Multi-View 3D Skin Feature Recognition and Localization for Patient Tracking in Spinal Surgery Applications

Francesca Manni (f.manni@tue.nl)
University of Technology Eindhoven: Technische Universiteit Eindhoven https://orcid.org/0000-0003-0470-2299

Marco Mamprin
University of Technology Eindhoven: Technische Universiteit Eindhoven

Ronald Holthuizen
Philips Medical Systems: Philips Healthcare

Caifeng Shan
Shandong University of Science and Technology

Gustav Burström
Karolinska University Hospital: Karolinska Universitetssjukhuset

Adrian Elmi-Terander
Karolinska University Hospital: Karolinska Universitetssjukhuset

Erik Edström
Karolinska University Hospital: Karolinska Universitetssjukhuset

Svitlana Zinger
University of Technology Eindhoven: Technische Universiteit Eindhoven

Peter H.N. de With
University of Technology Eindhoven: Technische Universiteit Eindhoven

Research

Keywords: patient tracking, spinal surgery, skin tracking, surgical guidance, feature localization, surgical guidance

Posted Date: October 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-87107/v1

Version of Record: A version of this preprint was published on January 7th, 2021. See the published version at https://doi.org/10.1186/s12938-020-00843-7.

RESEARCH

Multi-view 3D skin feature recognition and localization for patient tracking in spinal surgery applications

Francesca Manni[1*†], Marco Mamprin[1†], Ronald Holthuizen[2], Caifeng Shan[3], Gustav Burström[4], Adrian Elmi-Terander[4], Erik Edström[4], Svitlana Zinger[1] and Peter H.N. de With[1]

[*]Correspondence: f.manni@tue.nl
[1]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
Full list of author information is available at the end of the article
[†]Equal contributor

Abstract

Background: Minimally invasive spine surgery is dependent on accurate navigation. Computer-assisted navigation is increasingly used in minimally invasive surgery (MIS), but current solutions require the use of reference markers in the surgical field for both patient and instruments tracking.

Purpose: To improve reliability and facilitate clinical workflow, this study proposes a new marker-free tracking framework based on skin feature recognition.

Methods: Maximally Stable Extremal Regions (MSER) and Speeded Up Robust Feature (SURF) algorithms are applied for skin feature detection. The proposed tracking framework is based on a multi-camera setup for obtaining multi-view acquisitions of the surgical area. Features can then be accurately detected using MSER and SURF and afterwards localized by triangulation. The triangulation error is used for assessing the localization quality in 3D.

Results: The framework was tested on a cadaver dataset and in eight clinical cases. The detected features for the entire patient datasets were found to have an overall triangulation error of 0.207 mm for MSER and 0.204 mm for SURF. The localization accuracy was compared to a system with conventional markers, serving as a ground truth. An average accuracy of 0.627 and 0.622 mm was achieved for MSER and SURF, respectively.

Conclusions: This study demonstrates that skin feature localization for patient tracking in a surgical setting is feasible. The technology shows promising results in terms of detected features and localization accuracy. In the future, the framework may be further improved by exploiting extended feature processing using modern optical imaging techniques for clinical applications where patient tracking is crucial.

Keywords: patient tracking; spinal surgery; skin tracking; surgical guidance; feature localization; surgical guidance

Background

The insertion of pedicle screws is a critical step in spine fixation surgery. Conventional open surgery is performed through a midline incision where the posterior aspect of the spine is exposed. However, there is a trend towards increased use of minimally invasive surgical techniques, due to reductions in blood loss, length of hospital stay and surgical site infections (1). Minimally invasive surgery (MIS) is performed through small skin incisions where the vertebrae are reached by use of

tubular retractors (2). Due to the reduced visibility during MIS procedures, intraoperative imaging such as fluoroscopy is frequently used. However, to reduce radiation exposure and increase accuracy, a number of computer-assisted navigation solutions have been devised (3–6). Clinical studies have shown that the use of intra-operative three-dimensional (3D) imaging coupled to a navigation system leads to higher accuracies than competing technologies (7). All navigation technologies require co-registration of the patient and the pre- or intraoperative images to allow tracking of both patient and surgical instruments relative to the medical images. Conventional navigation solutions typically include infra-red camera systems tracking a dynamic reference frame attached to a vertebra (8). The navigation technology used in this study is an augmented-reality surgical navigation (ARSN) system relying on adhesive optical skin markers for motion tracking and compensation (9; 10). Four high-resolution optical cameras are integrated in the flat detector of a C-arm with cone-beam computed tomography (CBCT) capability. The markers are recognized by the cameras and their relative positions in space is used to create a virtual reference grid, which is co-registered to the patient during CBCT acquisition. Optical markers attached to the patient's skin have been used for respiratory motion tracking (11) and for medical imaging applications (12). The use of optical markers for motion tracking is combined with digital image correlation and tracking techniques (13–16). Recently, Xue *et al.*, demonstrated that ink dots on the skin could be video tracked with high precision and that the post processing retrieved more detailed information compared to marker-based methods (17). Similarly, direct tracking of spine features and tracking of skin features using hyper spectral cameras for spine surgery, have recently been proven feasible.

In this study, a new marker-less tracking technology using grey-scale video cameras based on skin feature detection was evaluated. Image analysis techniques were applied to detect and track natural features of the skin. There are several advantages when refraining from using optical markers for motion tracking. First, the workflow of the procedure can be improved by simplifying the protocol for patient preparation and by increasing the reliability of tracking during the surgical procedure. Second, the risk of losing sight of the markers can be abolished when skin features can be used as a reference. The well-known feature detection algorithms, Maximally Stable Extremal Regions (MSER) and Speeded Up Robust Features (SURF) were applied to detect and extract skin features such as moles and pigment spots (18–21). The proposed 3D localization framework, used multi-view geometry principles to perform image rectification, enhance feature detectability, improve feature matching and calculate and assess each triangulated feature. The Sum of Squared Differences (SSD) was used as a feature matching metric on scan lines between multiple view acquisitions. To remove 3D outliers, a second feature selection step was applied, specifically performed for the z-coordinate mismatch after the triangulation of all pairs of matched features. The final inliers were evaluated by computing the overall mean triangulation error. In summary, the contribution of this paper is an alternative to marker-based tracking. We hypothesize that the camera feed provides enough details for skin feature detection and tracking. The sub-millimeter localization accuracy achieved was sufficient for surgical navigation. The framework included 3D reconstruction and feature localization over multiple-

view acquisitions. It was validated on eight clinical spinal surgery cases performed in an academic tertiary medical center.

Results

Optical data for assessing the 3D localization were collected from two sources: a cadaver study and a prospective clinical observational study. The cadaver study was performed according to all applicable laws and directives. The clinical study was approved by the local ethics committee and all enrolled patients signed informed consent. The data that support the findings of this study are generated by Philips Electronics B.V., Best, The Netherlands and Karolinska University Hospital, Stockholm, Sweden. All images of the datasets were acquired at the same UHD resolution of 2,592 pixels by 1,920 lines.

The first dataset consisted of one multi-view acquisition, thus 4 images, of a cadaver. MSER and SURF were applied to several selected regions to perform the first multi-view experiment for skin localization. The second dataset consisted of image data from eight patients included in a spine navigation study and taken during the surgical procedures. The data was used to perform two different experiments, first a skin feature localization and later an optical marker localization and the ground-truth comparison. All patients were classified by the physicians regarding Fitzpatrick Skin Type I, II or III. The first feasibility study was performed by analyzing the skin of the cadaver. The localization system was used for four flat selected regions with the c1|c3 camera pair. The features detected for this dataset were triangulated with a mean triangulation error of 0.239 mm for MSER and 0.218 mm for SURF. Due to its intrinsic functional operation, the MSER algorithm detects multiple blobs located at the same coordinates. This explains why MSER seems to have the capability to detect more features than SURF. The discarded-feature ratio of the matched features to the selected inliers was 3.96 ± 0.80 and 2.93 ± 0.45 for MSER and SURF, respectively.

The clinical dataset involved patients undergoing open surgical procedures via mid-line incisions along the spine. Several plane regions were carefully selected for some patients (2nd, 4th, 7th), where the skin was partially covered by blood. The performance results of the localization framework for all the eight patients in the study are shown in Tables 1 and 2. Figure 1 shows two examples of MSER and SURF feature detection and corresponding matches at the same location for an image pair. On a total amount of 4,934 (MSER) and 1,727 (SURF) features, a mean triangulation error of 0.207 and 0.204 mm were reached for MSER and SURF, respectively. An important observation was that 75% of the detected features had a triangulation error within 0.3 mm (Figure 2), appropriate for spinal surgery applications. The discarding ratio of the matched features to the selected inliers in this case was 3.73 ± 2.69 and 2.61 ± 1.70 for MSER and SURF, respectively.

Marker localization and ground-truth comparison

The marker detection and ground-truth comparisons were performed by applying both MSER and SURF feature detection algorithms, to detect the optical markers positioned on the patient. The mean triangulation error of the tracked markers was 0.290 mm for MSER and 0.303 mm for SURF, as shown in Table 3. Table 3 portrays that an average Euclidean distance of 0.627 mm for MSER and of 0.622 mm for SURF are reached, in relation to the ground truth.

Figure 1 Two examples of MSER and SURF feature detections (a and b) and corresponding matches (c and d) at the same location for an image pair.

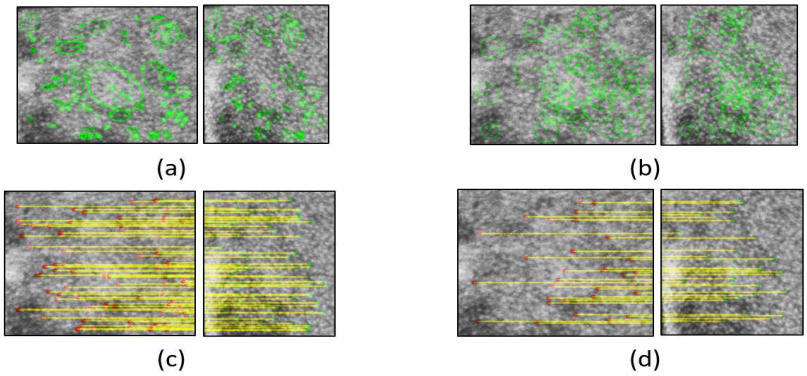


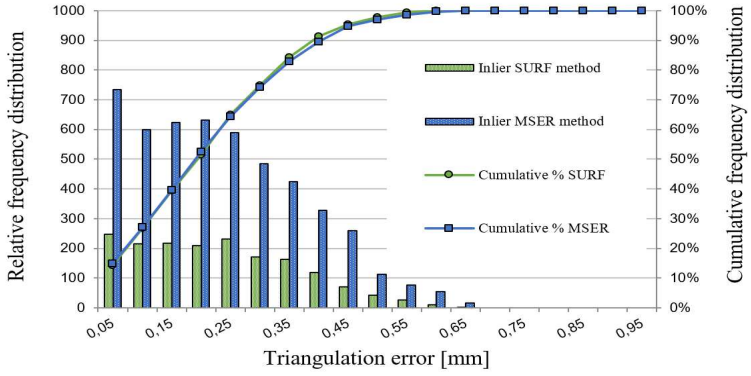
Table 1 Triangulation error analysis - MSER

	MSER								
Patient no.	1	2	3	4	5	6	7	8	TOT
Analyzed regions	1	1	1	7	11	3	7	3	ALL
Area [cm^2]	57.8	18.6	27.6	183.9	232.8	75.5	80.0	29.5	705.6
Number of features	366	125	60	1372	1511	264	1083	153	4934
Mean [mm]	0.167	0.162	0.176	0.191	0.269	0.164	0.170	0.214	0.207
Std deviation [mm]	0.110	0.088	0.099	0.122	0.159	0.109	0.123	0.128	0.139
Features/cm^2	6.3	6.7	2.2	7.5	6.5	3.5	13.5	5.2	7.0

Table 2 Triangulation error analysis - SURF

	SURF								
Patient no.	1	2	3	4	5	6	7	8	TOT
Analyzed regions	1	1	1	7	11	3	7	3	ALL
Area [cm^2]	45.2	18.9	29.1	145.5	157.4	48.0	49.8	20.7	514.7
Number of features	130	78	42	577	457	83	299	61	1727
Mean [mm]	0.184	0.148	0.126	0.199	0.268	0.172	0.160	0.195	0.204
Std deviation [mm]	0.113	0.091	0.103	0.125	0.148	0.110	0.120	0.117	0.134
Features/cm^2	2.9	4.1	1.4	4.0	2.9	1.7	6.0	2.9	3.4

Figure 2 Triangulation error distribution (MSER and SURF)- Patient data.



Discussion

This feasibility study proposes a new innovative, accurate and unobtrusive alternative approach for skin feature localization which can be used for patient tracking in surgical navigation. This result was achieved with the direct detection of features

Table 3 Traingulation error for the optical markers

Marker detection	**Triangulation error**		**Euclidian distance**	
Method	MSER	SURF	MSER	SURF
Amount of features	52	43	52	43
Mean [mm]	0.290	0.303	0.627	0.622
Standard deviation [mm]	0.124	0.133	0.101	0.107

on the patient's skin using high-resolution grayscale cameras and the subsequent analysis of the captured multi-view images. The framework is based on MSER and SURF feature detection algorithms to localize the visual skin features.

The accuracy, of roughly 0.6 mm at skin level, achieved with the current framework should be seen in light of previous results obtained by the ARSN system relying on adhesive skin markers. In a recent study using ARSN, Burström *et al.*, demonstrated a technical accuracy of 0.94 ± 0.59 mm and 1.97 ± 1.33 mm, respectively for cadaveric and clinical cases of pedicle screw placement (22). The difference in accuracy, could be attributed to the actual placement of pedicle screws in a clinical situation. Several studies indicate that a key issue in accurate pedicle screw placement is a well-chosen entry point and correct entry into the pedicle (4). A study using a similiar apporach as here for spine feature tracking reached an accuracy of 0.5 mm (23) while a more comparable study using hyperspectral cameras for skin feature detection showed an accuracy of less than 0.5 mm (24). The results obtained by the proposed framework, using grey-scale cameras, are thus well in line with these previous results. A marker-less tracking framework has the advantage of building a virtual reference grid that cannot be dislodged or completely obscured during surgery. Furthermore, compared to conventional dynamic reference frames that track a single vertebra, a marker-less framework can track the entire spine and compensate for inherent movements within the spinal column during surgery. In this study, several datasets were used for validation. MSER showed a better capability of detecting features of observable anatomical skin details. It was visually verified that MSER provided a higher number of detected features contributing to a better plane selection in the 3D outlier removal step. The multi-camera system enabled triangulation of each feature, to obtain an accurate 3D triangulation performance. This performance can be potentially evaluated by automatically computing the triangulation error continuously and in real-time within a software function for navigation. The proposed framework may simplify the existing patient preparation procedure and improve the reliability of the tracking process by relying on skin features instead of optical markers or reference frames.

Limitations

The limitation of this study is the small sample size and the retrospective set-up. The results were validated on 8 clinical cases. The addition of more cases would strengthen the conclusions. A prospective study, comparing different modes of patient tracking (different types of optical markers versus different frameworks for feature tracking) is warranted. Another important issue is to reach enough computational power for real-time tracking of movements and surgical navigation.

Conclusion

This study demonstrates the feasibility of skin feature localization by exploiting an optical multi-view, grey-scale, camera system combined with image analysis and tracking techniques. The system has been tested on several patients undergoing spinal surgery with sub-millimeter accuracy. This study can be the basis for future surgical applications where optical patient tracking is required.

Methods

Image preprocessing

The principles of multi-view geometry (25) are based on assuming a pinhole camera model,applied for correction of camera images with respect to intrinsic parameters. A preprocessing step, the Contrast-Limited Adaptive Histogram Equalization (CLAHE) algorithm, was used to maximize the detection of skin features and reduce the noise during the acquisition (26). The simple computation of the fundamental matrix F with the normalized eight-point algorithm was used for image rectification (27). The fundamental matrix F was defined as:

$$x'^T F x = 0. \tag{1}$$

For any pair of matching points x' and x there are two images in the same coordinate system. The obtained pixel points were imposed on the corrected input images, to enable feature detection.

Feature Detection, Extraction and Matching

Let us consider an image pair captured with different cameras c_i and c_j, from different views: I_{c_i} and I_{c_j} . For both images, a corresponding set of $n(c_i)$ and $n(c_j)$ features respectively, were extracted and saved in a dedicated object ensemble F_{c_i} and F_{c_j} , to capture all information of the detected features $f(c, n)$:

$$F_{ci} = f(c_i, 1), f(c_i, 2), ..., f(c_i, n(c_i)), \tag{2}$$

and

$$F_{cj} = f(c_j, 1), f(c_j, 2), ..., f(c_j, n(c_j)). \tag{3}$$

MSER and SURF algorithms were applied for blob-similar feature design and feature detection. Afterwards, the SURF feature descriptor was applied for feature extraction. The regions of interest (ROIs) on the skin were selected manually and saved as bounding-box coordinates for future iterations. A manual selection was performed because the regions of the same subject were located within different views. Unfortunately, it was not possible to manually select precisely the same region boundary within multiple views. For this reason, the matching process contained some outliers, which were filtered out at a later step of the processing. Attention was paid to select regions where the skin was as flat as possible. For every detected feature, it was necessary to extract a feature vector, known as descriptor that provided information about the feature, in this case the pixels surrounding the

center of the blob. For this purpose, the SURF algorithm was used to extract the feature vector (20). This method was adopted since it offers high reproducibility, even under different viewing conditions. The descriptor vectors were then saved in the following set of dedicated descriptor vectors:

$$\Delta_{ci} = \delta(c_i, 1), \delta(c_i, 2), ..., \delta(c_i, n(c_i)), \tag{4}$$

and

$$\Delta_{cj} = \delta(c_j, 1), \delta(c_j, 2), ..., \delta(c_j, n(c_j)) \tag{5}$$

Where every descriptor $\delta(c, n)$ consisted of a SURF descriptor vector and the y-coordinate of a specific feature n from a generic camera c. Using the previous dedicated descriptor vectors Δ_{ci} and Δ_{cj} , the feature matching step performed a matching between the (c_i)feature detected in one view with respect to the$n(c_j)$ feature from another view and provided an index of correspondences between the two dedicated descriptor vectors Δ_{ci} and Δ_{cj}, and the feature vectors F_{cj} andF_{cj}. These correspondences were achieved by computing the SSD between the SURF descriptor vectors of those features lying within the scan lines of interest. At this point, fusing the epipolar constraint was crucial since it leads the matching process between features that were shifted along an epipolar line for a specific range. Thanks to this top-bottom scan-line stereo matching, matches between features that lie on different epipolar lines were omitted, to reduce the computational cost and maximize the chance of a good match. This step returned two vectors with the indexes related to the matched features. Using these indexes, it was possible to build two new dedicated object ensembles M_{ci} and M_{cj} of equal size, with matched features.

Feature triangulation and outlier removal

In computer vision, triangulation allows determination of the 3D position of a point, given that the positions of the same points are matched in at least two alternative views (25). This was achieved with two or more lines projected from each camera center to the respective point on the camera plane. Consequently, the projected lines did not always intersect in the same 3D point. It was important to evaluate and quantify the accuracy of this method. In this feasibility study, the triangulation error was used as evaluation metric to obtain an index of the triangulation accuracy and then the 3D point locations were used to perform a benchmark against an existing tracking system. The triangulation error was computed by calculating the location of the shortest distance between the two projected lines. The center of the line segment was the triangulated point, and the length of the line segment is the triangulation error, expressed in millimeters (or in micrometers). The triangulation function of the ARSN system returns the 3D Cartesian coordinates of the triangulated point and the corresponding triangulation error through the projections from the camera centers, resulting in:

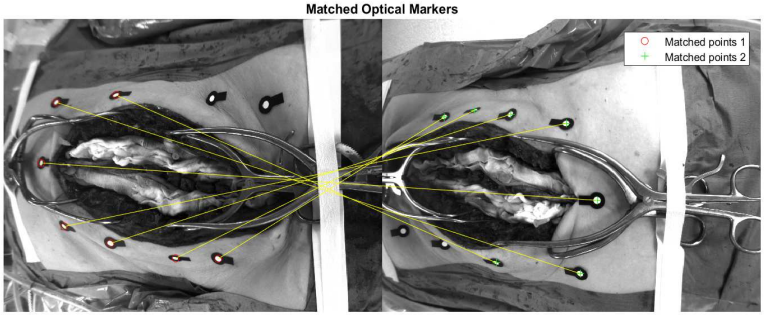
$$\Psi(c_i|c_j, n) = x((c_i|c_j), n), y_(c_i|c_j, n), z(c_i|c_j, n), e_(c_i|c_j, n)), \tag{6}$$

where $c_i|c_j$ denotes the pair of cameras used for the triangulation, $x(c_i|c_j, n)$, $y(c_i|c_j, n)$, $z(c_i|c_j, n)$ are the 3D Cartesian coordinates of the triangulated feature n and $e(c_ic_j, n)$ is the triangulation error of the triangulated feature n with cameras $c_i|c_j$. These 3D coordinates were stored in a dedicated object vector $\Psi c_i|c_j$ for further analysis. The triangulation was achievable between at least two camera pairs. In fact, it was possible to obtain the rectification for the six camera pairs: $c_i|c_j$=$c_1|c_2$, $c_1|c_3$, $c_1|c_4$, $c_2|c_3$, $c_2|c_4$, $c_3|c_4$ and then detect, match and triangulate the features with at least one of these pairs, mainly determined by choosing a more favorable line-of-sight that avoids occlusion of the analyzed regions. The overall mean triangulation error was computed by averaging the triangulation errors of all the features detected over multiple regions. At this point, most of the matches were obtained by only relying on the epipolar constraint. Using the outlier removal step a 3D constraint was imposed, by approximating throe skin surface to a simple planar representation, in which all the detected features were situated. Now, M-estimator SAmple and Consensus (MSAC) (28) was used to fit a plane to the 3D point cloud and remove triangulated points that lying below or above a certain maximum distance from the plane model (expressed in mm). The reference orientation constraint was inferred from the mechanical parameters of the C-arm, once the position of the multi-view camera system with respect to the skin surface was obtained. It was important to quantify the considered region of interest for algorithm comparison and the applicability of the approach. For this purpose, a simple method for skin-area estimation was applied by considering the triangulated features included in the approximated skin-model plane. A flat mesh plot was created from the coordinates obtained by intersecting the normal projections of the features onto the model plane. The overall area was computed as the sum of all the Delaunay triangulation areas included within the boundaries of the region (29). The size of the analyzed region was then expressed in square centimeters.

Marker localization and ground-truth comparison

The accuracy of the proposed framework was evaluated by detection 3D triangulation of the optical markers, which were considered as ground truth. The detection and matching of the optical markers were performed in two different views of the patients. For this purpose, a preprocessing step consisting of a segmentation mask and a simple binarization thresholding was applied. At the end, the MSER and SURF methods were used to perform the feature detection of the bright circular region of every optical marker. The final matching process for a pair of images is shown in Figure 3. When the marker matching was performed, their triangulation was computed, and their 3D location was obtained. The benchmarking was performed with respect to the ground truth of each optical marker provided by the ARSN system. The Euclidean distances in 3D between the detected marker and the corresponding ground-truth marker coordinates were computed to assess the accuracy of the localization.

Figure 3 Matching of the optical markers.



Ethics approval and consent to participate
This study was approved by the local ethics committee and all enrolled patients signed informed consent.

Consent for publication
All case reports consent for publication. That is included in the informed consent.

Competing interests
The authors who are affiliated to clinical and academic institutions (FM, MM, GB, AET, EE, SZ, PHNW) have no financial interests in the subject matter, materials, or equipment or with any competing materials and did not receive any payments as part of the study. Karolinska University hospital and Philips Healthcare have a major collaboration agreement. The authors affiliated with Philips Healthcare (RH, CS) are employees of Philips and thereby have financial interests in the subject matter, materials and equipment. The extent of influence on the data, manuscript structure and manuscript conclusions by these authors and/or Philips Healthcare was limited to technical knowledge and input. Authors without conflicts of interest had full control of all data collection, data analysis and information submitted for publication and over all conclusions drawn in the manuscript. The prototype system described in this article is currently a research prototype and is not for commercial use.

Funding
This work has been financed by the European H2020-ECSEL Joint Undertaking under grant agreement n° 692470, identified as ASTONISH project.

Author's contributions
conceptualization, supervision, writing, review and editing, F.M;investigation, software, conceptualization and writing M.M; conceptualization, supervision, review and editing, S.Z. and R.H.; data, review and editing, G.B., E.E., A.E.T.; supervision, project administration, review, C.S.,supervision, review and editing, P.H.N.d.W. All authors have read and agreed to the published version of the manuscript.

Acknowledgements
We acknowledge: Philips Electronics B.V., Best, The Netherlands, and Karolinska University Hospital for the dataset and the support on conducting this study.

Availability of data and material
Restrictions apply to the availability of these data, which were used under license within a European project (see Acknowledgements) for the current study, and so are not publicly available.

Author details
[1]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands. [2]Philips Healthcare, Best, The Netherlands. [3]Shandong University of Science and Technology, Qingdao, China. [4]Department of Neurosurgery , Karolinska University Hospital and Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden.

References
1. Yoon, J.W., Wang, M.Y.: The evolution of minimally invasive spine surgery: Jnspg 75th anniversary invited review article. Journal of Neurosurgery: Spine **30**(2), 149–158 (2019)
2. Mobbs, R.J., Sivabalan, P., Li, J.: Technique, challenges and indications for percutaneous pedicle screw fixation. Journal of Clinical Neuroscience **18**(6), 741–749 (2011)
3. Devito, D.P., Kaplan, L., Dietl, R., Pfeiffer, M., Horne, D., Silberstein, B., Hardenbrook, M., Kiriyanthan, G., Barzilay, Y., Bruskin, A., *et al.*: Clinical acceptance and accuracy assessment of spinal implants guided with spineassist surgical robot: retrospective study. Spine **35**(24), 2109–2115 (2010)
4. Burström, G., Nachabe, R., Persson, O., Edström, E., Terander, A.E.: Augmented and virtual reality instrument tracking for minimally invasive spine surgery: a feasibility and accuracy study. Spine **44**(15), 1097–1104 (2019)

5. Hott, J.S., Deshmukh, V.R., Klopfenstein, J.D., Sonntag, V.K., Dickman, C.A., Spetzler, R.F., Papadopoulos, S.M.: Intraoperative iso-c c-arm navigation in craniospinal surgery: the first 60 cases. Neurosurgery **54**(5), 1131–1137 (2004)
6. Elmi-Terander, A., Nachabe, R., Skulason, H., Pedersen, K., Söderman, M., Racadio, J., Babic, D., Gerdhem, P., Edström, E.: Feasibility and accuracy of thoracolumbar minimally invasive pedicle screw placement with augmented reality navigation technology. Spine **43**(14), 1018 (2018)
7. Elmi-Terander, A., Burström, G., Nachabé, R., Fagerlund, M., Ståhl, F., Charalampidis, A., Edström, E., Gerdhem, P.: Augmented reality navigation with intraoperative 3d imaging vs fluoroscopy-assisted free-hand surgery for spine fixation surgery: a matched-control study comparing accuracy. Scientific Reports **10**(1), 1–8 (2020)
8. Van de Kelft, E., Costa, F., Van der Planken, D., Schils, F.: A prospective multicenter registry on the accuracy of pedicle screw placement in the thoracic, lumbar, and sacral levels with the use of the o-arm imaging system and stealthstation navigation. Spine **37**(25), 1580–1587 (2012)
9. Edström, E., Burström, G., Nachabe, R., Gerdhem, P., Elmi Terander, A.: A novel augmented-reality-based surgical navigation system for spine surgery in a hybrid operating room: design, workflow, and clinical applications. Operative Neurosurgery **18**(5), 496–502 (2020)
10. Elmi-Terander, A., Burström, G., Nachabe, R., Skulason, H., Pedersen, K., Fagerlund, M., Ståhl, F., Charalampidis, A., Söderman, M., Holmin, S., *et al.*: Pedicle screw placement using augmented reality surgical navigation with intraoperative 3d imaging: a first in-human prospective cohort study. Spine **44**(7), 517 (2019)
11. Dieterich, S., Tang, J., Rodgers, J., Cleary, K.: Skin respiratory motion tracking for stereotactic radiosurgery using the cyberknife. In: International Congress Series, vol. 1256, pp. 130–136 (2003). Elsevier
12. Helm, P.A., Teichman, R., Hartmann, S.L., Simon, D.: Spinal navigation and imaging: history, trends, and future. IEEE transactions on medical imaging **34**(8), 1738–1746 (2015)
13. Burström, G., Buerger, C., Hoppenbrouwers, J., Nachabe, R., Lorenz, C., Babic, D., Homan, R., Racadio, J.M., Grass, M., Persson, O., *et al.*: Machine learning for automated 3-dimensional segmentation of the spine and suggested placement of pedicle screws based on intraoperative cone-beam computer tomography. Journal of Neurosurgery: Spine **31**(1), 147–154 (2019)
14. Wang, F., Behrooz, A., Morris, M.: High-contrast subcutaneous vein detection and localization using multispectral imaging. Journal of biomedical optics **18**(5), 050504 (2013)
15. Yang, R., Wang, Z., Liu, S., Wu, X.: Design of an accurate near infrared optical tracking system in surgical navigation. Journal of Lightwave Technology **31**(2), 223–231 (2012)
16. Asrar, M., Al-Habaibeh, A., Houda, M.: Innovative algorithm to evaluate the capabilities of visual, near infrared, and infrared technologies for the detection of veins for intravenous cannulation. Applied optics **55**(34), 67–75 (2016)
17. Xue, Y., Cheng, T., Xu, X., Gao, Z., Li, Q., Liu, X., Wang, X., Song, R., Ju, X., Zhang, Q.: High-accuracy and real-time 3d positioning, tracking system for medical imaging applications based on 3d digital image correlation. Optics and Lasers in Engineering **88**, 82–90 (2017)
18. Donoser, M., Riemenschneider, H., Bischof, H.: Shape guided maximally stable extremal region (mser) tracking. In: 2010 20th International Conference on Pattern Recognition, pp. 1800–1803 (2010). IEEE
19. Donoser, M., Bischof, H.: Efficient maximally stable extremal region (mser) tracking. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 1, pp. 553–560 (2006). Ieee
20. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: European Conference on Computer Vision, pp. 404–417 (2006). Springer
21. Manni, F., Mamprin, M., Zinger, S., Shan, C., Holthuizen, R., de With, P.: Multispectral image analysis for patient tissue tracking during complex interventions. In: 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 3149–3153 (2018). IEEE
22. Burström, G., Nachabe, R., Homan, R., Hoppenbrouwers, J., Holthuizen, R., Persson, O., Edström, E., Elmi-Terander, A.: Frameless patient tracking with adhesive optical skin markers for augmented reality surgical navigation in spine surgery. Spine (2020)
23. Manni, F., Elmi-Terander, A., Burström, G., Persson, O., Edström, E., Holthuizen, R., Shan, C., Zinger, S., van der Sommen, F., *et al.*: Towards optical imaging for spine tracking without markers in navigated spine surgery. Sensors **20**(13), 3641 (2020)
24. Manni, F., van der Sommen, F., Zinger, S., Shang, C., Holthuizen, R., Lai, M., Buström, G., Hoveling, R.J., Edström, E., Elmi-Terander, A., *et al.*: Hyperspectral imaging for skin feature detection: Advances in markerless tracking for spine surgery. Applied Sciences **10**(12), 4078 (2020)
25. Multiple View Geometry in Computer Vision. Cambridge university press
26. Zuiderveld, K.: Contrast limited adaptive histogram equalization. Graphics gems, 474–485 (1994)
27. Hartley, R.I.: In defense of the eight-point algorithm. IEEE Transactions on pattern analysis and machine intelligence **19**(6), 580–593 (1997)
28. Torr, P.H., Zisserman, A.: Mlesac: A new robust estimator with application to estimating image geometry. Computer vision and image understanding **78**(1), 138–156 (2000)
29. Aurenhammer, F., Klein, R., Lee, D.-T.: Voronoi diagrams and delaunay triangulations (2013)

Figures

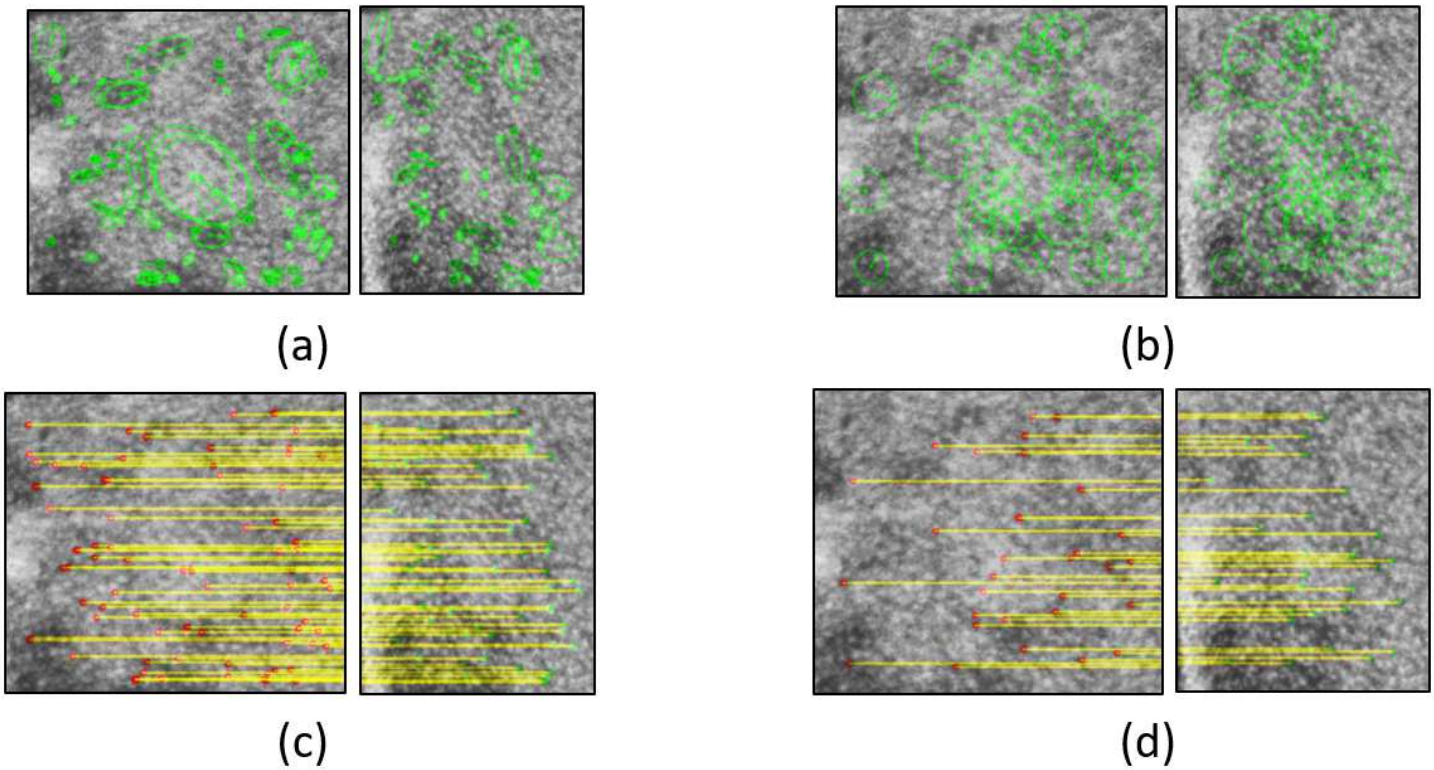


Figure 1

Two examples of MSER and SURF feature detections (a and b) and corresponding matches (c and d) at the same location for an image pair.

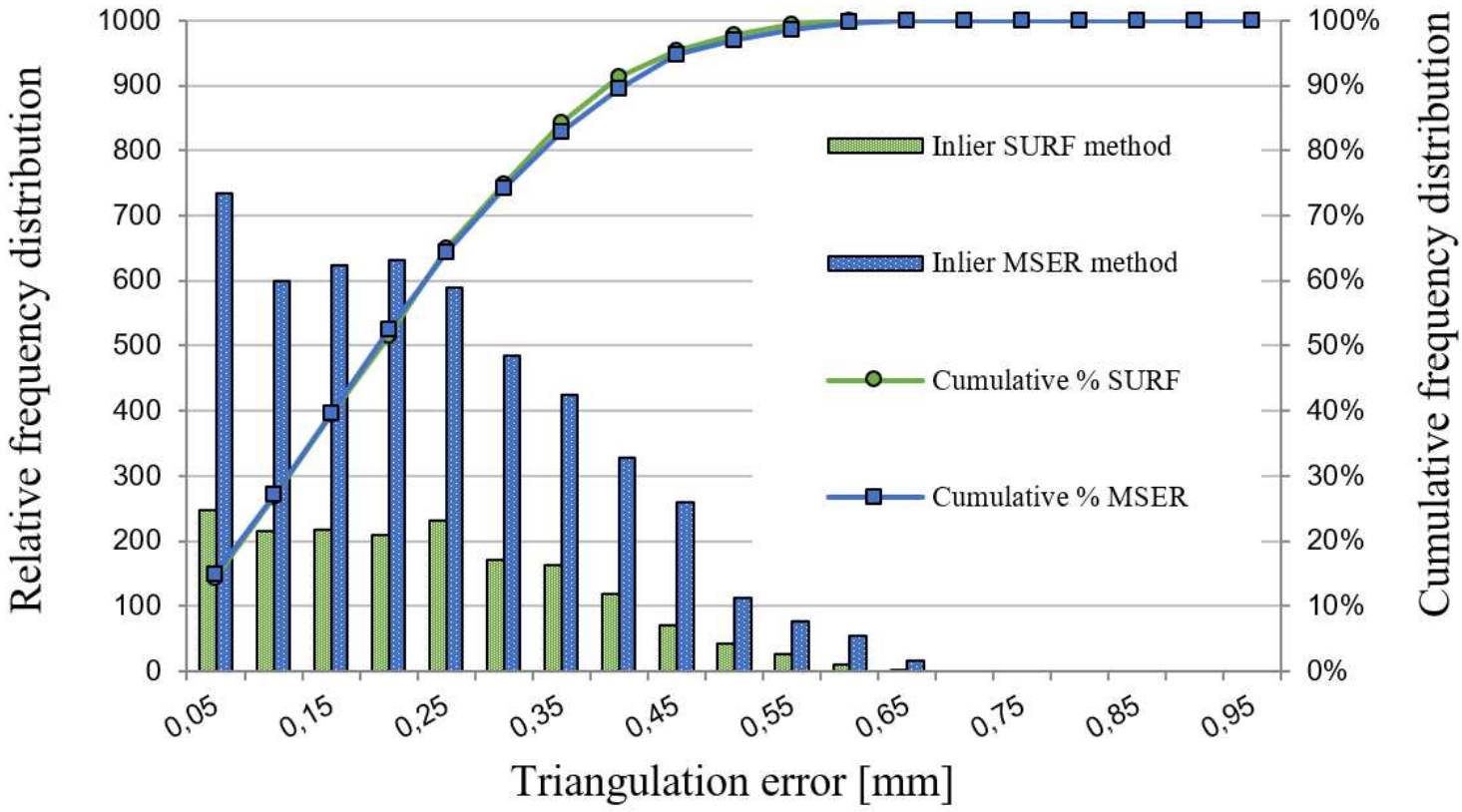


Figure 2

Triangulation error distribution (MSER and SURF)- Patient data.

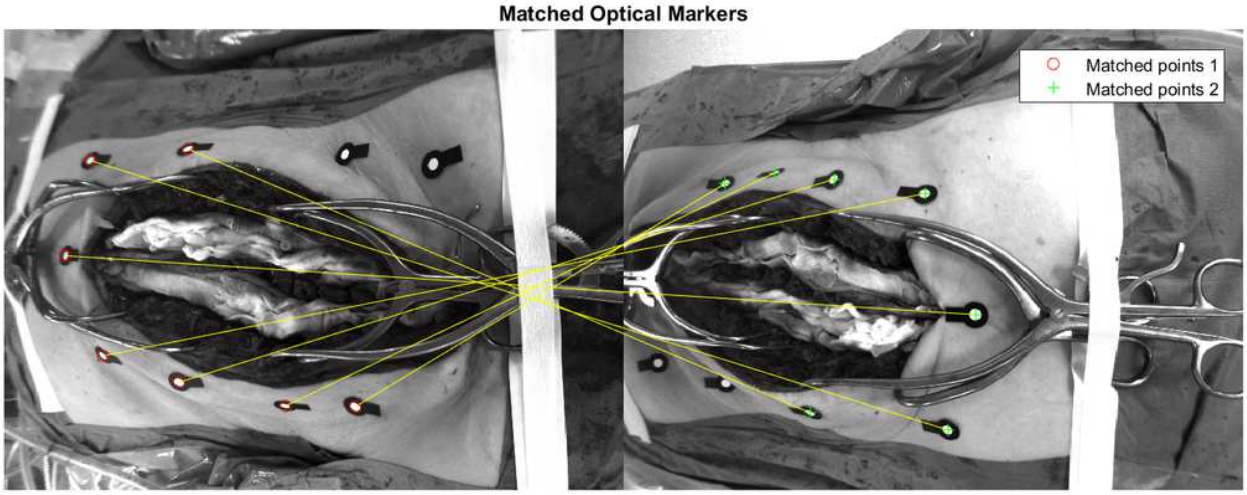


Figure 3

Matching of the optical markers.