{ij}$ is denoted as $\boldsymbol{S}_{ij}(r)$. Let $\boldsymbol{\pi}_{lij} : \mathbb{R}^3 \rightarrow \boldsymbol{D}_{lij}$ and $\boldsymbol{\pi}_{rij} : \mathbb{R}^3 \rightarrow \boldsymbol{D}_{rij}$ be the projection mapping from 3D point $\boldsymbol{X}_{ij}$ to 2D image point $\boldsymbol{p}_{lij}$ and $\boldsymbol{p}_{rij}$ in the left and right images, respectively. $\boldsymbol{\pi}_{lij}^{-1} : \boldsymbol{D}_{lij} \rightarrow \mathbb{R}^3$ and $\boldsymbol{\pi}_{rij}^{-1} : \boldsymbol{D}_{rij} \rightarrow \mathbb{R}^3$ are an inverse projection from the 2D image points $\boldsymbol{p}_{lij}$ and $\boldsymbol{p}_{rij}$ to the marker surface $\boldsymbol{S}_{ij}(r)$. As the 3D space scene includes the reflective spherical marker and the background, the background is located at infinity, and its projection image in the image plane is totally black. Then, the reprojection image $\hat{\boldsymbol{I}}_{kij} : \boldsymbol{D}_{kij} \subset \mathbb{R}^2$ (where $k = \{l, r\}$) can be obtained via the inverse projection point $\boldsymbol{p}_{kij}$ to the 3D real world. Visual rays emit from the optical center O_k and back project to the scene of the 3D real world. Then, the visual rays intersect with the surface of the reflective marker at a point or intersect with the background at infinity. If a ray intersects with the reflective marker, the image point of the intersection point between the ray and the surface of the reflective marker is assigned to 1. Otherwise, the image point is assigned to 0 when the ray intersects with the background. Thus, the reprojection image $\hat{\boldsymbol{I}}_{kij}$ from the surface $\boldsymbol{S}_{ij}(r)$ of the sphere $\boldsymbol{V}_{ij}$ to the image plane, which is visible to the camera k, is generated as follows

$$\hat{\boldsymbol{I}}_{kij}(p) = \begin{cases} 1 & \boldsymbol{\pi}_{kij}^{-1}(p) \in \boldsymbol{S}_{ij}(r) \\ 0 & \text{else,} \end{cases} \quad k = l, r. \tag{11}$$

The binary images $\boldsymbol{H}_{lij}$ and $\boldsymbol{H}_{rij}$ of image $\boldsymbol{I}_{lij}$ and $\boldsymbol{I}_{rij}$ are defined as $\boldsymbol{H}_{lij} : \boldsymbol{D}_{lij} \rightarrow \{0, 1\}$ and $\boldsymbol{H}_{rij} : \boldsymbol{D}_{rij} \rightarrow \{0, 1\}$ whose value is assigned to 1 as it lies inside or on the bright blob and 0 otherwise. The binary image $\boldsymbol{H}_{kij}$ is then denoted as

$$\boldsymbol{H}_{kij}(p) = \begin{cases} 1 & p \in \boldsymbol{I}_{kij} \\ 0 & \text{else,} \end{cases} \quad k = l, r \tag{12}$$

Let $\phi_k : \boldsymbol{D}_{kij} \rightarrow \{0, 1\}$ be a binary function that measures the inconsistency between the binary image $\boldsymbol{H}_{kij}$ and reprojection image $\hat{\boldsymbol{I}}_{kij}$, 1 for $\boldsymbol{H}_{kij}(p)_{kij} \neq \hat{\boldsymbol{I}}_{kij}(p_{kij})$, and 0 for $\boldsymbol{H}_{kij}(p)_{kij} = \hat{\boldsymbol{I}}_{kij}(p_{kij})$. Then, the error term $\boldsymbol{E}_{ij}^S(r)$ measured silhouette reprojection error for image point pair $\hat{\boldsymbol{I}}_{lij} \leftrightarrow \hat{\boldsymbol{I}}_{rij}$ is thus expressed as

$$\boldsymbol{E}_{ij}^S(r) = \int_{\boldsymbol{D}_{lij}} \phi_l(\boldsymbol{H}_{lij}(p), \hat{\boldsymbol{I}}_{lij}(p))dp + \int_{\boldsymbol{D}_{rij}} \phi_r(\boldsymbol{H}_{rij}(p), \hat{\boldsymbol{I}}_{rij}(p))dp, \tag{13}$$

where dp is the area measured in the image plane. We minimize Eq. (13) by forcing the silhouette reprojection image of the virtual sphere's surface to be consistent with the observed marker image, which can acquire the virtual sphere's radius r approximated to the real radius for the correct correspondences because it has lower error when the reprojection image approximates the observed marker image.

As the OTS is essentially a binocular vision system, the solution of the stereo-match problem is the establishment of an objective function that integrates the silhouette reprojection error and epipolar constraint. Given the marker image pair $\boldsymbol{I}_{li} \leftrightarrow \boldsymbol{I}_{ri}$, the objective function is presented in the following form:

$$\boldsymbol{E}_{ij}(r, p_{lij}, p_{rij}) = \boldsymbol{E}_{ij}^d + \boldsymbol{E}_{ij}^S \text{ subject to } \widetilde{\boldsymbol{p}}_{rij}'^{\mathrm{T}} \boldsymbol{F} \widetilde{\boldsymbol{p}}_{lij}' = 0, \tag{14}$$

where $\boldsymbol{E}_{ij}^d$ is related to Eq. (10). Thus, the complete objective function for the jth group can be expressed as follows:

$$\boldsymbol{E}_j = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{E}_{ij}. \tag{15}$$

Intuitively, the group that includes the correct corresponding image pairs has the maximum silhouette consistency with the observed images and corresponds to the minimum of error defined in Eq. (15). Thus, the stereo-match problem is reduced such that only determining the group with the minimum error among all the combination groups is the sole concern.

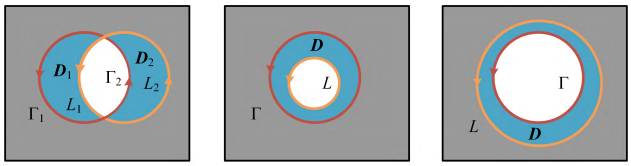

FIGURE 4. Relative position between the reprojection image and observed image. (a) The reprojection image intersects with the observed image; (b) the reprojection image is inside the observed image; (c) the reprojection image covers the observed image.

D. OBJECTIVE FUNCTION MINIMIZATION

To minimize Eq. (13), we must establish the relationship between the parameter set $\{r, p_{lij}, p_{rij}\}$ and the observed markers. First, we should parameterize the relationship between the reprojection images and the observed images. The relative position between the reprojection images and the observed images includes three situations, as shown in Fig. 4. Obviously, the silhouette reprojection error can be indicated

in the area filled with blue, and it is expressed as

$$\boldsymbol{E}_{ij}^{S}(r) = \int_{\boldsymbol{D}_1} dp + \int_{\boldsymbol{D}_2} dp \tag{16}$$

where $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are the total area of the observed and reprojection images subtracted to the intersection part. Following Green's theorem, the above equation can be reformulated as follows

$$\int_{\boldsymbol{D}_1} dp + \int_{\boldsymbol{D}_2} dp = \frac{1}{2}\int_{\Gamma_1 - L_1}(xdy - ydx) + \frac{1}{2}\int_{L_2 - \Gamma_2}(xdy - ydx) \tag{17}$$

where $\Gamma_1 + \Gamma_2 = \Gamma$ is the contour of the observed image, and $L_1 + L_2 = L$ is the contour of the reprojection image. Obviously, when the contour L is close to the contour Γ, the left term of Eq. (17) is decreased to zero. Thus, the problem reduces to find the optimal contour L_{opt} that approximates the contour Γ, which is essentially a least squares problem. Suppose that we have determined all the points p_k on the contour Γ and searched the corresponding points $\hat{p}_k(r, p_{lij}, p_{rij})$ on the contour L. Then, minimizing Eq. (13) may reformulate, thereby minimizing the function presented in the following form

$$\frac{1}{N_1}\sum_{k}^{N_1}\|p_{lk} - \hat{p}_{lk}(r, p_{lij}, p_{rij})\|^2 + \frac{1}{N_2}\sum_{k}^{N_2}\|p_{rk} - \hat{p}_{rk}(r, p_{lij}, p_{rij})\|^2, \tag{18}$$

where N_1 and N_2 are the number of points on the contours in the left and right images, respectively. Assigning correspondences between the point sets of p_k and $\hat{p}_k(r, p_{lij}, p_{rij})$ is essential in minimizing the above expression. The point p_k on the contour Γ of the marker blob is easy to obtain using the method proposed in the Part A of Section III. The corresponding point $\hat{p}_k(r, p_{lij}, p_{rij})$ on the contour L then can be determined based on the point p_k. One of the best choices is selected the closest point from the point p to the contour L as the corresponding point $\hat{p}_k(r, p_{lij}, p_{rij})$. Thus, the optimization method can be used for the minimization of the pixel distance between the observed point and reprojection point in the 2D image plane. Nevertheless, the final goal of this paper is to determine the real location and radius of the marker in 3D space and the real corresponding relationship of the image points in two image planes. The drawback of minimizing the pixel distance in the 2D image plane is the inconsistencies with the inverse projection distance error in 3D space that different inverse projection distance errors may relate to the same pixel distance error [32], [33]. Therefore, to increase the robust of the stereo-match method, we propose a strategy to optimize the distance error in 3D space. The detailed optimization method is described below.

Considered the intersection point Q of two rays and denoted $Q = [x_Q, y_Q, z_Q]^{\mathrm{T}}$ as the coordinate in camera frame system. Assume that the corresponding marker centroid is denoted as q^d and a point on its contour is denoted as p^d.

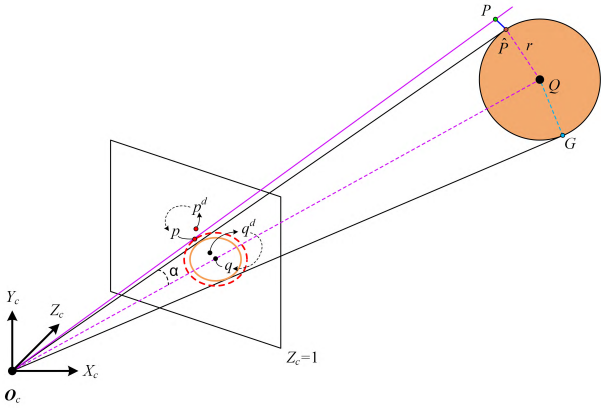

FIGURE 5. Optimizing corresponding points between the contour of the observed image and the contour of the reprojection image in 3D space.

Let $p = [u_p, v_p]^{\mathrm{T}}$ and $q = [u_q, v_q]^{\mathrm{T}}$ be the undistorted coordinates of q^d and p^d, respectively, as shown in Fig. 5. The problem of minimizing the 3D distance error is equivalent to finding the closest point from the optical ray $\overrightarrow{O_cp}$ to the projection contour of the sphere. The projection contour on the sphere is a circle in 3D space. This circle is the set of the tangent point between the rays emitted from the camera's optical center and the spherical marker. Denote a point on the projection contour of the sphere as $G = [x, y, z]^{\mathrm{T}}$. The geometrical relationship between the optical ray $\overrightarrow{O_cG}$ and the line $\overrightarrow{QG}$ is satisfied by $\overrightarrow{O_cG} \perp \overrightarrow{QG}$. Thus, the analytic expression of the projection contour on the sphere is easy to present as follows

$$\begin{cases} \|\overrightarrow{O_cG}\|^2 + \|\overrightarrow{QG}\|^2 = \|\overrightarrow{O_cQ}\|^2 \\ \|\overrightarrow{QG}\|^2 - r^2 = 0. \end{cases} \tag{19}$$

The closest point on the projection contour to the optical ray $\overrightarrow{O_cp}$ is denoted as $\hat{P} = [x_{\hat{p}}, y_{\hat{p}}, z_{\hat{p}}]^{\mathrm{T}}$, which must lie on the plane Π formed by two optical rays $\overrightarrow{O_cp}$ and $\overrightarrow{O_cQ}$. Let $\xi = [x_\Pi, y_\Pi, z_\Pi]^{\mathrm{T}}$ be an arbitrary point on the plane Π. Algebraically, the plane Π is expressed as

$$(\overrightarrow{O_cp} \times \overrightarrow{O_cQ}) \cdot \overrightarrow{O_c\xi} = 0. \tag{20}$$

Clearly, by combining Eqs. (19) and (20), the expression of the closest point $\hat{P} = [x_{\hat{p}}, y_{\hat{p}}, z_{\hat{p}}]^{\mathrm{T}}$ can be solved from the equations

$$\begin{cases} \|\overrightarrow{O_c\hat{P}}\|^2 + \|\overrightarrow{Q\hat{P}}\|^2 = \|\overrightarrow{O_cQ}\|^2 \\ \|\overrightarrow{Q\hat{P}}\|^2 - r^2 = 0 \\ (\overrightarrow{O_cp} \times \overrightarrow{O_cQ}) \cdot \overrightarrow{O_c\hat{P}} = 0. \end{cases} \tag{21}$$

Denote the solution of Eq. (21) as $[x_{\hat{p}}, y_{\hat{p}}, z_{\hat{p}}]^{\mathrm{T}} = M(r, x_Q, y_Q, z_Q, u_p, v_p)$. Basing on Eqs. (1) and (4), we can express the point $Q = [x_Q, y_Q, z_Q]^{\mathrm{T}}$ as a matrix function of parameter vector of $[u_{lq}, v_{lq}, u_{rq}, v_{rq}]^{\mathrm{T}}$. Thus, $[x_{\hat{p}}, y_{\hat{p}}, z_{\hat{p}}]^{\mathrm{T}} = M(r, x_Q, y_Q, z_Q, u_p, v_p)$ can be reformulated

as $[x_{\hat{p}}, y_{\hat{p}}, z_{\hat{p}}]^{\mathrm{T}} = M'(r, u_{lq}, v_{lq}, u_{rq}, v_{rq}, u_p, v_p)$. The perpendicular point from the closest point $\hat{P}$ to the ray $\overrightarrow{O_c p}$ is denoted as $P = [x_p, y_p, z_p]^{\mathrm{T}}$ such that it can be expressed by the point $\hat{P}$ and the ray $\overrightarrow{O_c p}$ as

$$\|\overrightarrow{O_c P}\|^2 + \|\overrightarrow{P\hat{P}}\|^2 = \|\overrightarrow{O_c \hat{P}}\|^2. \tag{22}$$

Thus, the goal is to minimize the function of distance error in 3D space as follows:

$$\boldsymbol{E}_{ij}^{S} = \frac{1}{N_1}\sum_{k=1}^{N_1}\|P_{lijk} - \hat{P}_{lijk}\| + \frac{1}{N_2}\sum_{k=1}^{N_2}\|P_{rijk} - \hat{P}_{rijk}\|, \tag{23}$$

where $\hat{P}_{lijk}$ is the kth point on the left projection contour of the sphere, $\hat{P}_{rijk}$ is the kth point on the right projection contour of the sphere, and P_{lijk} and P_{rijk} are the corresponding points of the points $\hat{P}_{lijk}$ and $\hat{P}_{rijk}$. Given an appropriate initial value to the parameter vector of $[r, u_{lq}, v_{lq}, u_{rq}, v_{rq}]^{\mathrm{T}}$, the parameter vector can be solved via the Levenberg-Marquardt (LM) algorithm. We summarize the MSRC algorithm in Algorithm 1.

Algorithm 1 Maximum Silhouette Reprojection Consistency (MSRC) Algorithm

Require: $\{q_{li}^d \mid i = 1, \cdots, N\}$, $\{q_{ri}^d \mid i = 1, \cdots, N\}$: the marker centroid point sets lay on the same epiploar line; $\{p_{lk}^d\}_i$, $\{p_{rk}^d\}_i$: the contour point sets of the markers q_{li}^d and q_{ri}^d;

Ensure: $\{q_{li}^d \leftrightarrow q_{ri}^d\}_{min}$: the correspondence of pair points of the group obtained the minimum of the silhouette reprojection error; $\boldsymbol{E}_{min}$: the minimum of the silhouette reprojection error;

compute the undistorted points q_{li}, q_{ri}, p_{lk}, and p_{rk} of the corresponding distorted points q_{li}^d, q_{ri}^d, p_{lk}^d, and p_{rk}^d according to Eq. (3);
compute $G = N!$ groups of point pairs that the point correspondence $\{q_{li}^d \leftrightarrow q_{ri}^d\}$ is unique for each group;
for each j **do**
  for each i **do**
    compute 3D point $\boldsymbol{Q}_{ij}$ through the optimal triangulation method [30];
    construct a virtual sphere $\boldsymbol{V}_{ij}$ centered on the 3D point $\boldsymbol{Q}_{ij}$ with a radius r;
    for each k **do**
      compute $\hat{P}_{lijk}$ and $\hat{P}_{rijk}$ according to Eqs. (1), (4), and (21);
      compute P_{lijk} and P_{rijk} according to Eq. (22);
    end for
    construct $\boldsymbol{E}_{ij}$ according to Eqs. (23) and (10);
    compute the optimal solution $\boldsymbol{E}_{ij}^{opt}$ by employing LM algorithm;
  end for
  compute $\boldsymbol{E}_j = \sum_i \boldsymbol{E}_{ij}^{opt}$;
  if j==1 **then**
    $\boldsymbol{E}_{min} = \boldsymbol{E}_1$;
    $\{q_{li}^d \leftrightarrow q_{ri}^d\}_{min} \leftarrow \{q_{li}^d \leftrightarrow q_{ri}^d\}_1$;
  else if $\boldsymbol{E}_j < \boldsymbol{E}_{min}$ **then**
    $\boldsymbol{E}_{min} = \boldsymbol{E}_j$;
    $\{q_{li}^d \leftrightarrow q_{ri}^d\}_{min} \leftarrow \{q_{li}^d \leftrightarrow q_{ri}^d\}_j$;
  end if
end for
return $\{q_{li}^d \leftrightarrow q_{ri}^d\}_{min}$ and $\boldsymbol{E}_{min}$.

IV. EXPERIMENTAL RESULTS

The proposed stereo-match algorithm for OTS combining silhouette reprojection error with epipolar constraint is conducted on both simulative and real data.

A. SIMULATION EXPERIMENTS

In the simulation experiments, the projection images of the reflective spheres on the left and right images are synthesized. We aim to assess the performance of MSRC method. Adopting the camera parameters based on real results of camera calibration is reasonable in creating synthetic images. The size of each simulative image is 1600×1200 pixels such that the principal point is defined as $[u_0, v_0]^{\mathrm{T}} = [800, 600]^{\mathrm{T}}$ pixel and the skew factor is assumed to be zero. The two focal lengths are defined as $f_u = 2700$ pixel and $f_v = 2700$ pixel in the u and v directions, respectively. The distortion coefficients are defined as $k_1 = 0.1\,\mathrm{mm}^{-1}$, $k_2 = 0.17\,\mathrm{mm}^{-1}$, and the distortion center is assumed to be the principal point. The rotation vector $\boldsymbol{r}$ and translation vector $\boldsymbol{t}$ between the camera coordinate frames is defined as $\boldsymbol{r} = [0.004, -0.380, -0.004]^{\mathrm{T}}$ and $\boldsymbol{t} = [350, 0, 0]^{\mathrm{T}}$, respectively. The reflective spheres with radius $r = 5.75$ mm are randomly distributed in the front of the two cameras within the field of view of OTS. We know the intrinsic and extrinsic parameters of OTS; thus, the spheres in 3D space that are coplanar with the two optical centers are easily projected to the left and right image planes to create the synthetic projection images. The simulative images are captured randomly in the position of which the distance from the markers to the camera is 800−1500 mm, and the distance from the markers to the XOZ plane of the left camera system is 0−400 mm. In the clinical application of OTS, more than four markers rarely appear in the same plane with two optical centers. Therefore, from a statistical and practical point of view, we consider the case of three markers in the same plane with two optical centers to validate the MSRC method in the simulation experiments.

Two main types of noise affect the detection of image contours, thereby affecting the robustness of the stereo-match results and the accuracy of the 3D point reconstruction. The first type of noise is the digitization error of the image contour, and the second type of noise is the identification error of the image contour. We consider these two types of noise. To model the digitization error of the image contour, we add Gaussian noise to the coordinates of contours of the projection images such that its mean is set to zero and its standard deviations range from 0 to 1 pixels. Then, these coordinates are directly used for the MSRC method. To model

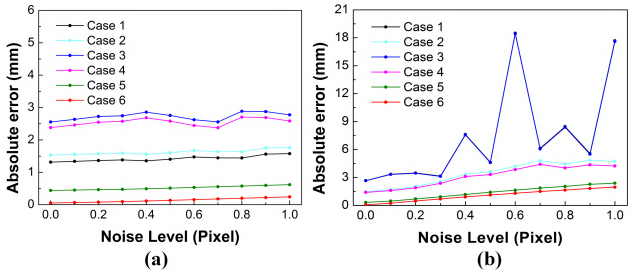

FIGURE 6. Effects of pixel noise on the detection of the image contour affect the average absolute error of the proposed stereo-match method. (a) The effect of the digitization error; (b) the effect of the identification error.

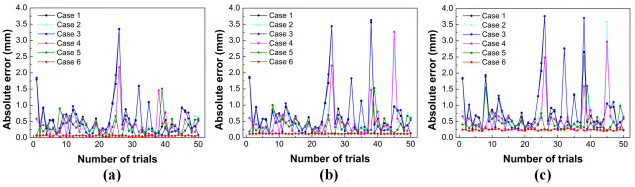

FIGURE 7. Absolute errors of fifty trials for three noise levels cause by the digitization error on the detection of the marker contour. (a) The noise level of zero pixels; (b) the noise level of 0.4 pixels; (c) the noise level of 1 pixel.

the detection error of image contour, a 3-by-3 Gaussian filter of mean 0 and standard deviations that range from 0 to 1 pixels is first added to the projection image areas, and then the marker contours are detected by the method described in the Part A of Section III. Fifty independent trials are performed for each noise level of each type of noise.

To clearly show the experimental results, we denote the correct correspondences as Case 6 and denote the false corresponding group as Cases 1−5, respectively. The average absolute errors of the objective function for six cases are shown in Fig. 6. Eleven noise levels (including a zero-noise level) are used for each type of noise. We selected three noise levels for each type of noise to show the experimental results: the noise level of zero pixel, the noise level of 0.4 pixels, and the noise level of 1 pixel, as shown in Figs. 7 and 8. The experimental results show that the proposed method can find the correct correspondences in all the cases. As evident in the figures, the absolute errors of the case of the correct correspondences are minimal among the six cases in 50 independent trials. The MSRC method obviously produces a smaller error for the correct stereo-match correspondences than the false stereo-match correspondences, because the parameters of the correct correspondences are best consistent with the ground truth.

To validate the influent of noise on the parameters involving in the MSRC method, we computed the errors of the parameters in obtaining the correct correspondences. The absolute errors of the marker radius r and the absolute errors of the 3D reconstructed point Q are calculated. Figure 9 shows the influence of different noise levels for each type of noise on the parameter vector of

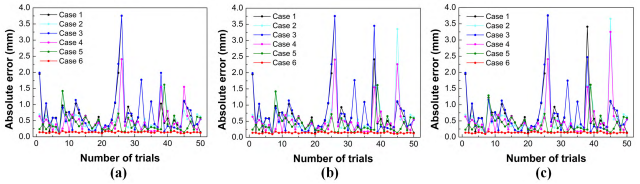

FIGURE 8. Absolute errors of fifty trials for three noise levels cause by the identification error on the detection of the marker contour. (a) The noise level of zero pixel; (b) the noise level of 0.4 pixel; (c) the noise level of 1 pixel.

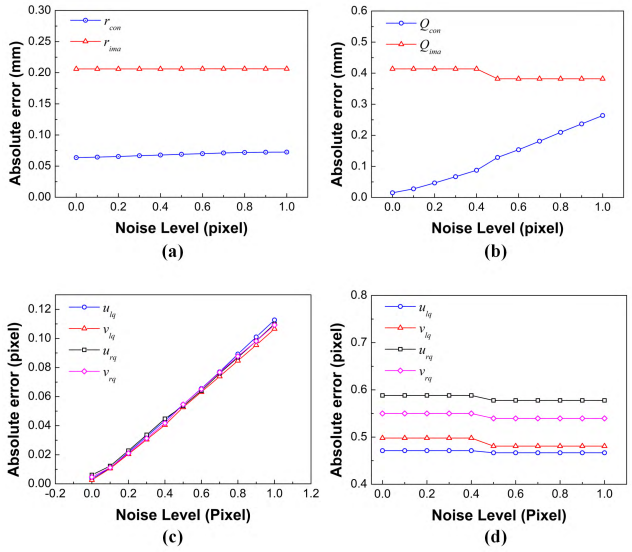

FIGURE 9. Effects of detection error of the marker contour on the parameters involving in the proposed stereo-match method. (a) Reconstructed radius r of the spherical marker; the red curve represents the effect of the digitization error and the blue curve represents the effect of the identification error; (b) the coordinates of the 3D reconstruction marker centers; the red curve represents the effect of the digitization error and the blue curve represents the effect of the identification error; (c) the pixel coordinates of the marker blob's centroids affected by the digitization error; (d) the pixel coordinates of the marker blob's centroids affected by the identification error.

$[r, Q, u_{lq}, v_{lq}, u_{rq}, v_{rq}]^{\mathrm{T}}$. As can be observed from Fig. 9(a), the influence of both noises on the marker radius is minimal, whereas the error caused by the edge detected error is larger than the error caused by the digitization error. The absolute error of the 3D reconstructed point Q increases nearly linearly with the noise level after the addition of the digitization error, as shown in Fig. 9(b). For the edge-detected error, the absolute error of the 3D reconstructed point Q is maintained at the same level for noise levels of 0−0.4 pixel, and it is reduced in the noise level of 0.5 pixel. Then, it changed slightly from noise levels of 0.5−1 pixels. The absolute errors of parameters $(u_{lq}, v_{lq}, u_{rq}, v_{rq})$ increase almost linearly with the noise level caused by the digitization error. The absolute errors of $(u_{lq}, v_{lq}, u_{rq}, v_{rq})$ caused by the edge detected error have the same tendency with the absolute error of the 3D reconstructed point Q; this situation can be explained by the digitization error involved in the edge error has a slight influence on the

parameters $(u_{lq}, v_{lq}, u_{rq}, v_{rq})$. The reducing error occurred in the noise level of 0.5 pixels can be ascribed to the digitization error in which the intersection of the sight ray and the image plane will take a pixel even when the real intersection part is less than half of one pixel. These results indicate that the edge detected error plays a more important role on the parameter vector of $[r, Q, u_{lq}, v_{lq}, u_{rq}, v_{rq}]^{\mathrm{T}}$ than the digitization error. For two types of noise, the maximum error of the marker radius is less than 0.22 mm, the maximum error of the 3D reconstructed point Q is less than 0.45 mm, and the maximum error of the parameters $(u_{lq}, v_{lq}, u_{rq}, v_{rq})$ are below 0.6 pixels. Nevertheless, the parameters calculated by the MSRC method are sufficiently accurate from a practical point of view. Therefore, the results of the simulative experiments demonstrate that the MSRC method is accurate and robust to both types of image noise.

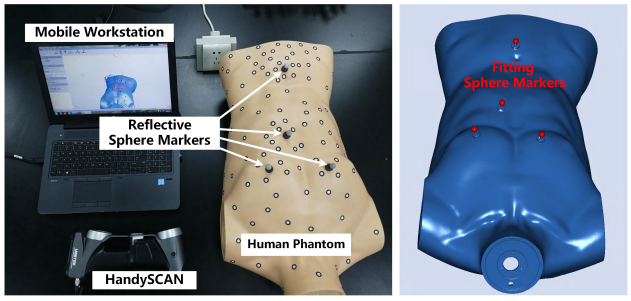

FIGURE 10. Scanning the entire human phantom with the four reflective spherical markers adhered on the surface. (a) Setup for the 3D scanning of the human phantom; (b) fitting the spherical markers from the point cloud data.

B. EVALUATION WITH REAL DATA

The MSRC method is applied in real image data. A human phantom with four reflective spherical markers adhered in the surface is made. A hand-held portable 3D scanner (HandySCAN; Creaform, Levis, Quebec, Canada) is used to acquire a point cloud of the skin surface of the entire human phantom that it has a very high accuracy (up to 0.1 mm), as shown in Fig. 10. The coordinates of the four spherical markers can be fitting from the point cloud data which are regarded as ground truth for the experiments. The mean fitting error of the spherical markers' raduises is less than 0.1 mm. We constructed an OTS that can track the reflective spherical markers at 60 frames per second to perform the experiments. The detailed design of the OTS is seen in our previous works [15], [34]. The human phantom is placed in the field of view of the OTS, the image pairs of three markers coplanar with the two optical centers are then captured by moving the position and orientation of the human phantom. Ghost markers can be divided into two categories: vertical and horizontal ghost markers, as showed in Fig. 3. If the line connecting the two ghost markers are more vertical to the line connecting the two sensors, these ghost markers are classified as vertical ghost markers, otherwise they are classified as horizontal ghost markers. A data set containing 30 image pairs is captured for each category of ghost markers. Another data set containing 30 image pairs is also captured as the usual testing data in which any two markers are not coplanar with the two sensors. We perform the MSRC method in these three data sets to investigate its validity and robustness, and evaluate its performance in comparison with the epipolar constraint (E-Constraint) method and the epipolar combined with ordering constraint (EO-Constraint) method.

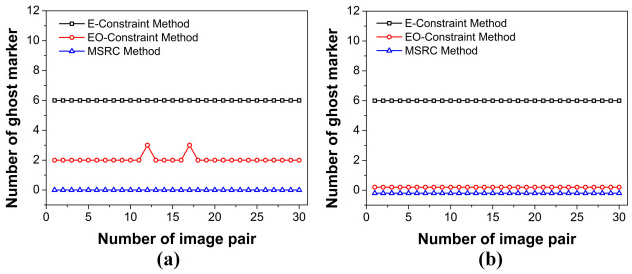

FIGURE 11. Appearance of ghost markers by implementing different methods on the two data sets. (a) Implementing different methods on the data set containing the vertical ghost markers; (b) implementing different methods on the data set containing the horizontal ghost markers.

For the usual testing data, all the three methods can find the correct corresponding points and reconstruct the coordinates of all the markers. The reconstruction coordinates of the spherical markers are registered with the coordinates of the fitting markers (ground truth data). The mean registration errors are 0.1 mm, 0.1 mm, and 0.1 mm for the E-Constraint method, the EO-Constraint method, and the MSRC method, respectively. Figure 11 shows the results of the appearance of ghost markers by performing the three methods on the data sets containing the vertical and horizontal ghost markers. The E-Constraint method reconstructs 3×3 markers (three real markers and six ghost markers) while testing it on these two data sets. The results indicate that the epipolar constraint is unable to eliminate the ghost markers when two or more markers coplanar with the two optical centers. The presence of ghost markers in this situation can be observed in the Polaris tracking system (Northern Digital, Inc., ON, Canada) which is applying the E-Constraint method [19], [27]. The EO-Constraint method is performed well on the data set containing the vertical ghost markers due to the images of markers are arranged in the same ordering along the epipolar line in the left and right images. However, ghost markers are presented when the EO-Constraint method is performed on the data set containing the horizontal ghost markers due to the orderings of points are diferent in the two images. In some situations (e.g. the results of the 12_{th} and 17_{th} image pairs), three ghost markers are presented as their orderings in the two images are totally different (e.g. $[p_{l1}, p_{l2}, p_{l3}] \leftrightarrow [p_{r3}, p_{r1}, p_{r2}]$). As expected, ghost markers do not presented when performing the MSRC method on the two data sets, which turn out that the MSRC method is able to eliminate ghost markers.

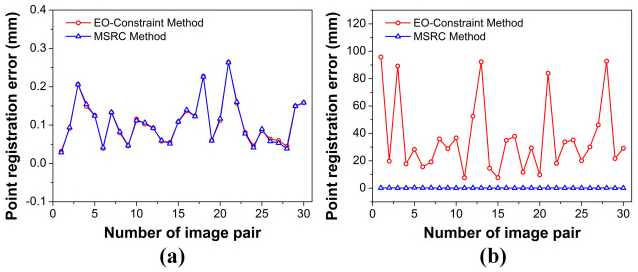

FIGURE 12. Point registration error of the EO-Constraint method and the MSRC method. (a) Point registration error obtained on the data set containing the vertical ghost markers; (b) point registration error obtained on the data set containing the horizontal ghost markers.

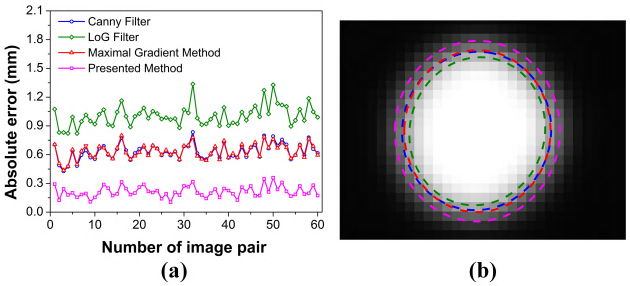

FIGURE 13. Absolute error between the radius of the reconstructed marker and the radius of the real spherical marker using the marker contours detected by Canny filter, Laplacian of Gauss filter, maximal gradient method and the presented method to apply on the two data sets containing ghost markers. (a) The absolute error; (b) one example of the detected contours.

Figure 12 shows the comparative results of the point registration error between the fitting markers and the reconstruction markers performed by the EO-Constraint and MSRC methods. The average errors of the point registration error are 0.11 mm and 0.11 mm for the EO-Constraint and MSRC methods performed on the data set containing the vertical ghost markers, respectively. Nevertheless, the point registration error ranges from 7.61 mm to 95.78 mm for the EO-Constraint method performed on the data set containing the image pairs of horizontal ghost markers, while the average error of the point registration error is 0.13 mm for the MSRC method. It is noticeable that the MSRC method proposed in this article finds the correct correspondences in all the cases relative to the E-Constraint and EO-Constraint methods, since the information including in the MSRC method is more close to the ground truth.

To further evaluate the availability of the MSRC method, the radiuses of the reconstruction spherical markers are compared with the real radius of the spherical marker. Canny filter, Laplacian of Gauss filter, and the method described in the Part A of Section III are applied to detect the contours of the marker images in the two data sets containing the vertical and horizontal ghost markers. These three contour-detecting methods (coded in C++ with the OpenCV library) are performed on an Intel(R) Xeon(R) E3-1225 V2 CPU (3.20 GHz) with 8 GB RAM running under Windows 7. The average running time of detecting all contours in an image pair is 0.0106 s, 0.0106 s, and 0.0069 s for Canny filter, Laplacian of Gauss filter, and the presented method, respectively. The contours with maximal gradient magnitude are obtained through our presented method for comparisons, denoted as maximal gradient method. Figure 13 shows the absolute errors between the radius of the reconstruction marker and the radius of the real spherical marker by performing the four contour-detecting methods on the data sets containing the vertical and horizontal ghost markers. The absolute error acquired by Laplacian of Gauss filter is higher than the other methods. The absolute error curve of the Canny filter is almost overlapping to the error curve of the maximal gradient method that the contours detected by these two methods are very close to each other, as shown in Fig. 13(b). The error curve acquired by the presented method indicates that this method can obtain a better accuracy than the other methods. In other words, the presented method is superior to the other methods both in the computing time and the contour-detecting accuracy. Thus, the MSRC method involved the presented contour-detecting approach can search the correct correspondences accurately and reconstruct the radius of the spherical marker within an error of 0.4 mm.

V. CONCLUSIONS

In this paper, a robust stereo-match method for OTS is proposed. OTS has been used in image-guided therapies for more than two decades. However, ghost markers remain a problem in this system even when costly professional IR marker-based tracking systems, such as Polaris (Northern Digital, Inc., ON, Canada), are used because some markers are coplanar with two optical centers. To date, many methods, including ordering and geometrical constraint, have been established to address the phenomenon of ghost markers that combine epipolar constraint and other constraints. Although these methods eliminate ghost markers in some instances, most of them are incompatible with many tracking systems and often require manual intervention. To eliminate ghost markers, we propose a robust stereo-match method combining maximum silhouette reprojection consistency with epipolar constraint for the OTS. Given the aim of searching the correct correspondences, an objective function is established for the minimization of the 3D distance from the inverse projective rays of the marker contour to the 3D reprojection contour on the virtual sphere surface. Additionally, we combine the epipolar constraint to increase the efficiency of the proposed method and the accuracy of the reconstructed 3D point. The subsequent procedure of parameter optimization is implemented by Levenberg-Marquardt algorithm. To verify the availability of the MSRC method, we use synthetic and real data to implement the MSRC method. The experimental results demonstrate the validity and robustness of this method.

Overall, the proposed stereo-match method effectively eliminates the ghost markers. A drawback of the proposed

method is that it consumes more computation time than non-optimization methods. Nevertheless, performing the computation of each marker pair on GPU kernels by using a parallel strategy is feasible because the computation step for each marker pair is independent. Thus, the computational cost can be reduced to an acceptable level for real-time implementation.

REFERENCES

[1] S. Tauscher, J. Tokuda, G. Schreiber, T. Neff, N. Hata, and T. Ortmaier, "OpenIGTLink interface for state control and visualisation of a robot for image-guided therapy systems," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, no. 3, pp. 285–292, May 2015.

[2] S. Pflugi *et al.*, "A cost-effective surgical navigation solution for periacetabular osteotomy (PAO) surgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 2, pp. 271–280, Feb. 2016.

[3] X. Chen *et al.*, "Development of a surgical navigation system based on augmented reality using an optical see-through head-mounted display," *J. Biomed. Inform.*, vol. 55, pp. 124–131, Jun. 2015.

[4] S. Pflugi *et al.*, "Augmented marker tracking for peri-acetabular osteotomy surgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 2, pp. 291–304, Feb. 2018.

[5] V. R. Kini, S. S. Vedam, P. J. Keall, S. Patil, C. Chen, and R. Mohan, "Patient training in respiratory-gated radiotherapy," *Med. Dosimetry*, vol. 28, no. 1, pp. 7–11, 2003.

[6] W. A. Tomé, S. L. Meeks, J. M. Buatti, F. J. Bova, W. A. Friedman, and Z. Li, "A high-precision system for conformal intracranial radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 47, no. 4, pp. 1137–1143, Jul. 2000.

[7] Y. Lyatskaya *et al.*, "Infrared-guided patient setup for lung cancer patients," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 71, no. 4, pp. 1124–1133, Jul. 2008.

[8] Y. Sun *et al.*, "Validation of anatomical landmarks-based registration for image-guided surgery: An *in-vitro* study," *J. Cranio-Maxillofacial Surg.*, vol. 41, no. 6, pp. 522–526, Sep. 2013.

[9] R. R. Shamir and L. Joskowicz, "Geometrical analysis of registration errors in point-based rigid-body registration using invariants," *Med. Image Anal.*, vol. 15, no. 1, pp. 85–95, Feb. 2011.

[10] R. R. Shamir, L. Joskowicz, S. Spektor, and Y. Shoshan, "Localization and registration accuracy in image guided neurosurgery: A clinical study," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 4, no. 1, pp. 45–52, Jan. 2009.

[11] T. H. Wagner *et al.*, "Optical tracking technology in stereotactic radiation therapy," *Med. Dosimetry*, vol. 32, no. 2, pp. 111–120, 2007.

[12] J. Burgner *et al.*, "A study on the theoretical and practical accuracy of conoscopic holography-based surface measurements: Toward image registration in minimally invasive surgery," *Int. J. Med. Robot Comput. Assist. Surg.*, vol. 9, no. 2, pp. 190–203, Jun. 2012.

[13] Q. Lin, R. Yang, K. Cai, P. Guan, W. Xiao, and X. Wu, "Strategy for accurate liver intervention by an optical tracking system," *Biomed. Opt. Express*, vol. 6, no. 9, pp. 3287–3302, Sep. 2015.

[14] S. L. Meeks *et al.*, "Optically guided patient positioning techniques," *Semin. Radiat. Oncol.*, vol. 15, no. 3, pp. 192–201, Jul. 2005.

[15] Q. Lin, K. Cai, R. Yang, H. Chen, Z. Wang, and J. Zhou, "Development and validation of a near-infrared optical system for tracking surgical instruments," *J. Med. Syst.*, vol. 40, no. 4, pp. 107–120, Apr. 2016.

[16] G. Soete *et al.*, "Initial clinical experience with infrared-reflecting skin markers in the positioning of patients treated by conformal radiotherapy for prostate cancer," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 52, no. 3, pp. 694–698, Mar. 2002.

[17] H. Yan *et al.*, "Investigation of the location effect of external markers in respiratory-gated radiotherapy," *J. Appl. Clin. Med. Phys.*, vol. 9, no. 2, pp. 57–68, Apr. 2008.

[18] N. Linthout, D. Verellen, K. Tournel, and G. Storme, "Six dimensional analysis with daily stereoscopic X-ray imaging of intrafraction patient motion in head and neck treatments using five points fixation masks," *Med. Phys.*, vol. 33, no. 2, pp. 504–513, Jan. 2006.

[19] G. Yan, J. Li, Y. Huang, K. Mittauer, B. Lu, and C. Liu, "Ghost marker detection and elimination in marker-based optical tracking systems for real-time tracking in stereotactic body radiotherapy," *Med. Phys.*, vol. 41, no. 10, pp. 101713-1–101713-10, Nov. 2014.

[20] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 650–656, Apr. 2006.

[21] Y. S. Heo, K. M. Lee, and S. U. Lee, "Robust stereo matching using adaptive normalized cross-correlation," *IEEE Trans. Pattern Anal.*, vol. 33, no. 4, pp. 807–822, Apr. 2010.

[22] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Jun. 2004.

[24] R. van Liere and J. D. Mulder, "Optical tracking using projective invariant marker pattern properties," in *Proc. IEEE Conf. Virtual Reality*, Los Angeles, CA, USA, Mar. 2003, pp. 191–198.

[25] F. Steinicke, C. P. Jansen, K. H. Hinrichs, J. Vahrenhold, and B. Schwald, "Generating optimized marker-based rigid bodies for optical tracking systems," in *Proc. 2nd Int. Conf. Comput. Vis. Theory Appl.*, Barcelona, Spain, 2007, pp. 387–395.

[26] A. Mossel, "Robust 3D position estimation in wide and unconstrained indoor environments," *Sensors*, vol. 15, no. 12, pp. 31482–31524, Dec. 2015.

[27] G. Stroian, T. Falco, and J. P. Seuntjens, "Elimination of ghost markers during dual sensor-based infrared tracking of multiple individual reflective markers," *Med. Phys.*, vol. 31, no. 7, pp. 2008–2019, Jul. 2004.

[28] M. Loaiza, A. Raposo, and M. Gattass, "A novel optical tracking algorithm for point-based projective invariant marker patterns," in *Proc. 3rd Int. Conf. Adv. Vis. (ISVC)*. Berlin, Germany: Springer, 2007, pp. 160–169.

[29] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. London, U.K.: Cambridge Univ. Press, 2003, pp. 20–84.

[30] R. I. Hartley and P. Sturm, "Triangulation," *Comput. Vis. Image Und.*, vol. 68, no. 2, pp. 146–157, Nov. 1997.

[31] M. Ribo, A. Pinz, and A. L. Fuhrmann, "A new optical tracking system for virtual and augmented reality applications," in *Proc. 18th IEEE Conf. Instrum. Meas. Technol.*, Budapest, Hungary, May 2001, pp. 1932–1936.

[32] F. Zhou, Y. Cui, B. Peng, and Y. Wang, "A novel optimization method of camera parameters used for vision measurement," *Opt. Laser Technol.*, vol. 44, no. 6, pp. 1840–1849, Sep. 2012.

[33] Y. Cui, F. Zhou, Y. Wang, L. Liu, and H. Gao, "Precise calibration of binocular vision system used for vision measurement," *Opt. Express*, vol. 22, no. 8, pp. 9134–9149, Apr. 2014.

[34] Q. Lin, R. Yang, K. Cai, X. Si, X. Chen, and X. Wu, "Real-time automatic registration in optical surgical navigation," *Infr. Phys. Technol.*, vol. 76, pp. 375–385, May 2016.

QINYONG LIN received the B.S. degree in biomedical engineering and the Ph.D. degree in biomedical engineering from the South China University of Technology, Guangzhou, China, in 2010 and 2016, respectively. He is currently a Post-Doctoral Researcher with the South China University of Technology.

His main research interests include computer visions, surgical robot, and image-guided navigation.

RONGQIAN YANG received the B.S. degree in electronic instrumentation and measurement from Nanchang Hangkong University, Nanchang, China, in 2001, the M.S. degree in communication and information systems from Jinan University, Guangzhou, China, in 2005, and the Ph.D. degree in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009.

He is currently an Associate Professor with the Department of Biomedical Engineering, South China University of Technology, Guangzhou, and a Visiting Associate Professor from the Department of Therapeutic Radiology, Yale University, New Haven, CT, USA. His main research areas include machine vision and biomedical instrument.

ZHESI ZHANG received the B.S. degree in biomedical engineering from South Central University for Nationalities, Wuhan, China, in 2015, and the M.S. degree in biomedical engineering from the South China University of Technology, Guangzhou, China, in 2018.

His research interests include robotics, calibration, and machine vision.

KEN CAI received the B.S. degree in measurement and control technology and instrumentation and the M.S. degree measurement and control technology and instrumentation from the Guangdong University of Technology, Guangzhou, China, in 2003 and 2006, respectively, and the Ph.D. degree in biomedical engineering from the South China University of Technology, Guangzhou, in 2011.

He is currently an Associate Professor with the School of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou. His main research areas include machine vision and biomedical instrument.

ZHIGANG WANG received the B.S. degree in biomedical engineering from Nanchang Hangkong University, Nanchang, China, in 2009, and the M.S. degree in biomedical engineering from the South China University of Technology, Guangzhou, China, in 2012. He has been with Guangzhou Aimooe Technology Co., Ltd.

His main research interests include surgical navigation systems and surgery robot.

MEIPING HUANG received the B.S. degree in medical radiography from Sun Yat-Sen University, Faculty of Medical Sciences, Guangzhou, China, in 1985, and the M.S. degree in imaging and nuclear medicine from Jinan University, Guangzhou, in 1999. She is currently the Professor of medicine with the Department of Catheterization Lab, Guangdong General Hospital, Guangzhou.

Her main research interests include cardiovascular imaging based on multimodal imaging diagnosis, surgery navigation in heart disease, and 3-D printing technology in medicine.

JINHUA HUANG received the B.S. degree in clinical medicine from the Jiangxi Medical College, Nanchang, China, in 1985, and the M.S. degree in interventional oncology from Shanghai Second Medical University, Shanghai, China, in 1992. After that, he was with MD Anderson, Houston, USA, in 2004, as an Academic Visitor.

He is currently the Professor of the Department of Minimally Invasive Interventional Therapy, Sun Yat-sen University Cancer Center, Guangzhou, China. Among his research about the basic and clinical study of interventional oncology, he concentrates himself in ablation therapy for solid tumor, especially large hepatocellular carcinoma.

YINWEI ZHAN received the B.S. degree in mathematics and the M.S. degree in computational mathematics from Jilin University, Changchun, China, in 1986 and 1988, respectively, and the Ph.D. degree in computational mathematics from the Dalian University of Technology, Dalian, China, in 1992. He held a post-doctoral position at Beijing Normal University from 1992 to 1994. Then, he joined Shantou University as an Associate Professor. From 2001 to 2004, he was a Post-Doctoral Researcher with CWI and Groningen University, The Netherlands.

Since 2015, he has been a Professor with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China. His main research interests include computer vision, HCI, machine learning, and VR/AR applications.

JIAN ZHUANG received the B.S. degree from the Department of Medicine, Southern Medical University, Guangzhou, China, in 1984, and the M.S. and Ph.D. degrees from the Department of Cardiac Surgery, Guangdong Cardiovascular Institute, Guangzhou, in 1989 and 2000, respectively. He is currently a Professor of medicine with the Department of Cardiac Surgery, Guangdong General Hospital, Guangzhou.

His main research interests include surgery navigation in heart disease and 3-D printing technology in medicine.