Epipolar Consistency in Fluoroscopy for Image-Based Tracking

André Aichert[1]
andre.aichert@cs.fau.de
Jian Wang[1]
jian.wang@cs.fau.de
Roman Schaffert[1]
roman.schaffert@fau.de
Arnd Dörfler[2]
arnd.doerfler@uk-erlangen.de
Joachim Hornegger[1]
joachim.hornegger@cs.fau.de
Andreas Maier[13]
andreas.maier@cs.fau.de

[1] Pattern Recognition Lab
Friedrich-Alexander Universität
Erlangen-Nürnberg, Germany

[2] Department of Neuroradiology
Universitätsklinikum Erlangen
Erlangen, Germany

[3] Graduate School in Advanced Optical Technologies (SAOT),
Erlangen, Germany

Abstract

Geometry and physics of absorption imaging impose certain constraints on X-ray projections. Recently, the Epipolar Consistency Conditions (ECC) have been introduced and applied to motion correction in flat-detector computed tomography (CT). They are based on redundant information in transmission images along epipolar lines. Unlike other consistency conditions for CT scans, they act directly on an arbitrary pair of X-ray images. This paper proposes an application of ECC to 3D patient tracking in interventional radiology. We evaluate the proposed method against 2D-3D registration with a previously acquired CT. Our experiments on synthetic data based on a patient CT and phantom data from an interventional C-arm demonstrate that our method is able to compensate online for rotations of up to ±10° and translations of ±25 mm between consecutive frames. We successfully track rotations of as much as 45° over 45 images. The outstanding property of the approach is that no 3D scan is required for tracking a 3D object in space. We show, that small rotations of about 3° in space and translations of about 50 mm can be tracked based on just two reference X-ray images. Since the proposed approach works directly on X-ray images, it exceeds regular 2D-3D registration with a CT in an order of magnitude in computational speed. We conclude that ECC are a simple and effective new tool for pre-aligment and online patient tracking for fluoroscopic sequences.

1 Introduction

Image guidance in interventional radiology typically relies on both preoperative CT and real-time fluoroscopic images. Correct alignment is either provided by the scan-

ner odometery or, in many cases, must be estimated using 2D-3D registration [8]. Patient motion during the procedure deteriorates the initial alignment. Depending on application, this may include non-rigid periodic effects such as breathing and cardiac motion[3, 5], as well rigid patient movement. In order to recover from prominent movements and re-positioning of the patient, 2D-3D registration is triggered to correct the alignment. In addition, tracking-based approaches allow incremental correction of smaller movements [12].

Flat-detector CT is another X-ray based modality affected by motion. Here, consistency conditions in the raw data domain of X-ray projections offer an attractive means to detect it. Notably, Debbeler *et al.* [4] observed that specific lines in projection images contain redundant information. They exploit this property for raw-data-driven re-calibration of the CT geometry. The method was extended to jitter and motion correction by Maass *et al.* [7] and consequently, Aichert *et al.* [1, 2] presented a derivation and alternative formulation of the Epipolar Consistency Conditions (ECC) using the epipolar geometry of an arbitrary pair of X-ray images. The latter formulation is not restricted to the trajectory of a CT scan. Epipolar geometry is the intrinsic geometry of two pinhole cameras and it therefore applies even to a small set of fluoroscopic images.

This paper investigates the application of ECC in interventional radiology to track patient movement. The idea is to directly exploit redundancies in a few reference X-ray images to estimate 3D motion relative to an unseen image. Our approach can be understood as a low-dimensional alternative to 2D-3D registration, which we consider the gold-standard. However, we do not need a CT to apply the proposed method, which has two major advantages. First, we are computationally much less expensive, since the algorithm works directly on a small set of 2D X-ray images. Second, two or three low-dose images expose the patient to less radiation and can can be acquired and readily updated more easily compared to a CT. We present two sets of experiments. One uses synthetic X-ray images generated from a clinical CT. The other uses real data of a pumpkin phantom acquired with an intervention C-arm. The tracking of two image sequences is complemented by random studies to investigate stability. We conclude that our method provides stable tracking even in the presence of large rotations and translations.

2 Methods

2.1 Epipolar Consistency Conditions

We begin with the introduction of ECC [2] between two X-ray images. Throughout this paper, we refer to "ray-sums" $I(\mathbf{x}) = \log(I_{\text{tube}}/\hat{I}(\mathbf{x}))$ when we speak of projection images. Here, I_{tube} is the initial intensity and $\hat{I}(\mathbf{x})$ is the intensity of an X-ray image at a pixel location $\mathbf{x} \cong (u, v, 1)^\top \in \mathbb{P}^2$ in homogeneous coordinates of the oriented projective plane, where $\cong$ denotes equality up to positive scalar multiples. Every image is associated with a projection matrix $\mathbf{P} \in \mathbb{R}^{3\times4}$, which describes the mapping of world points $\mathbf{X} \cong (X, Y, Z, 1)^\top \in \mathbb{P}^3$ to the image $\mathbf{x} \cong \mathbf{PX}$.

The ECC state that integrating over lines in projection images is *almost* the same as integrating over planes of absorption coefficients through the object. Epipolar lines are pairs of special lines in two images, whose corresponding plane is the same. Thus

there are two redundant ways to compute the plane integral from either of the two images. Epipolar Consistency can thus be quantified by taking several epipolar planes and measuring the difference between redundant line integrals. A line $\mathbf{l} \cong \text{line}(\alpha, t) = (-\sin(\alpha), \cos(\alpha), -t)^\top \in \mathbb{P}^2$ in an image can be expressed by its angle α between the u-axis and its distance to the origin t. Let the epipolar plane $\mathbf{E} \cong (n_X, n_Y, n_Z, d)^\top \in \mathbb{P}^3$ have normal $\mathbf{n} = (n_X, n_Y, n_Z)^\top$ of unit length and signed distance to the origin d, then it is related to the epipolar lines as [6]

$$\mathbf{E} \cong \mathbf{P}_0^\top \mathbf{l}_0 \cong \mathbf{P}_1^\top \mathbf{l}_1. \tag{1}$$

Debbeler *et al.* [4] suggested pre-computing all line-integrals of a projection image I using the Radon transform $\rho_I(\mathbf{l}) = \rho_I(\alpha, t)$ (note the slight abuse of notation) and takes the first derivative of the Radon transform $\frac{d}{dt}\rho_I$ in $t-$direction to account for non-parallel projections [2]. Given two views of the same object or patient denoted by the tupels of image and projection matrix $V_0 = (I_0, \mathbf{P}_0)$ and $V_1 = (I_1, \mathbf{P}_1)$, along with the derivative of their Radon transform $\frac{d}{dt}\rho_{I_0}$ and $\frac{d}{dt}\rho_{I_1}$, the consistency condition for two corresponding epipolar lines $\mathbf{l}_0, \mathbf{l}_1 \in \mathbb{P}^2$ can be expressed by

$$\frac{d}{dt}\rho_{I_0}(\mathbf{l}_0) - \frac{d}{dt}\rho_{I_1}(\mathbf{l}_1) \approx 0. \tag{2}$$

2.2 Cost Function and Optimization

We assume a rigid body motion and express its 6 degrees of freedom in a parameter vector as $\phi = (r_X, r_Y, r_Z, t_X, t_Y, t_Z)^\top$, where r_X, r_Y, and r_Z are three Euler angles defining a rotation matrix $\mathbf{R}_\phi$ and a translation vector $\mathbf{t}_\phi = (t_X, t_Y, t_Z)^\top$. The rigid transformation of the object is then given as a homography in projective three-space which can be right-multiplied to the respective projection matrix

$$\mathbf{T}_\phi \cong \begin{bmatrix} \mathbf{R}_\phi & \mathbf{t}_\phi \\ \mathbf{0}^\top & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4}. \tag{3}$$

Given a set of reference views, i.e. a set of tupels of X-ray images and projection matrices $\mathcal{V} = \{V_1, V_2, \dots\} = \{(I_1, \mathbf{P}_1), (I_2, \mathbf{P}_2), \dots\}$ our goal is to optimize a cost function for the consistency with an input image I_0 and its projection matrix $\hat{\mathbf{P}}_0 = \mathbf{P}_0\mathbf{T}_\phi$, given our current motion estimate $\mathbf{T}_\phi$. Suppose we have computed for each reference view $V_i \in \mathcal{V}$ a set of equiangular epipolar planes $\hat{\mathcal{E}}_i = \{\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_2, \dots, \hat{\mathbf{E}}_N\}$ which intersect the detectors and contain both (finite) positions of the X-ray sources $\hat{\mathbf{C}}_0 \cong \text{null}(\hat{\mathbf{P}}_0) \cong \mathbf{T}_\phi^{-1}\text{null}(\mathbf{P}_0)$ and $\mathbf{C}_i \cong \text{null}(\mathbf{P}_i)$, where $\text{null}(\cdot)$ denotes the kernel of a matrix. For an algorithm to compute these planes, we refer the reader to Aichert *et al.* [2]. We can then measure consistency after transformation between the view $\hat{V}_0 = (I_0, \hat{\mathbf{P}}_0)$ and all reference views based on Equation 2 using the sum of squared difference between the derivative of line integrals in all views

$$M_\phi = \sum_{V_i \in \mathcal{V}} \frac{1}{|\hat{\mathcal{E}}_i|} \sum_{\hat{\mathbf{E}}_k \in \hat{\mathcal{E}}_i} \left(\frac{d}{dt}\rho_{I_0}\left(\hat{\mathbf{P}}_0^{+\top}\hat{\mathbf{E}}_k\right) - \frac{d}{dt}\rho_{I_i}\left(\mathbf{P}_i^{+\top}\hat{\mathbf{E}}_k\right)\right)^2, \tag{4}$$

where $\cdot^+$ denotes the pseudo-inverse. This expression is a metric of the consistency between the input $\hat{V}_0$ with respect to all reference views in $V_i \in \mathcal{V}$ taking into account

the epipolar planes $\hat{\mathbf{E}}_k \in \hat{\mathcal{E}}_i$. In Equation 4, all quantities which depend on the parameter vector $\boldsymbol{\phi}$ are indicated with a hat ($\hat{}$). Our goal is now to find the set of parameters $\boldsymbol{\phi}^\star = \text{argmin}_{\boldsymbol{\phi}} M_{\boldsymbol{\phi}}$, which minimizes the cost function $M_{\boldsymbol{\phi}}$ and therefore maximizes the consistency between the input view $\hat{V}_0$ with respect to the reference views $\mathcal{V}$. It has previously been suggested to find $\boldsymbol{\phi}^\star$ using either grid-search [7] or a non-linear local optimization strategy [1, 2]. In our experiments, we relied on SBPLX [10], a local gradient-free optimizer, since grid-search would not allow real-time results because a fine grid requires too many individual evaluations. The specific choice of optimization technique seems not be too important, since we obtained similar results with the familiar downhill simplex and BOBYQA [9] optimizers.

2.3 Center of Rotation

Since we require that all projections show the same object, it is possible to determine a point close to its center by computing the closest point $\mathbf{X} \in \mathbb{R}^3$ to the principal rays of the reference projections. If no additional information about the geometry of the object is available this point can serve as the center of rotation. For finite source positions, it is easily determined by minimizing the sum of algebraic distances of that point and all X-ray sources in an orthogonal projection in direction of the respective principal ray. Let $\mathbf{m}_i^3 \in \mathbb{R}^3$ be the direction of the i-th principal ray, which can be found in the lower left three elements of the projection matrix $\mathbf{P}_i$ [6] and $\mathbf{C}_i = \text{null}(\mathbf{P}_i) \in \mathbb{R}^3$ be the i-th source position, then $\mathbf{O}_i = \mathbf{I} - \mathbf{m}_i^3 \cdot \mathbf{m}_i^{3\top}$ with the 3×3 identity matrix $\mathbf{I}$ is the orthogonal projection to a plane through the origin orthogonal to the principal ray $\mathbf{m}_i^3$. The point $\mathbf{X}$ should minimize the distance of the projections within that plane

$$\sum_{V_i \in \mathcal{V}} \|\mathbf{O}_i\mathbf{X} - \mathbf{O}_i\mathbf{C}\|. \tag{5}$$

This can be written as a linear least-squares problem. In case of a small number of reference views, we can even solve for $\mathbf{X}$ directly:

$$-\sum_{V_i \in \mathcal{V}} \left(\mathbf{m}_i^3 \cdot \mathbf{m}_i^{3\top}\right)\mathbf{X} = \sum_{V_i \in \mathcal{V}} \left(\mathbf{C}_i - \mathbf{m}_i^3 \cdot \mathbf{m}_i^{3\top}\mathbf{C}_i\right). \tag{6}$$

In practice, this step is important because the world coordinate system is usually located either at a corner of the table on which the patient is lying or in the isocenter of the scanner. For a stable optimization, however, the center of rotation should be located in the center of the structure of interest.

3 Experiments and Results

3.1 Simulation Study Based on a Patient CT

We validated our method using a simulation study of noise-free, absorption-only digitally reconstructed radiographs from a real patient head CT. A salient basis of evaluation is the re-projection error, which we compute based on the corners of cube of side length 10 cm (compare Figures 1, 2). We extracted 15 source and detector positions from real images acquired with an interventional C-arm, see Figure 1, left. The epipolar lines in the first image of the sequence for three reference images are

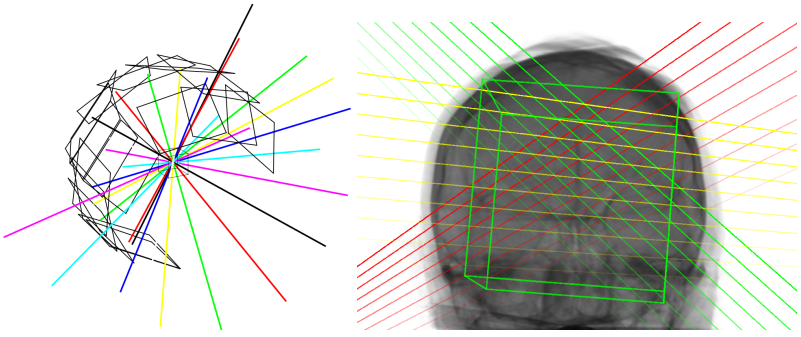

Figure 1: Left: Visualization of 15 source and detector positions around a cube of 10 mm side length, which can actually be reached on a clinical C-arm system; Right: Digitally reconstructed radiograph of a patient CT showing epipolar lines for three reference views;.

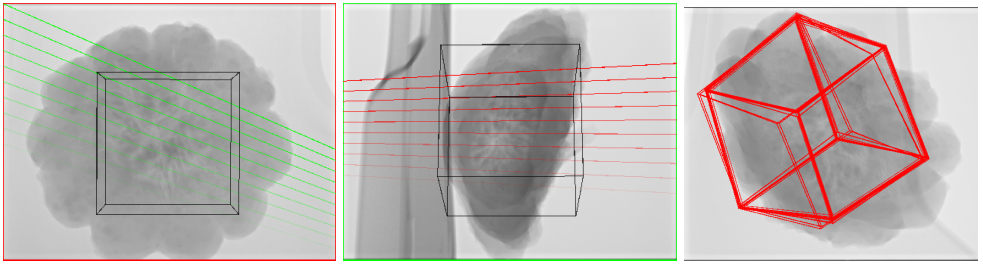

Figure 2: Left, Center: reference images (X-ray intensity before logarithm). Epipolar lines are shown in the color of corresponding frame. Right: Visualization of 25 iterations of a random study of $\pm 10°$ and ± 25 mm showing two outliers.

visualized in Figure 1, center. It is an important observation, that a translation in horizontal direction would only marginally affect the line-integrals of the yellow lines. It is therefore important to have sets of epipolar lines at an angle close to 90°, which assures, that any object motion is observable in at least one of the reference views. Similar to 2D-3D registration [11], an estimation of motion of the object in view direction from just one input image is an ill-posed problem.

We manually created a continuous motion over 300 frames of various translations and rotations of more than 90° and 30° about different axes. Individual frames have a resolution of 1240×960. The radon transforms were computed with 1024×1024 bins. To cover such a large variation in orientation, we selected 5 of the set of 15 reference views from Figure 1, left. We then tracked the motion incrementally by minimizing Equation 4 for each frame in the sequence. We used the ground truth projection matrix as an initial guess for the first image. Supplementary material provides a video sequence of the 300 frames with the 10 cm box overlay. The box is projected once with ground truth (green) and once with estimated projection matrices (red). In this paper, we present Figure 3, left, which shows the ground truth (green) and estimated (magenta) distance of the source to the object along with the angles of rotation relative to the CT. Note that in-plane translations are not shown and that

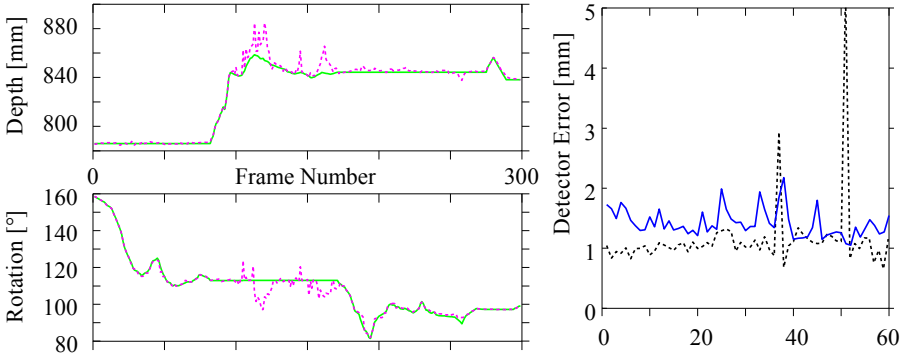

Figure 3: Left: Source-object distance and rotation angle in degrees of the artificial motion used in the simulation study. Both ground truth (green) and estimate (magenta) are shown; Right: The same for second sequence and only two reference images with disturbances of ±1° and ±25 mm.

the angle alone is not a complete representation of the rotation (varying rotation axis is not visualized). We achieved a mean re-projection error of 3.50 pixels or 2.15 mm on the detector, where the mean 3D error in depth was 2.2 mm and the mean error in rotation was 1.3°. Despite we observe 2% outliers of more than 10 pixels, tracking was recovered in consecutive frames without re-initialization. Notably, this experiment demonstrates, that the proposed method can in fact track a 3D rigid motion even with large rotations and translations in all spatial directions using nothing but five reference X-ray images as input. Observe, that the quick translation of the source away from the object was recovered correctly, if only after several outlier frames.

3.2 Pumpkin Phantom Study on Clinical C-Arm

3.2.1 Experimental set-up.

To prove applicability on real data, we acquired 15 images of a pumpkin at a resolution of 2480×1920 with the same source and detector geometry shown in Figure 1, left. 2×2 binning was applied as a first step but errors are reported w.r.t. the original resolution. To simulate motion, we suspended the pumpkin from the ceiling and recorded two short fluoroscopic sequences of 60 frames each. They show a swinging motion of the pumpkin. The first is dominated by rather extreme rotations of about 45°, while the second is dominated by translations of about ±50 mm and only about 3° of rotation. We generated gold-standard projections matrices by 2D-3D registration of individual frames with a CTusing a variation of the algorithm presented in [12]. The reference implementation required several hours to align the CT with all images. Judging from the overlay and smoothness of the sine-like patterns due to the swinging motion, precision appears to be high for the second sequence, while the the first produced two outliers and several unrealistic measurements of depth (i.e. sudden changes of a few centimeters, see green line in Figure 4, top left).

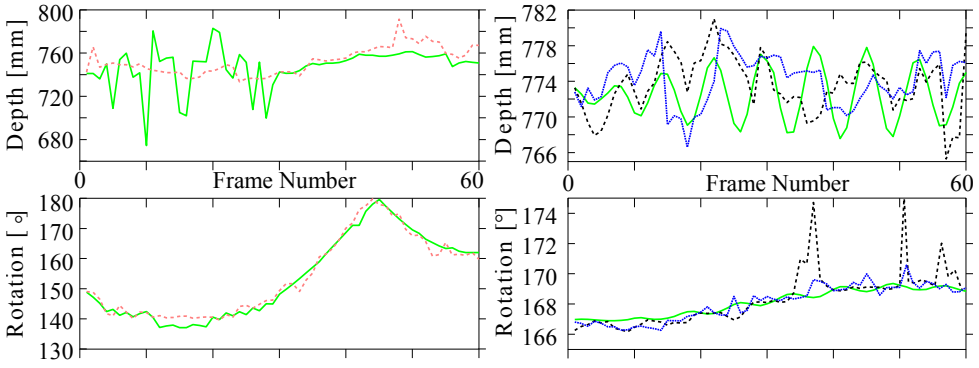

Figure 4: Source-object distance and rotation angle in degrees of reference registration (green) and our method. Left: First sequence, our method in dashed red; Right: second sequence comparison of out method for two (dashed black) and three (blue) reference views.

3.2.2 Tracking

The acquisition of reference views may be associated with additional effort for the clinician. It is therefore important to investigate the performance for very few reference images. We begin by selecting four images to track the first sequence. Three of them are shown in Figure 2. In Figure 4, left, the source-object distance (top) and angle (bottom) is shown for the 2D-3D registration (green) and our method (dashed red). Note that our method actually produces more realistic results for the source-object distance, while the results for the angular parameters are comparable. We achieved an average agreement of our method with the reference up to 1.7° and in terms of re-projection 18.9 pixles or 2.9 mm. We processed the second sequence using the same four reference images to achieve an average disagreement of 0.4° and 9.2 pixels or 1.4 mm. The disagreement of source-object distance was 3.1 mm on average. In a second experiment, we reduced the input to two carefully selected reference images. As a rule of thumb, we observed that views rotated about two orthogonal axes by at least 30° relative to the input image gave best results. We achieved an agreement of up to 0.7° and 7.5 pixels or 1.2 mm, except for two outliers. Detailed results are shown in Figure 4, right and the re-projection error in Figure 3, right. The video in the supplementary material shows several tracked sequences of the pumpkin. The same overlay of the 10 cm cube is used to visualize accuracy. Observe that the rotation in the second sequence is much more extreme than what can be expected in medical data. In spite of that, tracking is never lost.

3.2.3 Random Studies

To get a more reliable estimate of the stability and accuracy of the algorithm, we performed random studies with a sample size of 10 for each of the 60 frames. We used the three reference images shown in Figure 2. For the first sequence, we disturbed the reference projection by uniformly distributed random offsets of as much as 10° and 25 mm and observed, that an initial re-projection error of almost 200 pixels could be recovered up to 18.8 pixels or 2.9 mm, excluding 30% outliers of beyond 50 pixels

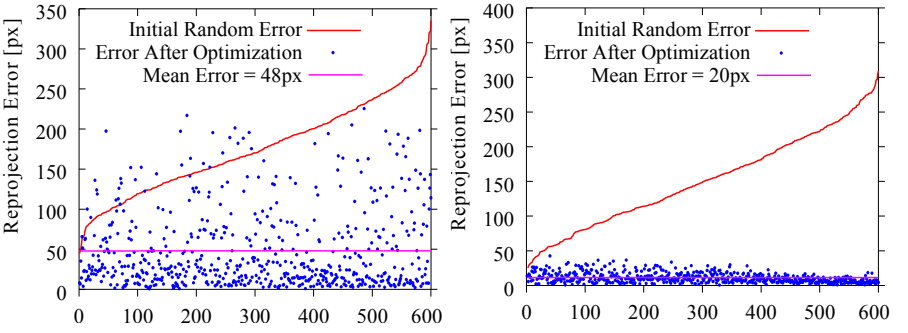

Figure 5: Left: Re-projection errors of the second pumpkin sequence tracked with 2 (dashed black) and 3 (blue) reference images; Right: Random study of 10 samples per 60 frames of the first sequence over random disturbances of ±10° and ±25 mm using three reference images.

error, compare Figure 4, center. This demonstrates the potential of the algorithm to recover even from sudden extreme movements without re-initialization. For the second sequence, we used only two reference views and disturbed the reference by 1° and 25 mm and consistently achieved a re-projection error of 11.6 pixels or 1.8 mm with no outliers, compare Figure 5, right. The latter demonstrates high stability and accuracy for the order of magnitude of motion expected in practice, using just two 2D reference images.

3.2.4 Computation Time

The proposed method is computationally very efficient. An evaluation of M_ϕ is associated primarily with sampling the pre-computed derivatives of Radon transforms $\frac{d}{dt}\boldsymbol{\rho}_i$ with a total of $2 \cdot \sum_{V_i \in \mathcal{V}} |\hat{\mathcal{E}}_i|$ memory accesses, which can be parallelized. Both the number of reference views $|\mathcal{V}|$ and the number of planes per view $|\hat{\mathcal{E}}_i|$ have a linear influence on the computation time. We observed that under-sampling in Radon space results in local minima in the cost function. Since this is undesirable, the number of planes depends directly on the size of the pre-computed radon transform. The number of bins of the Radon transforms can be adjusted, such that computation times meet practical constraints.

The random studies of 600 samples presented in this paper were computed in under five minutes, despite our prototype runs on a mobile Intel i3 CPU for 1024×1024 Radon bins. A higher resolution Radon transform and more epipolar planes produce a smoother cost function. Along with more exhaustive search, this drastically increases the success rate, unfortunately at the cost of computation time. This performance trade-off should be evaluated once a GPU implementation exists.

3.2.5 Shortcomings

There are several assumptions involved in the ECC. Among those, first, an absorption-only model for X-ray physics (Beer-Lambert law of attenuation) is assumed when computing the "ray-sums", which neglects scatter and other non-linear effects of in-

tensity. Second, the method uses very little information about the object. It is essential, that reference views are chosen, which ideally produce sets of orthogonal epipolar lines in the input image. Third, and most notably, we present an image-based method. This means, that the object seen by the input and reference views must actually show the same part of the same object in all views. The method will not be applicable, for example, if there is a highly non-rigid motion of the object, or - obviously - if the images are dominated by other objects or show different parts of the same object. In our experiments, the pumpkin moves relative to the table. The method proved to be robust to this inconsistency. An investigation into the amount of consistency required for the ECC to be effective has yet to be run.

4 Discussion

This paper presents a novel approach to the problem of patient tracking in fluoroscopy. For the first time, we perform 3D tracking of an object solely based on the consistency between 2D X-ray shots. The core idea is that before an intervention under fluoroscopy, between two and five reference images of the region of interest are acquired from different angles. The major contribution of this paper is the realization, that the Epipolar Consistency Conditions (ECC) can be applied not only for motion correction in FDCT, but with the more general formulation of Aichert et al. [2], can be applied to problems in interventional radiology as well. We suggest, that an optimization based on the ECC between an unseen image and the reference images enables us to track a rigid 3D patient pose. The method is fast, simple and relatively robust, despite it does not require a CT scan or other prior information.

We present a proof-of-concept implementation and validate using a synthetic data based on a real patient CT. Additionally, we evaluate the performance of the method with a real pumpkin phantom acquired on an interventional C-arm. In both cases, we are able to incrementally track the object pose, despite extreme motions of about 90° in simulated and 45° in real pumpkin data within 45 images. The method provides a mean error of under 3 mm in our experiments and was even able to recover from outliers. We were able to show that small rotations of about 3° and translations of several centimeters can be compensated for in real data with as little as two reference images. In a random study, we observed stability and high accuracy of < 2 mm in a range of $\sim 1°$ rotation about all axes and ~ 25 mm translations between consecutive frames. With three reference images, even extreme rotations of $\pm 10°$ and translations of 50 mm could be recovered in about 70% of cases. These results are in fact comparable to some 2D-3D registration methods, despite the proposed method uses only a few 2D X-ray images as input and is real-time capable. Future work could also investigate an exhaustive search to determine an initial pose for 2D-3D registration without prior knowledge. Additionally, some type of temporal regularization, for example a Kalman filter, would make the method even more resilient to outliers.

To conclude, we present a new method to track motion in X-ray images. It is able to determine the 3D pose, although it uses only a small set of 2D reference projections. In a phantom experiment, we were able to show that it is real-time-capable and very robust. Even after sudden and large motions, tracking was recovered without the need for re-initialization. An application specific evaluation, such as neurosurgery, with real patient data is in order.

References

[1] André Aichert, Nicole Maass, Yu Deuerling-Zheng, Martin Berger, Michael Manhart, Joachim Hornegger, Andreas K Maier, and Arnd Doerfler. Redundancies in X-ray images due to the epipolar geometry for transmission imaging. In Frederic Noo, editor, *Proceedings of the third international conference on image formation in x-ray computed tomography*, pages 333–337, 2014.

[2] André Aichert, Martin Berger, Jian Wang, Nicole Maass, Arnd Doerfler, Joachim Hornegger, and Andreas Maier. Epipolar Consistency in Transmission Imaging. *IEEE Trans Med Imaging*, Apr 2015.
(Epub ahead of print http://dx.doi.org/10.1109/TMI.2015.2426417).

[3] Alexander Brost, Rui Liao, Norbert Strobel, and Joachim Hornegger. Respiratory motion compensation by model-based catheter tracking during EP procedures. *Med Image Anal*, 14(5):695–706, Oct 2010.

[4] Christina Debbeler, Nicole Maass, Matthias Elter, Frank Dennerlein, and Thorsten M. Buzug. A new ct rawdata redundancy measure applied to automated misalignment correction. In *Proceedings of the Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, page 264, 2013.

[5] Jonathan Hadida, Christian Desrosiers, and Luc Duong. Stochastic 3D motion compensation of coronary arteries from monoplane angiograms. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pages 651–658. Springer, 2012.

[6] Richard I. Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.

[7] Nicole Maass, Frank Dennerlein, André Aichert, and Andreas Maier. Geometrical Jitter Correction in Computed Tomography. In Frederic Noo, editor, *Proceedings of the third international conference on image formation in x-ray computed tomography*, pages 338–342, 2014.

[8] P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš. A review of 3D/2D registration methods for image-guided interventions. *Med Image Anal*, 16(3):642 – 661, April 2012.

[9] Michael JD Powell. The bobyqa algorithm for bound constrained optimization without derivatives. 2009.

[10] Thomas Harvey Rowan. Functional stability analysis of numerical algorithms. 1990.

[11] Žiga Špiclin, Boštjan Likar, and Franjo Pernuš. Fast and robust 3D to 2D image registration by backprojection of gradient covariances. In *Biomedical Image Registration*, pages 124–133. Springer, 2014.

[12] Jian Wang, Anja Borsdorf, Benno Heigl, Thomas Kohler, and Joachim Hornegger. Gradient-based differential approach for 3-D motion compensation in interventional 2-D/3-D image fusion. In *3D Vision (3DV), 2014 2nd International Conference on*, volume 1, pages 293–300. IEEE, 2014.