3D Tracking and Positioning of Surgical Instruments in Virtual Surgery Simulation

Zhaoliang Duan
School of Computer, Wuhan University, Wuhan 430072, China
Email: dzlwhu@gmail.com

Zhiyong Yuan, Xiangyun Liao, Weixin Si, Jianhui Zhao
School of Computer, Wuhan University, Wuhan 430072, China
Email: zhiyongyuan@whu.edu.cn, xyunliao@gmail.com, wxsics@gmail.com, jianhuizhao@whu.edu.cn

Abstract—3D tracking and positioning of surgical instruments is an indispensable part of virtual Surgery training system, because it is the unique interface for trainee to communicate with virtual environment. A suit of 3D tracking and positioning of surgical instruments based on stereoscopic vision is proposed. It can capture spatial movements of simulated surgical instrument in real time, and provide 6 degree of freedom information with the absolute error of less than 1 mm. The experimental results show that the 3D tracking and positioning of surgical instruments is highly accurate, easily operated, and inexpensive. Combining with force sensor and embedded acquisition device, this 3D tracking and positioning method can be used as a measurement platform of physical parameters to realize the measurement of soft tissue parameters.

Index Terms—Virtual surgery simulation system, Stereoscopic vision, Surgical instruments, 3D tracking and positioning

I. INTRODUCTION

With advances in robotics and information technology, computer graphics (CG) and virtual reality (VR) have been increasingly applied to the field of medicine [1]. As the cutting-edge interdisciplinary research field of information and medical sciences, research on virtual surgery simulation system has significant application value for reducing surgery risks, cutting training cost and protecting human health [2]. With the help of virtual surgery training platform, trainee surgeons can skillfully master the operations of surgical instruments, general procedure of surgery and anatomy of diseased region or organ.

The accurate displacement and force response of tissue model is a key part of VR based surgery training simulation system [3]. To meet this requirement, we must model the non-linear heterogeneous nature of soft tissue. The interaction of surgical instrument and virtual scene decides the accuracy and effectiveness of displacement measurement and force response.

In virtual surgery simulation, the interaction of surgical instrument and virtual scene mainly contains construction of surgical instruction, rendering of collision detection and rendering of the interaction and simulation [4]. In order to simulate the interaction between surgical instrument and virtual organ tissue vividly in virtual surgery simulation, the surgical instrument must be tracked and located accurately in real time.

Currently, there have been some available three dimensional (3D) trackers in the field of virtual reality. According to their physical properties, they are roughly classified into five subcategories: mechanical tracker [5], magnetic tracker [6-8], ultrasonic tracker [9], optical tracker [10-11] and hybrid tracker [12]. Some of them can provide high positioning accuracy, such as [6] and [10], and have been used in some medical applications. However, these existing devices are very expensive, therefore can only be popularized in a limited number of medical centers and research institutes. A 3D surgical instrument tracking and positioning method with a high performance-price ratio has been highly desirable for computer-based virtual surgery simulation systems [13].

In order to make this goal come true, we present a method based on stereoscopic vision for 3D tracking and positioning of surgical instruments based on stereoscopic vision. This method employs three cameras to capture the motion images of simulated surgical instrument in real time. After a series of computer processing, including camera calibration, reconstruction of 3D coordinates of markers on simulated surgical instruments and so on, we can obtain the six degree of freedom information of simulated surgical instruments, thereby positioning the instrument. At the end of this paper, we apply the presented method to accomplish interactive virtual organ tissue deformation simulation. The experimental results show that it is feasible and effective in virtual surgery simulation systems.

The rest of the paper is organized as follows. Section 2 describes the methodology for tracking and positioning...
surgical instruments. Section 3 gives the implementation of the presented method in details. Section 4 provides experimental results, and then section 5 concludes this paper.

II. METHODOLOGY FOR 3D TRACKING AND POSITIONING

A. Binocular Vision System

As an ideal linear camera model, pinhole model is the basic imaging model in computer vision. Fig. 1 illustrates the abstract graph of pinhole model. Where plane \( C \) is the imaging plane, point \( O \) is the camera optical center in plane \( C \). The coordinate system \( O_{1}−UV \) in plane \( C \) is the image coordinate system. Point \( O_{1} \) is the projection of point \( O \) in plane \( C \). Axis \( X_{1} \) and axis \( Y_{1} \) are respectively parallel with axis \( u \) and axis \( v \) in image coordinate system. Axis \( Z_{1} \) is the camera optical axis, perpendicular to plane \( C \). Camera coordinate system consists of point \( O \) and axis \( X_{1}, Y_{1}, Z_{1} \). \( OO_{1} \) is the camera focal length \( f \). To describe the position of the camera and surroundings, we adopt a reference coordinate system \( O_{w}−X_{w}Y_{w}Z_{w} \), which is called world coordinate system.

![Figure 1. Pinhole model](image1)

For a point \( P \) in world coordinate system, we can obtain its approximate imaging position \( \hat{p} \) in an image, that is the intersection point of \( PO \) and plane \( C \).

Binocular vision system is the simplest stereoscopic vision system consisting of two cameras. As shown in Fig. 2, \( P(x_{w},y_{w},z_{w}) \) is a point in world coordinate system, its two image points in image planes of two camera are \( p_{l}(u_{l},v_{l}) \) and \( p_{r}(u_{r},v_{r}) \), naming \( p_{l}(u_{l},v_{l}) \) and \( p_{r}(u_{r},v_{r}) \) “conjugate point”. The extension line of photo center \( O_{l} \) and point \( p_{l} \) intersect with the extension line of photo center \( O_{r} \) and point \( p_{r} \), this intersection point \( P \) is. The 3D coordinates of point \( P \) in world coordinate system can be obtained by the internal and external parameters of two cameras.

Generally, binocular vision system consists of five modules: image acquisition, camera calibration, feature extraction, stereo matching and 3D reconstruction [14].

B. Basic Principle

Our presented method is to utilize cameras to recover 3D coordinates of two markers on the simulated surgical instruments. Aiming at this goal, each marker must be covered by at least two cameras. If we use two cameras to track simulated surgical instrument, we can detect four feature points, the corresponding image regions of markers within the motion image of simulated surgical instrument, at a time, then we classify these four feature points into two pairs of identical points, and calculate their image coordinates respectively. Along with the camera parameters, 3D coordinates of two markers can be obtained through least square method [15]. Fig. 3 shows the flowchart of the presented method.

![Figure 2. Binocular vision system](image2)

Considering the virtual surgery simulation system is extremely strict with precision, we employ three cameras to capture the movement of simulated surgical instrument...
for the purpose of minimizing the system error introduced by image acquisition. Three cameras construct three pairs of cameras groups, and each camera group includes two cameras. As for each marker, three pairs of cameras groups obtain three groups of 3D coordinates. By calculating their average values, we get the final and more accurate 3D coordinates of markers.

C. System Construction

The 3D tracking and positioning apparatus of surgical instrument based on our proposed method consists of a simulated surgical instrument, three cameras and a computer. Fig. 4 illustrates the hardware distribution of our developed 3D tracking and positioning apparatus.

The gray circular area is active region of simulated surgical instrument. The degree of included angle constructed by any two cameras and the center of gray circular area is 120.

The cameras are Basler ac A1300-30gc cameras with frame rate 30 Frames/sec and the highest resolution 1092*962. The cameras are connected to the computers. The main parameters of Basler acA1300-30gc camera are shown in Table I.

<table>
<thead>
<tr>
<th>TABLE I. THE MAIN PARAMETERS OF BASLER ACA1300-30GC CAMERA</th>
</tr>
</thead>
<tbody>
<tr>
<td>The highest resolution</td>
</tr>
<tr>
<td>Optical Size</td>
</tr>
<tr>
<td>Pixel size (micrometer)</td>
</tr>
<tr>
<td>Sensor type</td>
</tr>
<tr>
<td>Frame rate</td>
</tr>
<tr>
<td>Data transmission</td>
</tr>
</tbody>
</table>

Two markers with 70 mm space distance are deployed on the simulated surgical instrument, and specific distribution of markers has been shown in Fig. 5. As for the actual simulated surgical instrument, its main body is white, while two markers are black.

III. IMPLEMENTATION

A. Camera Calibration

Camera calibration is the most basic step in stereoscopic vision [16]. Its purpose is to obtain camera parameters, i.e., the mapping relation between the 3D coordinates of point \( P(x_w, y_w, z_w) \) in word coordinate system and 2D coordinates point \( p(x, y) \), the projection of point \( P \) on image plane of camera, in image coordinate system. The mapping relation can be described by the equation 1 simply:

\[
\begin{bmatrix}
    x_w \\
    y_w \\
    z_w
\end{bmatrix}
= \begin{bmatrix}
    m_{11} & m_{12} & m_{13} & m_{14} \\
    m_{21} & m_{22} & m_{23} & m_{24} \\
    m_{31} & m_{32} & m_{33} & m_{34}
\end{bmatrix}
\begin{bmatrix}
    x \\
    y \\
    1
\end{bmatrix}
= M
\begin{bmatrix}
    x \\
    y \\
    1
\end{bmatrix}
\tag{1}
\]

Where \( z_w \) is \( Z \) coordinate of point \( P \) in camera coordinate system, \( M \) is so called projection matrix determined by camera parameters. Once we obtain camera parameters, we do not need to calculate them again until camera is moved. Generally speaking, the detailed calibration process includes the following five steps.

- Step 1. Generation of planar calibration plate.
- We adopt a regular 7×7 black-and-white checkerboard as the pattern on the calibration plate, as shown in Fig. 6, the size of each checker is 30×30 mm.
• Step 3. Corner detection
In this paper, we employ the function, cvFindChessboardCorners(), from OpenCV library to detect corners of calibration plate image [17]. For an input image \( I(u,v) \)'s pixel \((x,y)\), if \( R \) is bigger than a given threshold, then pixel \((x,y)\) is a Corner.

\[
R = \det M - k(\text{trace}M)^2
\]

\[
= \begin{vmatrix} I_x \end{vmatrix}^2 + \begin{vmatrix} I_y \end{vmatrix}^2 - k(\begin{vmatrix} I_x \end{vmatrix} + \begin{vmatrix} I_y \end{vmatrix})^2
\]

Where \( k \) is a coefficient, generally is 0.04. \( I_x \) and \( I_y \) are first gray gradients:

\[
I_x = \frac{\partial I}{\partial x}, \quad I_y = \frac{\partial I}{\partial y}
\]

\( M \) is a real symmetric matrix:

\[
M = \begin{bmatrix} I_x^2 & I_x I_y & I_y^2 \\ I_x I_y & I_y^2 & I_y^2 \\ I_x I_y & I_y^2 & I_x^2 \end{bmatrix}
\]

\( \det M \) is \( M \)'s determinant, \( \text{trace}M \) is the trace of \( M \).

\( \begin{vmatrix} \cdot \end{vmatrix} \) is Gauss smoothing operator. After that, cvFindCornerSubPix() is used to get more accurate image coordinates of corners. Fig. 8 shows the corner detection results of calibration plate images in Fig. 6.

• Step 4. Corner matching
Actually, the distributions of corners in different rectangular arrays are not one-to-one corresponding, so corner matching is indispensable. In our paper, we design two basic transformation functions performing on the rectangular corner array: clockwise rotation function and horizontal flip function. The different combination of these two functions can achieve all transformation of the rectangular array needed in experiment.

• Step 5. Calculation of camera parameters
When it comes to calculation of camera parameters, the two relatively popular algorithms are Tsai two-step method [18] and Zhang’s algorithm [19], and they are both highly accurate and robust. Compared with Tsai two-step method, Zhang’s algorithm expects camera is supposed to capture calibration plate from different viewpoints, but the camera and calibration plate should be fixed all the time and cannot be moved in our application. In this situation, we choose Tsai two-step method to calculate the camera parameters.

The calibration results of the camera capturing Fig. 7(a) are shown in TABLE II.

<table>
<thead>
<tr>
<th>TABLE II. CALIBRATION RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foal length ( f ) (mm)</td>
</tr>
<tr>
<td>Radial distortion coefficient kappa(1/m²)</td>
</tr>
<tr>
<td>Translation vector ( T ) (mm)</td>
</tr>
<tr>
<td>Scale factor ( s )</td>
</tr>
<tr>
<td>Optical center coordinates</td>
</tr>
</tbody>
</table>
| Rotation matrix \( R \) | \[
\begin{bmatrix}
0.629319 & 0.328383 & 0.777092 \\
0.417971 & -0.847033 & 0.328383 \\
-0.655178 & 0.536929 & 0.531459
\end{bmatrix}
\]

B. Reconstruction of 3D coordinates of markers
In order to recover 3D coordinates of two markers on simulated surgical instrument, we first need to extract corresponding feature points, and then match feature points from three cameras to form identical points. After these, least square method is used to calculate the 3D coordinates of markers.

• Picture Preprocessing

Owing to effect of illumination, it is unavoidable to generate shadows of simulated surgical instrument and hands of operator in motion images of simulated surgical instrument. Besides, the noises are also inevitable to be introduced during camera imaging. Therefore, image preprocessing is quite necessary.

For any motion image of simulated surgical instrument, this process mainly includes the following steps:

1. Step 1. Convert the RGB image into gray image, the conversion formula is shown as following:

\[
\text{Gray} = \begin{bmatrix} R \\ G \\ B \end{bmatrix} \begin{bmatrix} 0.299 \\ 0.578 \\ 0.114 \end{bmatrix}
\]

2. Step 2. Subtract the gray image of corresponding background image to remove background and it can be expressed by the following formula:

\[
I_{\text{new}} = \max(0, I_{\text{bg}} - I_{\text{old}})
\]
Where \( I_{\text{old}} \) is a captured gray image, \( I_{\text{bkg}} \) is background image and \( I_{\text{new}} \) is a processed image according to formula (6).

(3) Step 3. Apply median filter to the residual image for noise removal and it can be expressed by the following formula:

\[
I(x, y) = \text{median} \{ I(i, j) \}_{(i,j) \in W_y}
\]

Where \( I \) is an input image, \( \hat{I} \) is an image applied median filter. As we seen in formula (7), to calculate the result of a pixel applied median filter, we must sort the pixels in \( W_y \), and we choose the intermediate value as the result.

(4) Step 4. Utilize threshold segmentation method based on statistics to complete initial shadow removal. The threshold value is set as 101 through statistics. Fig. 9 shows the result of image preprocessing. Although most of background and shadows has been removed, there are still redundant areas.

- Feature points extraction and stereo matching
  
  The purpose of feature point extraction is to further remove redundant areas, and finally obtain two dimensional (2D) image coordinates of marker by calculating the barycentre of corresponding feature point. Generally speaking, feature points in an image obey Gaussian distribution, so their grey histograms will appear Gaussian peaks [20]. The sum of each row is shown in Fig. 10. Assume the top left corner of an image is origin, the width direction is X axis, the height direction is Y axis. Then we have integral projection along Y axis that is adding all pixels’ gray value in a row, the formula is:

\[
y(j) = \sum_{i=0}^{w} I(i, j)
\]

Where \( j \) is the \( j \)th row of a image, \( w \) is the width of a image.

The integral projection of Fig. 9.b on Y axis forms a blue curve. Two highest peaks represent the projections of two feature points on Y axis, others areas correspond to the projections of redundant areas. If we set a proper threshold value, it is quite easy to isolate projections of feature points on Y axis. In the same way, we also can get their projections on X axis. Through experiments, we find that the threshold value is 2000. In this way, we can determine the feature points. At last, we take the coordinates of barycentre of each feature point as its 2D image coordinates. The extraction results of the feature points in Fig. 9(b) are shown as the crosses in Fig. 11.

- Calculation of 3D coordinates of markers
  
  Suppose the feature points of marker 1, \( Q(x_1, y_1, z_1) \), are \( q_1(x_1, y_1) \), \( q_2(x_2, y_2) \) and \( q_3(x_3, y_3) \) in images taken by three cameras, \( C_1 \), \( C_2 \), \( C_3 \), respectively. After camera calibration, we have known corresponding three projection metrics \( M^1 \), \( M^2 \) and \( M^3 \). We can obtain the following equation.

\[
\begin{bmatrix}
  x_1 \\
  y_1 \\
  1
\end{bmatrix}
= \begin{bmatrix}
m_{11} & m_{12} & m_{13} & m_{14} \\
m_{21} & m_{22} & m_{23} & m_{24} \\
m_{31} & m_{32} & m_{33} & m_{34}
\end{bmatrix}
\begin{bmatrix}
x_w \\
y_w \\
1
\end{bmatrix}
= M_{f,1}
\begin{bmatrix}
x_w \\
y_w \\
1
\end{bmatrix}
\]

(9)

During the actual surgery, marker 1, as shown in Fig. 11, always moves above marker 2. As for any image, the feature point with larger Y coordinate therefore is image region of marker 1, and another one is image region of marker 2. In this way, the extracted feature points are simply matched.

Because of its importance, the process of marker extraction and stereo matching is shown in detail as follows.
As for camera group including \( C_1 \) and \( C_2 \), we set \( i = 1, 2 \), and let \( x_i, y_i, z_i \) take the places of \( x_x, y_x, z_x \), respectively. Thus, we get linear system as follows.

\[
\begin{align*}
(x_{1m} - m_1)x + (x_{2m} - m_1)y + (x_{3m} - m_1)z &= m_1 - x_m \\
(y_{1m} - m_1)x + (y_{2m} - m_1)y + (y_{3m} - m_1)z &= m_1 - y_m \\
(z_{1m} - m_1)x + (z_{2m} - m_1)y + (z_{3m} - m_1)z &= m_1 - z_m
\end{align*}
\] (10)

The linear system above includes four equations and three unknowns \( x_w, y_w, z_w \); as for camera group including \( C_1 \) and \( C_3 \), we get \( x_{13}, y_{13}, z_{13} \) and \( x_{23}, y_{23}, z_{23} \). At last, the final 3D coordinates of maker 1 can be calculated as follows.

\[
\begin{align*}
x_w &= \frac{x_{12} + x_{13} + x_{23}}{3} \\
y_w &= \frac{y_{12} + y_{13} + y_{23}}{3} \\
z_w &= \frac{z_{12} + z_{13} + z_{23}}{3}
\end{align*}
\] (11, 12, 13)

In the same way, we also can obtain the 3D coordinates of maker 2. Thus, we complete the 3D tracking and positioning of simulated surgical instrument.

IV. EXPERIMENTAL RESULTS

The detailed configuration of our experimental platform is as follows: Computer: Intel Core Duo CPU @2.66GHz, 2GB memory; Basler acA1300-30gc Camera: 1092 × 962 resolution, 30FPS; Software: Microsoft Visual C++.net 2005.

The 3D tracking and positioning equipment is shown in Fig. 12. Three cameras are connected to the computers and transmit the captured pictures to computers. Then we process the captured pictures and calculate marked points' 3D coordinates.

![Figure 12. 3D tracking and positioning equipment](image)

The absolute error between the corners' calculated 3D coordinates and the theoretical 3D coordinates is shown in Fig. 13.

![Figure 13. Interactive organ tissue deformation simulation](image)

According to Fig. 13, the average absolute error of reconstructed corners’ 3D coordinates is 33.40 ÷ (36 × 3) = 0.31mm. The error mainly derived from the following three aspects:

- Production and placing of the calibration plate
- The error of corner detection algorithm
- The error of camera calibration algorithm

As the absolute error of reconstructed corners is less than 1mm, it can meet the required precision of the applied field.

The image coordinates and space coordinate of surgical instrument are shown in TABLE III.

<table>
<thead>
<tr>
<th>Image coordinate</th>
<th>Space coordinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>510.01, 226.23</td>
<td>285.76, 407.24</td>
</tr>
<tr>
<td>477.09, 380.54</td>
<td>346.31, 595.17</td>
</tr>
<tr>
<td>226.67, 663.15</td>
<td>258.64, 336.43</td>
</tr>
<tr>
<td>625.61, 384.06</td>
<td>317.18, 508.61</td>
</tr>
<tr>
<td>674.41, 170.26</td>
<td>259.24, 262.90</td>
</tr>
<tr>
<td>634.01, 333.78</td>
<td>321.10, 438.32</td>
</tr>
<tr>
<td>456.09, 104.61</td>
<td>273.32, 277.64</td>
</tr>
<tr>
<td>401.13, 258.51</td>
<td>350.13, 488.40</td>
</tr>
<tr>
<td>429.32, 190.17</td>
<td>556.72, 239.15</td>
</tr>
<tr>
<td>636.14, 366.27</td>
<td>621.82, 246.19</td>
</tr>
<tr>
<td>533.67, 261.04</td>
<td>524.30, 268.42</td>
</tr>
<tr>
<td>507.17, 438.30</td>
<td>580.19, 447.20</td>
</tr>
</tbody>
</table>

As we mentioned above, the actual distance between the two makers on the simulated surgical instrument is fixed and it measures 70 mm. Fig. 14 provides the
distances calculated by the 3D coordinates of the two makers. The Fig. 15 provides the absolute error and standard deviation calculated by the 3D coordinates of the two makers.

![Figure 14. Standard distance and actual distance](image1)

![Figure 15. Absolute error and standard deviation](image2)

As can be seen, the standard deviation of our developed apparatus is 0.371632mm, less than 1mm, which totally satisfies the precision requirement of current virtual surgery simulation systems.

There mainly exist three types of errors in our developed apparatus. The first one is the algorithm error which comes from the implementation of algorithms. The second error is the generation and placement of planar calibration plate and the third one is the generation of simulated surgical instrument manually. Overall, our developed apparatus is precise, and the major reason for the errors is the fabrication errors, which can be decreased by using professorial calibration plate and machined simulated surgical instrument.

V. CONCLUSION AND FUTURE WORK

In this paper, we present a method based on stereoscopic vision to construct a suit of 3D tracking and positioning apparatus of surgical instruments. It consists of a simulated surgical instrument, three cameras and a computer. Three cameras are used to capture the motion images of simulated surgical instrument in real time. After a series of computer processing, we can obtain the six degree of freedom information of simulated surgical instruments with the absolute error of less than 1 mm, thereby positioning the instrument. Then, we analyze the sources of error, and integrate the developed apparatus into soft tissue deformation simulation in virtual surgery. The experimental results show that the proposed method is highly accurate and easily operated, even if it is inexpensive. In future work, we will utilize this tracking and positioning method and equipment to capture the soft tissue warping images and calculate the soft tissue parameters by using least square method. Also we will perfect our tracing and positioning method and construct parameter measurement platform, proving the effectiveness of it.

ACKNOWLEDGMENTS

This work was fully supported by a grant from the National Natural Science Foundation of China (Grant No. 61070079).

REFERENCES


Weixin Si is currently a fresh-year Master’s student in Computer Science from Wuhan University in China. He has received his BS degree in computer science and technology from Wuhan University in China in June 2011. Si is recommended for admission to Wuhan University to continue his study in the next 2 years. His research interests include computer simulation and virtual reality, image processing, and pattern recognition, etc. He has published 9 papers in related conference proceedings and journals.

Jiahui Zhao received the B.Sc. degree in Computer Engineering from Huazhong University of Science and Technology in 1997, the M.Sc. degree in Computer Science from Huazhong University of Science and Technology in 2000, and the Ph.D. degree in Computer Science from Nanyang Technological University in 2004. From 2003 to 2006, he worked as a Research Assistant/Associate in Hong Kong University of Science and Technology. Currently, he is working as an Associate Professor in Computer School of Wuhan University. His research interests include computer graphics and digital image processing. He has published over 50 papers in related conference proceedings and journals.

Zhaoliang Duan is currently a fresh-year Master’s student in Computer Science from Wuhan University in China. He has received his BS degree in computer science and technology from Wuhan University in June 2011. He is recommended for admission to Wuhan University to continue his study in the next 2 years. His research interests include computer simulation and virtual reality. He has published 10 papers in related conference proceedings and journals.

Weixin Si is currently a fresh-year Master’s student in Computer Science from Wuhan University in China. He has received his BS degree in computer science and technology from Wuhan University in China in June 2011. Si is recommended for admission to Wuhan University to continue his study in the next 2 years. His research interests include computer simulation, virtual reality, image processing, and pattern recognition, etc. He has published 7 papers in related conference proceedings and journals.

Jiahui Zhao received the B.Sc. degree in Computer Engineering from Huazhong University of Science and Technology in 1997, the M.Sc. degree in Computer Science from Huazhong University of Science and Technology in 2000, and the Ph.D. degree in Computer Science from Nanyang Technological University in 2004. From 2003 to 2006, he worked as a Research Assistant/Associate in Hong Kong University of Science and Technology. Currently, he is working as an Associate Professor in Computer School of Wuhan University. His research interests include computer graphics and digital image processing. He has published over 50 papers in related conference proceedings and journals.