OPTICAL TRACKING FOR MEDICAL APPLICATIONS

Z. Szabó

Faculty of Biomedical Engineering, Czech Technical University in Prague,
Kladno, nám. Sítná 3105, 272 01, Czech Republic

szabo@fbmi.cvut.cz

Abstract: Optical tracking system is described for medical application, where the used medical tool is equipped with special marks. The position and orientation of the tool is measured based on different views from the cameras. An example of 3D free-hand ultrasound is presented, where the multi-camera system is used for ultrasound probe localization. The resolution of two-camera system is evaluated for the specific geometric arrangement of the cameras as well as the simple marker recognition algorithm is described.

Introduction

Tracking of medical tool movements is one of the major tasks in number of medical application. There are different methods how this tracking may be performed. Common tracking systems use magnetic or ultrasonic trackers as well as mechanical devices. The drawbacks of these systems are their principles of work. Typically, the user has to be linked to a measurement instrument, either by cable or by a mechanical linkage, which is more restraining for the user. Furthermore, while mechanical tracking systems are extremely precise, magnetic and acoustic tracking systems suffer from different sources of distortions. For this reason, an optical tracking is an alternative solution which overcomes many of the drawbacks of conventional tracking systems.

Materials and Methods

This work is focused on optical tracking with cameras providing position determination for medical application in the case, when there is always a free line of sight between the markers and cameras. The optical tracker does not involve any magnetic field for determination of position data and consequently does not permit any deformation of these data in the presence of metallic structures. Their principle is based on the analysis of 2-dimensional projections of image features received by CCD cameras.

The 3D Free-hand ultrasound is one of the applications of optical tracking, where conventional 2D diagnostic ultrasound machines are extended with ultrasound probe localization.

As the clinician moves the probe, its position and orientation are recorded as well as the 2D ultrasound image slices [2]. The camera system localizes the tip of the ultrasound probe by tracking the motion of markers that are mounted on the body of the device. The 2D slices, together with the information about their positions and orientations, constitute an irregularly sampled 3D dataset describing the volume scanned by the clinician. Afterwards the geometry of complex three dimensional anatomies is visualized (Figure 1).

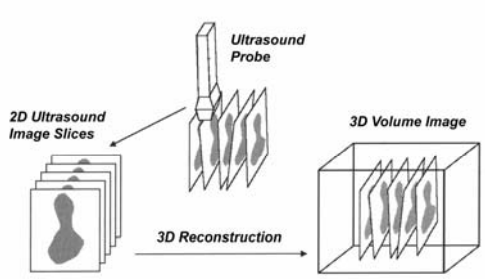

Figure 1: 3D Free-hand ultrasound image reconstruction

Before the camera system can be used to track markers (mounted on medical tool), the system first needs to be calibrated. This is done by placing a calibration frame in front of the cameras and by running the calibration software on the computer. Several methods for camera calibration are presented in the literature. We have used Matlab toolbox for performing calibration procedure based on direct linear transformation (DLT). This mathematical model was used in order to calculating intrinsic (focal length, location of the image center, effective pixel size, distortion coefficient of the lens) and extrinsic (rotation matrix, translation vector) camera parameters.

Camera calibration

The goal of camera calibration is to compute the components of some matrix $\boldsymbol{M}$ that represents how some known 3D points are transformed to 2D image plane.

If we have homogenous coordinates of n object points and we know their images under the projective transformation, which is represented by a matrix $\boldsymbol{M}$, we can write the following equation for all of these points

$$[a_{i1}, a_{i2}, a_{i3}, a_{i4}]\boldsymbol{M} = [b_{i1}, b_{i2}, 1] \qquad (1)$$

With merging these equations, we obtain a system of linear equations from which we can retrieve $\boldsymbol{M}$.

ISSN: 1727-1983

$$\begin{bmatrix} a_{11}, a_{12}, a_{13}, a_{14} \\ a_{21}, a_{22}, a_{23}, a_{24} \\ \vdots \ \vdots \ \vdots \ \vdots \\ a_{n1}, a_{n2}, a_{n3}, a_{n4} \end{bmatrix} \boldsymbol{M} = \begin{bmatrix} b_{11}, b_{12}, 1 \\ b_{21}, b_{22}, 1 \\ \vdots \ \vdots \ \vdots \\ b_{n1}, b_{n2}, 1 \end{bmatrix} \quad (2)$$

which we write

$$\boldsymbol{AM} = \boldsymbol{B} \quad (3)$$

We can retrieve $\boldsymbol{M}$ by multiplying both sides of the equation by left inverse (The matrix $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ is a left inverse of $\boldsymbol{A}$) so that

$$\boldsymbol{M} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{B} \quad (4)$$

If $n \geq 6$ (at least six corresponding points are known), this system is overdetermined and we can seek the best possible solution in the sense of least squares. The result gives us the matrix $\boldsymbol{M}$, called as projection matrix. The projection matrix $\boldsymbol{M}$ is composed from matrices $\boldsymbol{K}$, $\boldsymbol{R}$ and vector $\boldsymbol{t}$

$$\boldsymbol{M} = \left[\boldsymbol{KR} \middle| -\boldsymbol{KRt}\right] \quad (5)$$

where $\boldsymbol{K}$ is called the camera calibration matrix (the coefficients of this matrix are so called intrinsic parameters of the camera), $\boldsymbol{R}$ is called rotation matrix (it expresses three elementary rotations of the camera) and $\boldsymbol{t}$ is the translation vector giving three elements of the translation of the origin of the world-coordinate system with respect to the camera coordinate system.

Note: The delimiter | in equation (5) denotes that the matrix is composed of two submatrices.

The calibration matrix $\boldsymbol{K}$ is upper triangular containing five constants of the camera focal distance, principal point coordinates in the image, scaling and the degree of skew of the coordinate axes in the image plane [1].

The rotation matrix $\boldsymbol{R}$ is orthogonal and contains constants expressing rotations along the axes x, y and z also called as pan, tilt and roll, respectively.

The list of intrinsic parameters:

- f – Focal distance.
- $[U_{0x}, U_{0y}]$ – Principal point coordinates, sometimes called the center of the image in camera calibration procedures. It is the intersection of the optical axis with the image plane.
- θ – Skew coefficient: The skew coefficient defining the angle between the x and y pixel axes.
- k – Scaling: The image scaling coefficients (radial and tangential distortions).

The list of external parameters:

- Rotations of the cameras.
- Translations of the cameras

The spatial configuration of the two cameras and the calibration planes is displayed in a form of 3D plot:

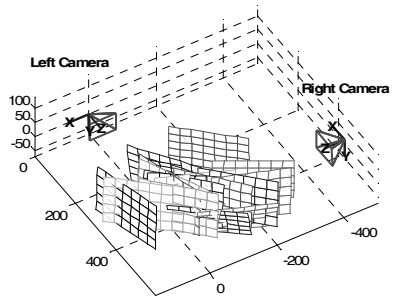

Figure 2: The spatial configuration of two cameras

Above we assumed ideal central projection, as a pinhole camera does, but in the case of real camera, the lens of the camera performs distortion. A typical lens performs distortion of several pixels which a human observer does not notice looking at a general scene. However, when the image is used for measurements, compensation for the distortion is necessary.

Let $\boldsymbol{P}$ be a point in 3D space having coordinates $[X_P; Y_P; Z_P]$ in the camera reference frame. The normalized coordinates of this point is than $[x_n; y_n] = [X_P/Z_P; Y_P/Z_P]$.

Considering lens distortion, the new normalized coordinates are defined as follows:

$$\begin{bmatrix} x_D \\ y_D \end{bmatrix} = \left(1 + k_1 r^2 + k_2 r^4 + k_5 r^6\right) \begin{bmatrix} x_n \\ y_n \end{bmatrix} + \boldsymbol{d}_t \quad (6)$$

where $r^2 = x^2 - y^2$ and d_t is the tangential distortion vector:

$$\boldsymbol{d}_t = \begin{bmatrix} 2k_3 xy + k_4\left(r^2 + 2x^2\right) \\ 2k_4 xy + k_3\left(r^2 + 2y^2\right) \end{bmatrix} \quad (7)$$

Vector $\boldsymbol{k} = [k_1, k_2, k_3, k_4, k_5]$ contains both radial and tangential distortion coefficients. The tangential distortion is due to imperfect centering of the lens components and other manufacturing defects in a compound lens. This distortion model was first introduced by Brown in 1966 [2].

For non wide-angle cameras, it is often not necessary to count with the radial component of distortion model beyond the 4th order (i.e. $k_5 = 0$). In distortion model used e.g. by Zhang [4] the last three distortion coefficients are set to zero. There is used 4th order distortion model with no tangential component.

Marker recognition and computation its position

The tested marker recognition algorithm consists of several steps. The first step is 2D cross correlation on the image using a pre-defined mask. Its output is a series of threshold points, which may be part of markers. The next step is to group these points and to

ISSN: 1727-1983

calculate the centroids of the markers. This gives a series of coordinate pares of possible markers.

There are two elementary ways of doing the cross correlation: time-domain cross correlation and frequency domain cross correlation. The first one requires a suitable mask, which design is crucial in the effectiveness of the marker recognition algorithm. When the mask is correlated with an image of a marker, the cross correlation value should generate a peak. The cross correlation is done on each pixel in the search area using the following equation:

$$R(h,k) = \sum_{y=0}^{y<l} \sum_{x=0}^{x<l} M(x,y) L(x+h, y+k) \qquad (8)$$

where *R(h,k) is* the correlation value of the image point with coordinates *(h,k)* inside the search area, *M* is the mask, *L* is the area of the image which must be correlated with the mask and *l* is the size of the mask. When we assume that multiplication and summation operations take the same processor clock cycles, then the computation complexity of this method is $2L^2M^2$ [7].

The second way of doing the cross correlation is the frequency domain cross correlation. A 2D FFT algorithm is first used to convert the search area to the frequency domain. This result is then multiplied by the 2D FFT of the mask. The 2D FFT of the mask is only done once, as it would not change. The product of the two FFTs is then converted back to the time-domain using a 2D IFFT algorithm. The computational complexity of the frequency domain cross correlation algorithm is $N^2(20\log_2 N+6)$, where *N* is the size of FFT ($N > L+M$) [7]. The advantage of the frequency domain method is that the number of calculations stays considerably the same for different sizes of search areas and mask sizes. The major drawback of this method is that it uses significantly more memory than the time domain method. Furthermore it is no as flexible as the time domain method.

When a cross correlation algorithm of a point has been calculated, it is necessary to see if that point is an overthresholded point. The *N* highest value overthreshold points are stored in a sorted list, which is steadily updated as a new correlation value is known. After the cross correlation process if finished, the overthresholded points must be processed. They are grouped into islands and each island is a possible marker. The algorithm considers points to belong to the same island if they touch each other. After an island of points has been isolated, centroid calculations must be made in that data to determine the position of the possible marker. The centroid of the island is calculated by calculating the centre of mass of the correlation values of the points. The following equations are used:

$$x_c = \frac{\sum_i x_i \sum_j R_{ij}}{W}, \quad y_c = \frac{\sum_j y_j R_{ij}}{W}, \quad W = \sum_{ij} R_{ij}, \qquad (9)$$

where *W* is the total weight of the island, (x_c,y_c) is the centroid coordinates of the island, R_{ij} is the correlation value of the thresholded point with coordinates (x_i,y_j). The results of the centroid calculations of each island are saved for further processing. An island with a heavy weight has a bigger chance of being a marker than an island with a light weight. After all the markers have been recognised and their coordinates are known, the markers should be classified. Classification depends on the used configuration of markers (depending on application) and is therefore not discussed here.

Camera system resolution

The resolution of the camera system is defined as the smallest change that can be detected by a sensor (in our case, by a CCD sensor inside the camera). Suppose two cameras in geometry as Figure 3 shows. When investigated points (markers) are not only on the plane perpendicular to the optical axes, but in 3D space, the derivation of the camera system resolution is a complex task. Also suppose next simplifications: the cameras are equally far from the investigated point; they are not turned around their optical axes, and the angle between them (between their optical axes) is γ.

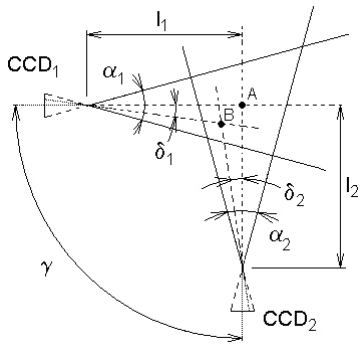

Figure 3: The configuration of cameras

When the investigated point is lying somewhere on the plane given by the two camera optical axes, then the maximum possible movement in the scene without being observed is *e*.

$$e = \sqrt{d_x^2 + d_y^2 + d_z^2 - 2d_x d_y \cos(180-\gamma)}\Big/2 \qquad (10)$$

for $\gamma \in (0; 90)$, where d_x, d_y, d_z are dimensions of 3D segment in the distance l_1 and l_2 from the cameras determining a space where the point can move without being observed. The half of the diagonal of this segment is defined as *e* [4].

An example how the error *e* is depending on the angle between cameras γ shows Figure 4. Note: the investigated point (point A) is lying on the intersection of optical axes. The angle γ is changing between 0° (when the camera optical axes are parallel) and 180° (when the cameras are looking against each other). Obviously, these two extremes will cause largest error (∞), where we cannot define the exact position of the point at all. The minimum error could be when the cameras are perpendicular to each other ($\gamma = 90°$) in the

ISSN: 1727-1983

point *A*. In other points this error will be naturally changed, because distances l_1 and l_2 will change.

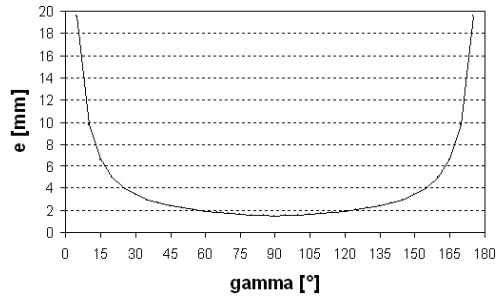

Figure 4: The function of *e*, when the distances from the cameras to the investigated point are fixed and only the angle between them γ is changing

Results and discussion

First experimental results are presented in this paper, where the resolution analysis of camera system due to finite number of sensing elements inside the cameras, focal distance and their location against the investigated point is shown. The possible error is the minimum when the camera positions are fixed and the angle between them is 90°.

The effect of distortions on the pixel image and the radial component versus the tangential component of distortion is visualized with arrows spread all in the image area, where each arrow represents the effective displacement of a pixel induced by the lens distortion. The cross indicates the center of the image, and the circle the location of the principal point.

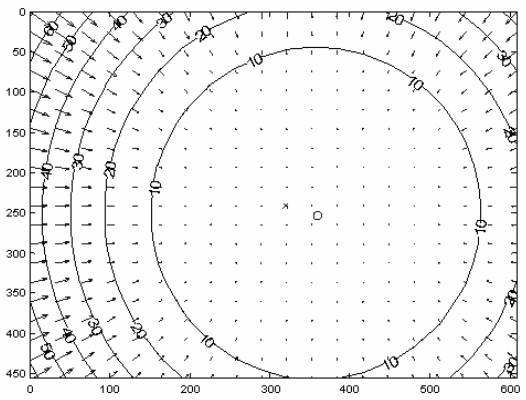

Figure 5: Distortion model of the camera

Two identical monochrome FireWire cameras were tested with the following features: 640 x 480 pixel resolution, up to 30 images/s, 1/4" CCD progressive scan. The parameters of lenses: H 2616 FICS – 3 Computar C/CS-Mount Standard Lenses with 2.6 mm focal length, 1.6 – 11 iris range, 128° horizontal angle of view. If both cameras are positioned 1m from the investigated point the possible error is 3 mm, but this value can be reduced by additional image processing.

The calibration results for both (left and right) cameras are summarized in the following table.

Table 1: Left and right camera calibration parameters

Intrinsic parameters of left camera	
Focal Length	2.6 mm
Principal point	[359.08 252.86]
Distortion coefficients	[-0.284 0.099 -0.001 -0.0006 0]
Intrinsic parameters of right camera	
Focal Length	2.6 mm
Principal point	[320.86 259.99]
Distortion coefficients	[-0.29 0.108 -0.0004 0.001 0]
Extrinsic parameters (position of right camera with respect to left camera)	
Rotation vector	[0.027 -1.585 -0.028]
Translation vector	[421.00 -6.26 512.75]

Acknowledgements

This work has been supported by the grant of Czech Science Foundation, no. 102/03/D030 and the grant of the Ministry of Education, Youth and Sports, Czech Republic, no. MSM6840770012.

References

[1] SONKA M., HLAVAC V., BOYLE R. (1999), 'Image Processing, Analysis, and Machine Vision', PWS Publishing, pp. 448-486

[2] NELSON T.R., DOWNEY D.B., PRETORIUS D.H., FENSTER A. (1999): 'Three-Dimensional Ultrasound', Lippincott Williams & Wilkins, Philadelphia, PA.

[3] BROWN D.C. (1966) 'Decentering Distortion of Lenses', *Photometric Engineering*, Vol. 32, No. 3, pp. 444-462

[4] SZABÓ Z., HUDEC R. (2004) 'Camera System Resolution Analysis for Ultrasound Probe Localization', MEDICON 2004, Mediterranean Conference on Medical and Biological Engineering, Ischia, Naples

[5] GENADE B. 'Marker recognition algorithm', Internet site address: http://mysite.mweb.co.za/residents/cyb00746/tracker/Marker_recognition.htm

ISSN: 1727-1983