

Markerless Endoscopic Registration and Referencing

Christian Wengert¹, Philippe C. Cattin¹, John M. Duff²,
Charles Baur³, and Gábor Székely¹

¹ Computer Vision Laboratory, ETH Zurich, 8092 Zurich, Switzerland
{wengert, cattin, szekely}@vision.ee.ethz.ch

² University Hospital of Lausanne, 1000 Lausanne, Switzerland

³ Virtual Reality and Active Interfaces, EPFL Lausanne, 1015 Lausanne, Switzerland

Abstract. Accurate patient registration and referencing is a key element in navigated surgery. Unfortunately all existing methods are either invasive or very time consuming. We propose a fully non-invasive optical approach using a tracked monocular endoscope to reconstruct the surgical scene in 3D using photogrammetric methods. The 3D reconstruction can then be used for matching the pre-operative data to the intra-operative scene. In order to cope with the near real-time requirements for referencing, we use a novel, efficient 3D point management method during 3D model reconstruction.

The presented prototype system provides a reconstruction accuracy of 0.1 mm and a tracking accuracy of 0.5 mm on phantom data. The ability to cope with real data is demonstrated by cadaver experiments.

1 Introduction

Over the past decade computer aided, navigated surgery has evolved from early laboratory experiments to an indispensable tool for many interventions in spine surgery [1]. Accurate patient registration with the pre-operative imaging modalities or referencing of the target anatomical structure is an essential component of these procedures. Many of today's approaches are invasive and use either mechanical fixation devices, fiducial markers, ionizing radiation or require tedious manual registration with a pointing device. While the potential of ultrasound imaging for non-invasive registration and referencing is currently explored [2] its usability is constrained by the presence of air-filled cavities in the body.

In this paper we present a novel approach that uses a tracked monocular endoscope to register and reference vertebrae during spine surgery. By tracking natural landmarks over multiple views, a 3D reconstruction of the surgical scene can be computed using photogrammetric methods. The proposed algorithm consecutively refines the 3D model with each additional view. This reconstruction provides quantitative metric data about the target anatomy which can be used for 3D-3D registration of the anatomy to the pre-operative data and for further referencing. Our aim is to create an intra-operative support environment, which

relies as much as possible on the tools and instrumentation used anyway during surgery and does not increase the invasiveness of the intervention just for navigation purposes.

The presented method is general and can be used for any rigid anatomical structure with sufficiently textured surfaces. In this paper, however, we concentrate on C2 vertebra referencing during C1/C2 transarticular screw placement, as this open intervention allows us to thoroughly test and validate our methods before applying them in minimal invasive spine surgery.

Various approaches have been proposed to provide the surgeon with more information during endoscopic interventions. Most of them are tailored to specific procedures and are of limited general applicability. In particular the following application fields can be identified: (a) registration and referencing, (b) navigation aids, and (c) augmented reality.

A fiducial marker based system to register freehand 2D endoscopic images to pre-operative 3D CT models of the brain has been presented in [3], and in [4] Thoranaghatte proposed a markerless referencing method for the spine. The use of markers, however, often forces the surgeon to increase the invasiveness of the intervention just to be able to perform referencing during the surgery. In [5], a hybrid tracking system is proposed using a magnetic tracking sensor and image registration between real and virtual endoscopic images to compute the pose and position of the camera in the CT coordinate frame.

A non-tracked calibrated endoscope for 3D reconstruction and motion estimation from endo-nasal images is used in [6] for registering the CT to the endoscopic video. Another navigation aid using photogrammetry during endoscopic surgery has been studied in [7]. They use the structural information to prevent the endoscope image from flipping upside-down while rotating the camera. In [8] the pose of a specially marked tool inside the surgical scene has been determined from monocular laparoscopic images and used to create 3D renderings from different views for the surgeon.

Augmented reality systems are not necessarily aiming at quantitative measurements but rather want to improve the visual perception for the surgeon by extending the image with additional information. In [9,10,11,12,13] externally tracked cameras are used to augment the surgeon's view by fusing pre-operative data with the actual endoscopic view. In contrast, [14] uses a stereo endoscope instead of a tracker for 3D-3D registration of the surgical scene with the pre-operative model data. An example for 2D-3D registration is presented in [15], where navigation is simplified by displaying the corresponding pre-operative CT or MRI slice next to the intra-operative endoscopic image.

2 Methods

In this section the experimental setup and the algorithms used for the 3D reconstruction based on the point database system and the 3D-3D registration between the computed model and the pre-operative data are presented.

2.1 Experimental Setup

The entire hardware setup is depicted in Fig. 1. For our experiments a 10 mm radial distortion corrected endoscope (Richard Wolf GmbH) with an oblique viewing angle of 25° was used. To avoid interlacing artifacts, we relied on a progressive frame color CCD camera with a resolution of 960×800 pixels and 15 fps. As the depth-of-field of the endoscope/camera combination is in the range of 3 – 8 cm the focal length of the camera can be kept constant during the entire procedure, allowing to avoid the recalibration of the system during surgery.

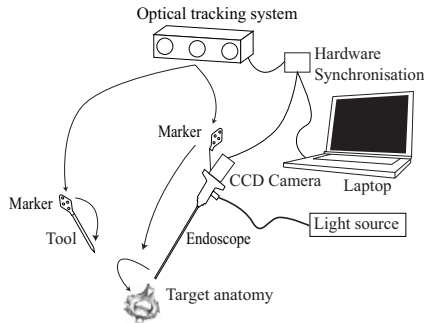


Fig. 1. Overview of the different components of the endoscopic navigation system

A marker is attached to the endoscope that is being followed by the active optical tracker (EasyTrack500, Atracsys LLC). The EasyTrack provides accurate position (less than 0.2 mm error) and orientation information in a working volume of roughly $50 \times 50 \times 1500 \text{ cm}^3$. An external hardware triggering logic ensures the synchronized acquisition of the tracking data and the camera images during dynamic freehand manipulation.

The entire setup is portable and can be installed in a sterile OR environment within minutes. The intrinsic camera and the extrinsic camera-marker calibration can be performed pre-operatively in one step by the surgeon without requiring assistance from a technician [16].

2.2 Point Database

Efficient and consistent data management is an important pre-requisite for the proposed near real-time 3D reconstruction method. All the captured views with their interest points, the views where a specific region is visible, and the continuously updated 3D point coordinates have to be stored and managed reliably and efficiently. In this section a novel approach for handling these data is presented. The algorithms described in the following section rely fully on this database.

The data structure to efficiently handle the related task is shown in Fig. 2 and consists of three parts. First, all the images with the camera pose, the location of the interest points and their feature descriptors are stored in the **Interest**

Point Database. Second, in the tracking phase the point correspondences between the last two views are established. If a feature was already detected in more than two views, the actual view is added to this feature. If the feature has only been present in the last two views, the newly matched points are added to the **Tracking Database**. Finally, during 3D reconstruction and model refinement the **3D Point Database** is filled by triangulating the matched image points from the **Tracking Database**. This database contains the 3D coordinates, the views these points were visible in and the locations of the corresponding 2D interest points. In order to maximize speed and minimize memory load, all references are realized using pointers.

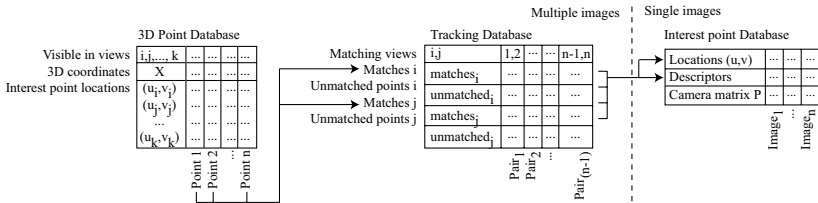


Fig. 2. Database structure for tracking, reconstruction and model refinement

2.3 Algorithm Outline

The method can be divided into three parts: (1) the tracking of feature points across the image sequence, (2) the 3D reconstruction and (3) the registration or referencing of the computed 3D structure.

As the endoscope rarely sees the whole scene at once, we need to find natural landmarks having distinctive features that can be tracked across multiple images and found again when reappearing. Figure 3 gives an overview of the steps performed in the feature tracking algorithm. To initialize the tracking two images are required. Starting with the first image, additional images are acquired until a baseline larger than the predefined threshold of 4 mm is detected. The baseline can be easily calculated using the extrinsic parameters reported by the external tracker. This minimum baseline restriction assures an accurate initial 3D reconstruction. From these two views interest points and their region descriptors are extracted using the method described in [17]. They are matched using the second-nearest-neighbor ratio of their 64-element feature descriptors. A filter using the robust RANSAC fundamental matrix estimation is then used to remove outliers. The remaining point correspondences are then used to triangulate an initial set of 3D points which are stored in the **3D Point Database**.

Once the point tracking is initialized, the trifocal tensor is calculated for the already known 3D points with the last two camera poses where the point was visible. The trifocal tensor is then used to predict the 2D positions of the 3D points in the current endoscopic view. If a point is indeed visible in the

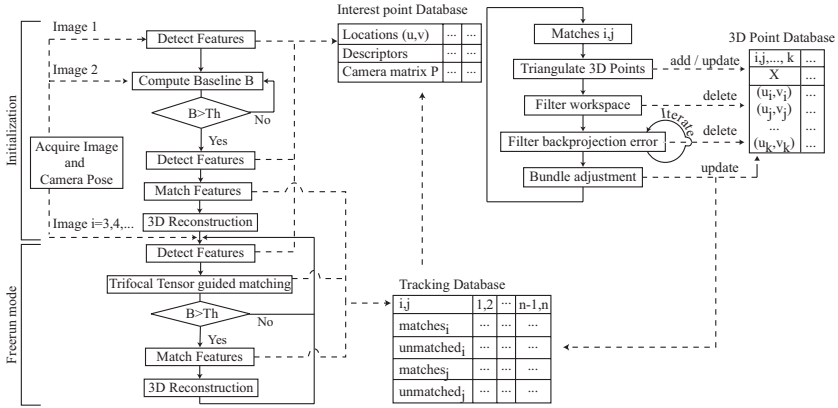


Fig. 3. Tracking and 3D reconstruction algorithm

current view, this view is added to the corresponding **3D Point Database** entry. If the baseline between the current view i and the last view $i - 1$ is less than the predefined threshold, the algorithm captures a new image and starts over with calculating the trifocal tensors. Otherwise, the still unmatched interest points in the current view i are checked if they can be matched to an interest point in the previous view $i - 1$. These new points are added to the **Tracking Database** and the triangulated 3D coordinates are stored in the **3D Point Database**.

In the feature point tracking phase (see Fig. 3 (left)) the **3D Point Database** is continuously updated. In the 3D reconstruction phase (right) the newly found 3D points are triangulated using the optimal triangulation method [18] that shifts the image coordinates to perfectly fulfill the epipolar constraint. The computed 3D points are then added to the **3D Point Database**. As we need accurate 3D points for further processing, two different filters are used to remove false results. The first filter eliminates all 3D points from the database, that are outside the working volume of the external tracking device. The second filter computes the backprojection error for each point in the last 5 views. Iteratively all points having a backprojection error larger than the $threshold = median + 3 \cdot \sigma$ are removed. The median and standard deviation are then recalculated for the remaining points. This filter stops once no more points have been eliminated from the database (usually after two iterations). After the filtering, the database contains a stable set of 3D points. The Levenberg-Marquardt (LM) algorithm is then used to optimize the structure parameters, over the last 5 views. The optimized 3D points are then updated in the **3D Point Database**.

In order to provide a good initialization for the ICP registration, the pre-segmented CT model is manually overlaid on the point cloud. Therefore the point cloud needs to be dense enough for identifying common structures of the two models.

3 Results

In order to assess and quantify the system's performance, we tested the algorithms on synthetic data (C2 vertebra phantom, mimicking blood spots and coagulation residuals) and on ex-vivo data. Three different accuracy tests and a feasibility study based on cadaver experiments were performed using only freehand captured images.

In the first experiment the quality of the metric 3D reconstruction was determined. Therefore, the 3D structure from 25 different sequences of a plastic vertebra were computed. The real thickness of the spinal process of the C2 was then compared to the reconstructed models. The obtained error was 0.10 ± 0.35 mm with the maximum error smaller than 0.8 mm. An example of the reconstructed C2 vertebra and the corresponding CT model is shown on Fig. 4a, b and c.

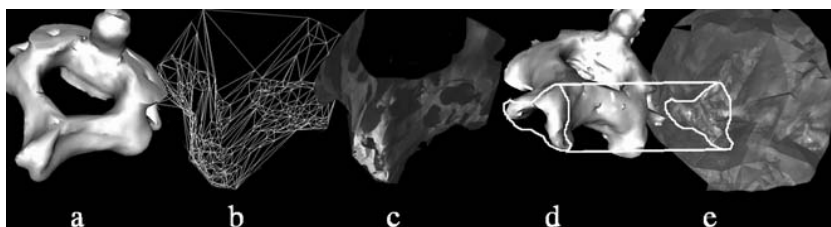


Fig. 4. a) Phantom CT model, b) wireframe and c) textured model of reconstructed phantom, d) Cadaver C2 CT model and e) corresponding textured reconstruction

In the second experiment the repeatability of the localization has been investigated. Therefore, the vertebra was fixed and 10 different sequences recorded followed by the 3D reconstruction. Each of the resulting models was then registered to the CT data using the point-to-point ICP. Then the spatial transformation between the registered models was measured, resulting in a deviation of 0.46 ± 0.31 mm in the position of the vertebra and $1.3 \pm 0.8^\circ$ for the rotation.

In the third experiment the precision of the quantified displacements has been assessed. The vertebra was positioned at different locations using a robot arm. At every distinct position the vertebra was reconstructed in 3D and registered to the CT. The distance between the registered CT models was then measured and compared to the ground truth from the robot thus covering the full error chain. This test has been performed separately in the xy -plane of the tracker's coordinate system with an error of 0.42 ± 0.28 mm as well as in its z -direction yielding an error of 0.49 ± 0.35 mm.

In order to verify the performance of the 3D reconstruction in a more realistic environment, we tested it on ex-vivo data from a cadaver experiment. On average we could identify 500 points on the cadaver's C2 vertebra, which proved to be sufficiently dense for the 3D reconstruction and compares well with the 800 points found on the synthetic body. We observed this similarity in the backprojection error, too, which was (2 ± 1) pixels in both cases, supporting our expectations to be able to reach similar accuracy as in the synthetic environment. Fig. 4d and

4e show the reconstructed pre-operative CT and intra-operative optical models, enabling a visually satisfactory overlay. These preliminary results indicate the usefulness of our method in an in vivo surgical environment.

The feature tracking and consecutive 3D reconstruction process currently uses approx. 1 s per frame. The final ICP registration step uses slightly less than 1 s, thus allowing the surgeon to reference the target anatomy in near real-time.

For the proposed system the following sources of inaccuracy have been identified: (1) intrinsic- and extrinsic errors in camera calibration, (2) inaccuracy of the external tracking device, (3) inaccuracies when matching features, and finally (4) errors from the ICP registration. The potential for improvements in intrinsic and extrinsic calibration is very limited, so further efforts for achieving even higher accuracy will have to concentrate on the last three components.

4 Conclusions

In this paper, a purely optical natural landmark based registration and referencing method using a tracked monocular endoscope was presented. The method is non-invasive and no fiducials or other synthetic landmarks are required. The metric 3D reconstruction accuracy is 0.1 mm. Rigidly registering the 3D reconstruction to the pre-operative 3D CT model using point-to-point ICP resulted in an error of < 0.5 mm proving the applicability of the proposed procedure for non-invasive registration or referencing during navigated spine surgery. Compared to similar work, a high accuracy is achieved, although it is difficult to compare these results due to the different setups and the lack of standardized benchmarks. The need for texture and distinguishable structure on the object for the registration are the two main short-comings of the algorithm. Currently motion in the scene cannot be handled, but it will be investigated to integrate parametrized motion models.

The current processing speed of one second per frame allows the surgeon to register the target anatomy within about thirty seconds. The current implementation, however, still leaves room for further improvement in processing speed.

As the next step, the cadaver experiments will be extended allowing quantitative comparisons with external tracking results, followed by in vivo validation during open C1/C2 transarticular screw placement surgery.

Acknowledgments

This work has been supported by the CO-ME/NCCR research network of the Swiss National Science Foundation (<http://co-me.ch>).

References

1. Schlenzka, D., Laine, T., Lund, T.: Computer-assisted spine surgery. *European Spine Journal* **9 (Suppl 1)** (2000) 57–64
2. Kowal, J., Amstutz, C., Ioppolo, J., Nolte, L.P., Styner, M.: Fast automatic bone contour extraction in ultrasound images for intraoperative registration. *IEEE Trans. Med. Imag.* (2003)

3. Dey, D., Gobbi, D., Slomka, P., Surry, K., Peters, T.: Automatic fusion of free-hand endoscopic brain images to three-dimensional surfaces: creating stereoscopic panoramas. *IEEE Trans. Med. Imag.* (2002)
4. Thoranaghatte, R.U., Zheng, G., Langlotz, F., Nolte, L.P.: Endoscope based hybrid-navigation system for minimally invasive ventral-spine surgeries. *Computer Aided Surgery* (2005) 351–356
5. Mori, K., Deguchi, D., Akiyama, K., Kitasaka, T., Maurer, C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. In: *MICCAI*. (2005) 543–550
6. Burschka, D., Li, M., Ishii, M., Taylor, R.H., Hager, G.D.: Scale-invariant registration of monocular endoscopic images to ct-scans for sinus surgery. *Medical Image Analysis* **9** (2005) 413–426
7. Koppel, D., Wang, Y.F.: Image-based rendering and modeling in video-endoscopy. In: *Proc. of the IEEE International Symposium on Biomedical Imaging: Macro to Nano*. Volume 1. (2004) 269 – 272
8. Caban, J.J., Seales, W.B.: Reconstruction and Enhancement in Monocular Laparoscopic Imagery. In: *Proc. of Medicine Meets Virtual Reality*. Volume 12. (2004)
9. Feuerstein, M., Wildhirt, S.M., Bauernschmitt, R., Navab, N.: Automatic Patient Registration for Port Placement in Minimally Invasive Endoscopic Surgery. In: *MICCAI*. (2005)
10. Bockholt, G., Bisler, U., Becker, A.: Augmented reality for enhancement of endoscopic interventions. In: *Virtual Reality Proceedings*. (2003)
11. Deligianni, F., Chung, A., Yang, G.Z.: Visual Augmentation for Virtual Environments in Surgical Training. In: *MICCAI*. (2003)
12. M. Scholz, e.a.: Development of an Endoscopic Navigation System Based on Digital Image Processing. *Computer Aided Surgery* **3** (1998)
13. Sauer, F., Khamene, A., Vogt, S.: An Augmented Reality Navigation System with a Single-Camera Tracker: System Design and Needle Biopsy Phantom Trial. In: *MICCAI*. (2002) 116–124
14. Mourgues, F., Devemay, F., Coste-Maniere, E.: 3D reconstruction of the operating field for image overlay in 3D-endoscopic surgery. In: *Proc. of the IEEE and ACM International Symposium on Augmented Reality*. (2001)
15. Muacevic, A., Muller, A.: Image-Guided Endoscopic Ventriculostomy with a New Frameless Armless Neuronavigation System. *Computer Aided Surgery* **4** (1999)
16. Wengert, C., Reeff, M., Cattin, P., Székely, G.: Fully automatic endoscope calibration for intraoperative use. In: *BVM*, Springer-Verlag (2006) 419–23
17. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: *Proc. of the ninth European Conference on Computer Vision*. (2006)
18. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press (2000)