Towards visual matching as a way of transferring pre-operative surgery planning

Stijn De Buck, Johan Van Cleynenbreugel,
Guy Marchal, and Paul Suetens

Faculties of Medicine and Engineering
Medical Image Computing (ESAT and Radiology)
University Hospital Gasthuisberg
Herestraat 49
B-3000 Leuven
`stijn.debuck@uz.kuleuven.ac.be`

Abstract. This paper presents a case study towards intra-operative visualization of pre-operative medical images and surgery planning information. We describe a two stage procedure to provide a surgeon with valuable information during surgery, by augmenting live video with a 3D visualization of surgery planning.
A first stage consists of automatic determination of the intrinsic parameters of the video camera by means of one image of a calibration object. It comprises an automatic ellipse extraction and a solution to the 2D-3D correspondence problem. With the intrinsic camera parameters known, we perform a second stage to compute the extrinsic camera parameters with respect to the patient under surgery, allowing us to position and visualize the medical images in the right place.
Techniques are investigated to visualize and manipulate the pre-operative medical images in order to cope with changes in the surgery scene.

1 Introduction

In an increasing number of surgeries, planning based on three dimensional reconstruction and visualization techniques is used. However intra-operatively such planning is shown on a monitor or as printouts besides the operation table. To benefit from this information the surgeon will have to transfer his field of view systematically back and forth between the images and the patient, thereby trying to integrate both. This way he tries to place his instruments as accurate as possible. Although other solutions do exist (navigation, mechanical drill guides), an integrated visualization of pre-operative and intra-operative images can alleviate this problem. This integration of real and virtual images is a typical augmented reality approach (e.g. [1],[2]). Augmented reality systems aim to place virtual objects in real space in such a way the user can not distinguish them from the real ones. This requires a modelling of the camera, a registration of the real world and a natural integration of the virtual objects (e.g. lighting).

Three different steps have to be attended to in an augmented reality system:

– First segmentation is needed to identify real objects in the scene. Possibly several real objects have to be distinguished (e.g. in surgery of joints, one may want to keep track of the two bone structures involved). Segmentation can be simplified through the use of markers allowing to search the image for known geometries, color, ...

– Next, the segmentation results (e.g. markers positions, corresponding points in 2 images, . . .) are employed to obtain calibration and registration of the camera.
– Finally a virtual camera is fed with the 'real' camera and object parameters thus obtained in order to render the virtual objects on the actual camera image.

Other work on transfer of planning by visualization has been reported in [3] and [4]. In their work, it is assumed that patient and intra-operative video camera(s) are not able to move relatively to each other. The approach discussed here differs in a number of technical realisations on automatic camera calibration and holds the potential to release the assumption. Indeed, *visual matching* is an approach to overlay synthetic and real images in an interactive way, thus coping with (small) movements of the patient with respect to the camera.

In this paper we will assume that the intrinsic camera parameters do not change during surgery. This allows to separate the calibration and registration part. During calibration, performed in a automatic way previously to the actual "surgery", we determine the parameters that are proper to the camera (section 2). These parameters are used for rendering the virtual scene, and for the registration of the pose of the camera relative to the patient (section 3). In section 4 we will highlight the visualization and user interaction aspects of our implementation.

2   Automatic camera calibration

Basically, camera calibration is the determination of the relation between 3D points and image coordinates observed by a camera. We can model this relation through several different models of which the complexity depends on the problem at hand and the limitations that are imposed when solving it. We will first discuss the calibration algorithm we used and next the procedure to gather 2D image pixels and corresponding 3D points, used by the algorithm.

2.1   Camera calibration algorithm

From the many algorithms for camera calibration already existing in literature ([5], [6], [7], [8], [9] and [10] with [7], [8], [9] and [10] even as freeware implementations available) we are applying the algorithm developed by R. Tsai ([7], [8]) and implemented by Willson [11].

However our camera model differs a bit, so we first discuss this issue together with our modifications to the algorithm.

Camera models. Mathematically, the pinhole camera, modelling ideal perspective projection, is expressed by an equation of the following form:

$$w_i = K(R^t W_i + T) \tag{1}$$

where w_i is an image point $\left[\,\mu u_i\ \mu v_i\ \mu\,\right]^t$, W_i a 3D point, K an upper triangle matrix, R a rotation matrix and T a translation vector. The matrix

$$K = \begin{bmatrix} \frac{f}{s_x d_x} & s & C_x \\ 0 & \frac{f}{d_y} & C_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

is called the calibration matrix and summarizes the intrinsic parameters of a camera. For a given camera and lens system these parameters are constant: when using zoom and focus they are likely to change. Not only f and s_x, which model the focal length and aspect ratio, but also C_x and C_y will change (cfr Willson [12]). The s appearing in K is a skew factor for the pixel skew and is mostly close to zero and further will be assumed zero.

The matrix R and vector T are determined by the cameras' extrinsic parameters and relate only to the position and orientation of the camera corresponding to the registration part. Zooming and focussing also would affect position and orientation of the camera (cfr Willson [12]) so we will not use these features.

The model of Tsai extends the pinhole model with radial distortion which is a non-linear distortion in the image plane proportional to the square of the distance of a pixel to the center of the image (C_x, C_y) and modelled by two radial distortion coefficients k_1 and k_2 .

Algorithm. Our renderer uses ideal perspective projection to visualize 3D objects. If we would fit our data with Tsai's method the rendering would show not only the normal fitting error but also an error due to the neglection of the radial distortion coefficients.

Tsai's method uses a two-stage calibration with an analytical calculation followed by non-linear optimisation. Since the analytical step is independent of the value of k_1 or k_2, we use the same analytical solution and omit the radial distortion coefficients in the non-linear optimisation function. This results in a best fit to the pinhole-model parameters.

2.2 Elliptical region detection and 2D-3D correspondence

For automatic calibration we use the calibration jig shown in figure 1. It consists of two perpendicular white planes in which circular objects are laid out in a regular pattern. These objects are "closed" black discs, except for the square configurations of four "open" black discs that appear on each plane near the lower corner of the jig. The color of these the open discs is used to distinguish the planes they belong to. In order for our method to be applicable these eight open discs should be visible to the camera.

Complying to the pinhole model, each disc is projected as an elliptical region. Next, we describe how to extract the centroids of the elliptical regions and how to find the correspondence between the centroids and the 3D locations of the discs on the calibration jig.
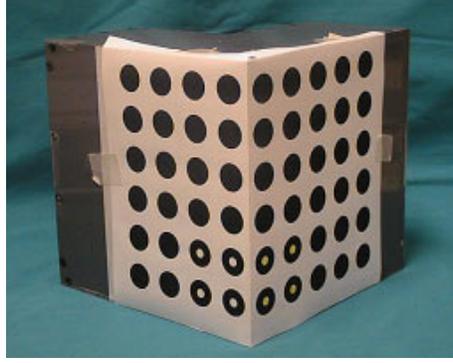
Fig. 1. The calibration jig consisting of two planar patterns

Elliptical region detection. One approach to the problem of recognizing ellipses is applying a Hough transform [13]. This would lead to a five dimensional search space without the guarantee that all the detected features are real ellipses in the image nor that all ellipses are found. Therefore we chose for a two step implementation. First we look for all the objects that could correspond to an ellipse; second, we filter all detected objects such that only the projections of circles remain. This second step is integrated in the 2D-3D correspondence solution.

To start the first stage the image histogram is normalized in order to be robust to variations in lighting. After thresholding and erosion/dilation, a region-based segmentation is done. For every region the bounding box, minor axis, major axis and Euler number are calculated. Only the regions of which the minor axis and major axis are between certain boundaries, are retained. We can employ such a constraint because the distance between camera and object is between known boundaries in a surgery environment.

The coordinates $(EC_x(i), EC_y(i))$ of centriod i are the first order moment divided by the area A of the centriods bounding box [2]:

$$EC_x(i) = \frac{\sum_{boundingbox} \rho(u, v)u}{A} \tag{3}$$

$$EC_y(i) = \frac{\sum_{boundingbox} \rho(u, v)v}{A} \tag{4}$$

with

$$\rho(u, v) = 1 - \frac{\text{the intensity of pixel}(u, v)}{\text{the maximum pixel intensity}}$$

We identify the eight centers originating from the open discs by their Euler number equal to 0. For each of these the pixels around the center are converted to the HSV-colourspace. Based on the hue-component and sufficient saturation we can separate the white from the yellow ones. We then know 4 points from each plane enabling us to solve the 2D-3D correspondence problem.

2D-3D correspondence. In order to solve the 2D-3D correspondence, we consider the image plane and the two planes of the calibration jig as 2D projective planes $\mathcal{I}$, $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively. Then the pinhole projection of each $\mathcal{P}_j$ onto $\mathcal{I}$ is modelled by a homography H_j of projective planes. We show that determining H_j for $j = 1, 2$ will solve the 2D-3D correspondence. Suppose the detected centers of the elliptical regions are described in $\mathcal{I}$ as $m_k = \begin{bmatrix} u_k & v_k & 1 \end{bmatrix}^t$ for $k = 1 \ldots N$, where N is the total number of detected centers. Suppose the known centers of the circular objects on calibration jig plane j are described in $\mathcal{P}_j$ as $m_k^j = \begin{bmatrix} x_k^j & y_k^j & 1 \end{bmatrix}^t$ for $k = 1 \ldots N_j$, where N_j is the total number of objects on plane j. It is good to keep in mind that to each m_k^j does correspond a known 3D point also.

If m in $\mathcal{I}$ corresponds to m^j in $\mathcal{P}_j$ then H_j satisfies

$$\lambda m = H_j m^j \tag{5}$$

So if we know "enough" corresponding pairs (m, m^j), we can solve these equations for H_j. Suppose there are r such pairs. Without loss of generality we can assume them to be (m_i, m_i^j) for $i = 1, \ldots, r$. This leads to a linear system of the following form (in which arbitrarily $\lambda_r = 1$) with $H_j = [h_{kl}]$ and $k, l = 1, \ldots, 3$:

$$
\begin{bmatrix}
E_1 & -m_1 & 0 & \cdots & 0 \\
E_2 & 0 & -m_2 & \cdots & 0 \\
\vdots & & & & \vdots \\
E_{r-1} & 0 & 0 & 0 & -m_{r-1} \\
E_r & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
h_{11} \\
\vdots \\
h_{33} \\
\lambda_1 \\
\vdots \\
\lambda_{r-1}
\end{bmatrix}
=
\begin{bmatrix}
0 \\
\vdots \\
0 \\
m_r
\end{bmatrix}
\text{ with } E_i =
\begin{bmatrix}
m_i^{jt} & 0^t & 0^t \\
0^t & m_i^{jt} & 0^t \\
0^t & 0^t & m_i^{jt}
\end{bmatrix}
$$

So what is the minimum r and how can the corresponding pairs be obtained? After all this whole procedure is meant to find correspondences, but it seems to be dependent on such correspondences. First, $r \geq 4$, since four non-collinear projective points span a basis of a projective plane and the mapping of a basis onto a basis determines the homography. Second due to our design of the calibration jig and the elliptical region detection, for a given plane j four corresponding pairs are readily determined. So we start from these four and obtain an estimate for H_j. This H_j is used to project an other known m_k^j onto $\mathcal{I}$. By taking the elliptical region that is closest to this estimated image point, we associate m_k^j with the center of this region and hence with an m_k. The more pairs that are obtained in this way the better the approximation of H_j from iteratively solving the linear system.

Theoretically this solution should be optimized further by a non-linear scheme ([14], [15]). The linear solution is however sufficient for our purpose of establishing 2D-3D correspondences as input for a subsequent camera calibration. Also, false responses to the elliptical region detection procedure are filtered out in this stage because no world plane coordinates are corresponding to them.

2.3 Accuracy

To determine the overall accuracy of our method we generated a sequence of 30 synthetical images of our calibration jig. After adding an amount of Gaussian noise, these images are input to our routine yielding the results in Table 1. The error depicted here is the difference between a projection of a set of 300 random points with the estimated and original camera parameters. The fact that we even have an error when no noise is added can be explained by discretisation errors of the software renderer (Open Inventor) that generates the synthetical images.

Table 1. Accuracy of the calibration method.

	Variance of the noise		
	No noise	0.01	0.1
Mean projection error (in pixels)	0.743	0.745	0.813
Standard Deviation (in pixels)	0.155	0.188	0.182

2.4 Conclusion

The elliptical region detection and the correspondence deliver us with pairs of 2D-3D coordinates. These are used as input for the modified calibration algorithm giving the camera's intrinsic parameters and its extrinsic parameters with respect to the calibration jig. We only retain the intrinsic parameters which we suppose to be constant throughout the sequel of the approach and use them to model a software camera.

3 Registration

In order to progress towards visual matching, we give some details on a laboratory setup (cfr fig. 2) enabling us to temporarily solve the initial registration problem required.

A plastic phantom skull, on which a number of radio-opaque markers are attached, was CT-scanned. From this CT-volume a 3D surface representation (mimicking the pre-operative planning) was derived (see figure 2). The same skull is then imaged by a camera intrinsically calibrated as described above. For the world 3D coordinate system of the skull, we use the one defined by the CT-image volume.

We need to know the relative position of the skull in the reference frame of the camera or, vice versa, the position of the camera with respect to the skull. This pose determination problem can be seen as an extrinsic camera calibration. The implementation by Willson provides in such a procedure. It requires 7 point matches between world and image coordinates. These are gathered manually by pointing out the radio-opaque markers which are very well visible on the video image also. By corresponding the 2D image coordinates thus obtained to the 3D coordinates of the respective markers from the CT, the camera pose is determined.

Although the procedure is rather coarse, the quality of the results obtained can be judged from fig. 4 (Left).
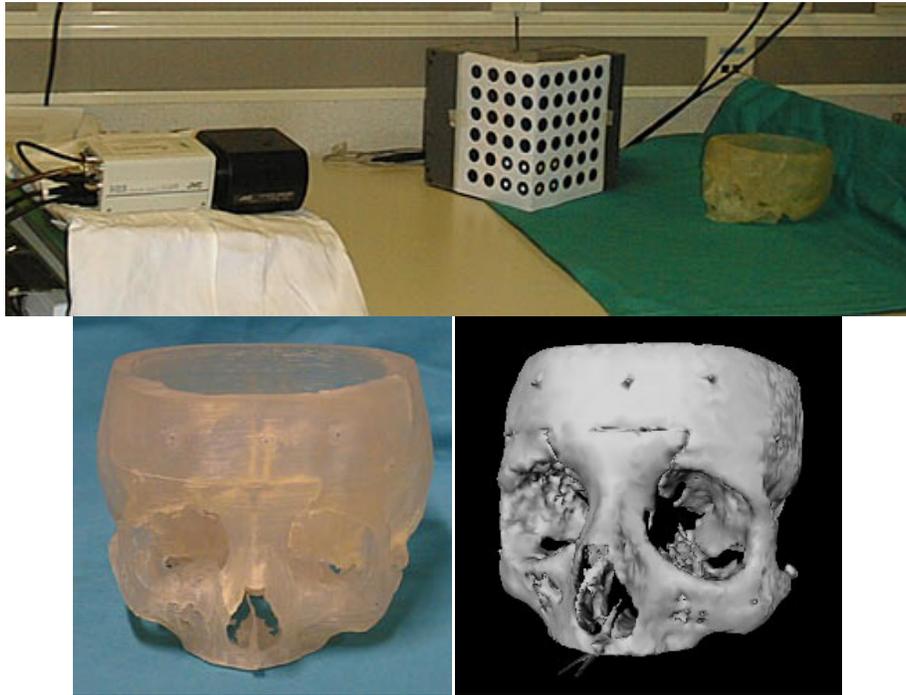
Fig. 2. Top: the laboratory setup we used to test the integrated visualization; Bottom left: the skull phantom; Bottom right: the 3D surface representation of the CT-volume of the skull phantom

4   Intraoperative visualization and visual matching

The visualization itself is developed in Open Inventor (on SGI Octane). The intrinsic and extrinsic parameters of the video camera (JVC, KY-F55B) are fed into a software camera. The viewport of the latter is overlayed on live-video input.

The system can be useful to surgeries where certain pre-operatively determined distances, angles, etc. have to be measured intra-operatively. The measures the surgeon needs during surgery are then planned in the CT-image volume and transferred to the intra-operative video.

We illustrate this procedure by showing a simulation on inserting zygomatic oral implants (osseo-integrating screws). This system is an extension of the work presented in [16]. Pre-operatively the surgeon simulates on the CT-image volume different positions to place the implant and selects the most optimal one (cfr fig. 3). The surgeon has to position his drill at the precise angle and location determined during planning, in order to insert the zygomatic implant. During surgery an integrated view of live video and the rendering of the virtual images by the software camera (cfr fig. 3) is shown on a workstation display. The surgeon can interactively decide which virtual images he wants to see: the CT-image volume, the planned drill axis, …(cfr fig. 3 and fig. 4). Visual matching capabilities are provided: by means of a 3D-mouse the surgeon can realign the virtual images with his patient after a small movement. This is shown in figure 4 (Right). Here the skull is slightly rotated and the virtual objects are realigned, making use of the 3D-mouse.
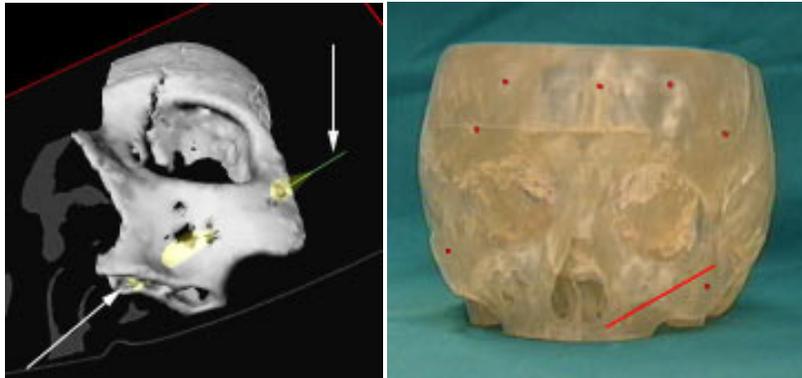

Fig. 3. Left: the planning of the placement of a zygomatic oral implant; Right: Transfer of the planned drill axis to the surgery environment

5   Discussion

In this paper we propose a procedure towards visual matching for intra-operative transfer of planning data.
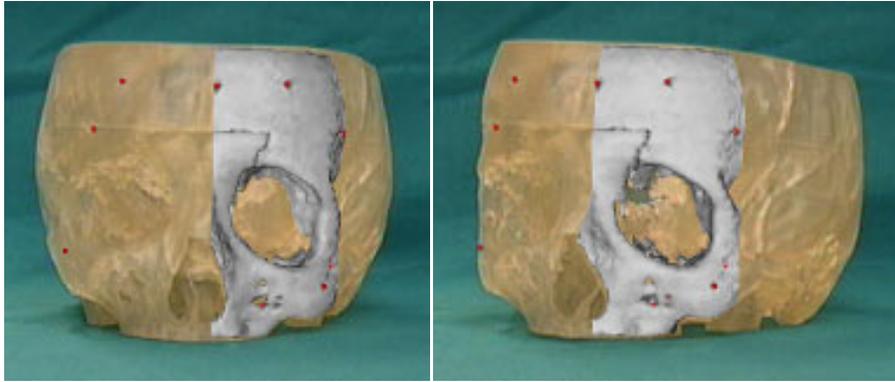
Fig. 4. Left: an integrated view of the skull, a partial CT-surface and markers after the calibration and registration procedure. Right: after the skull is slightly rotated, the virtual objects were realigned by visual matching

Visual matching can only be applied when the virtual world is perceived in the same way as the real world. Modelling of the video camera is a means to fulfill this requirement. The automatic camera calibration procedure we propose offers a robust, fast solution to this problem. So it can be applied without much overhead and specific knowledge in a surgery environment. Furthermore we obtain subpixel accuracy on synthetical images, which is sufficient for our application, and a low sensitivity to noise.

The seperation of extrinsic and intrinsic parameters makes visual matching capabilities like a 3D-mouse possible. In turn, this technique together with the surgeons' clinical experience permits him to adapt the integrated view to his needs in an intuitive manner.

At the moment we are further investigating how depth impression can be induced in order to give the surgeon a better perception of the scene recorded by the camera(s). Techniques like stereo vision and stereoscopic display of the augmented intra-operative video images look promising towards that goal.

Acknowledgements

References

1. J. Vallino. Introduction to augmented reality.
   http://www.cs.rit.edu/~jrv/research/ar/introduction.html.
2. J.P. Mellor. Enhanced reality visualization in a surgical environment. In *MIT AI-TR*, 1995.
3. Colchester et al. Development and preliminary evaluation of VISLAN, a surgical planning and guidance system using intra-operative video imaging. *Medical Image Analysis*, 1(1):73–90, March 1996.

4. Grimson et al. An automatic registration method for frameless stereotaxy, image guided surgery and enhanced reality visualization. *IEEE Transactions on Medical Imaging*, 15(2):129–140, April 1996.

5. D. Vandermeulen. *Methods for registration, interpolation and Interpretation and of Threed-imensional medical image data for use in 3D-display, 3D-modelling and therapy planning.* PhD thesis, KUL, 1991.

6. J.P. Tarel and A. Gagalowicz. Calibration de caméra à base d'ellipses. In *9ème congrès Reconnaissance des Formes et Intelligence Artificielle*, Paris, France, 1994. AFCET.

7. R.Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *CVPR86*, pages 364–374, 1986.

8. R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *RA*, 3(4):323–344, 1987.

9. J. Heikkilä and O. Silvén. A four-step camera calibration procedure with implicit image correction. In *CVPR97*, pages 1106–1112, 1997.

10. J. Heikkilä and O. Silvén. Camera calibration and image correction using circular control points. In *SCIA97*, 1997.

11. R.G. Willson. Freeware implementation of R. Tsai's camera calibration algorithm. http://www.cs.cmu.edu/afs/cs.cmu.edu/user/rgw/www/TsaiCode.html.

12. R.G. Willson. Modeling and calibration of automated zoom lenses. In *CMU-RI-TR*, 1994.

13. Van Gool L. and Proesmans M. *An introduction to image processing and analysis*. 1998. Course text ESAT KULeuven.

14. Z. Zhang. A flexible technique for camera calibration. Technical Report MSR-TR-98-71, Microsoft, 1998.

15. R. Horaud and G. Csurka. Self-calibration and euclidean reconstruction using motions of a stereo rig. In *ICCV98*, pages 96–103, 1998.

16. Verstreken K., Van Cleynenbreugel J., K. Martens, G. Marchal, D. van Steenberghe, and P. Suetens. An image-guided planning system for endosseous oral implants. *IEEE Transactions on Medical Imaging*, 17(5):842–852, October 1998.