

Frame Localisation Optical Projection Tomography

Craig Russell^{1,2} and Eric Rees^{1,✉}

¹ Dept. of Chemical Engineering and Biotechnology, Cambridge University, Cambridge, U.K.

² European Bioinformatics Institute

1 OPT, reconstruction, CT, 3D-imaging,
2 Correspondence: ctr26@ebi.ac.uk

3 We present a tomographic reconstruction algorithm, which is
4 applied to Optical Projection Tomography (OPT) images, that
5 is robust to mechanical jitter and systematic angular and spatial
6 drift. OPT relies on precise mechanical rotation and is less me-
7 chanically stable than large-scale computer tomography (CT)
8 scanning systems, leading to reconstruction artefacts. The al-
9 gorithm uses multiple (5+) tracked fiducial beads to recover the
10 sample pose and the image rays are then back-projected at each
11 orientation. The quality of the image reconstruction using the
12 proposed algorithm shows an improvement when compared to
13 the Radon transform. Moreover, when adding a systematic spa-
14 tial and angular mechanical drift, the reconstruction shows a
15 significant improvement over the Radon transform.

16 Sharpe *et. al* proposed OPT (1) using visible light to im-
17 age transparent or translucent mesoscopic samples, with mi-
18 crometer resolution. OPT addresses the scale gap between
19 photographic techniques (for samples typically larger than
20 10 mm), and light microscopy techniques (samples smaller
21 than 1 mm) to image biological samples in the 1 mm to
22 10 mm range.

23 OPT is based on computerised tomography techniques (2) in
24 which a set of projections of a specimen are imaged as the
25 specimen travels through a full rotation. Typically, a Radon
26 transform is then used to transform this set of images into
27 a 3D image stack in Cartesian coordinates (X, Y, Z). The
28 Radon transform relies heavily on the assumption of circular
29 motion with constant angular steps about a vertical axis. This
30 work presents an improved reconstruction algorithm that is
31 robust to spatial and angular mechanical drifts during acqui-
32 sitions, as well as to inconsistent angular steps. The proposed
33 algorithm triangulates points between image pairs to extract
34 camera pose using the theoretical framework used in stereo-
35 scopic imaging.

Stereoscopic imaging

36 When the features or fiducial markers in one view are
37 uniquely identifiable, the stereoscopic imaging of scenes al-
38 lowing for the triangulation of individual features in three di-
39 mensional space (known as world points), see Figure 2 for
40 the coordinate system which describes this geometry. Trian-
41 gulation requires that each feature is detected in both images
42 of a stereo imaging system and for these detections to be cor-
43 rectly associated with one another. This is known as the cor-
44 respondence problem. Various methods exist to ensure that

46 features are detected from image data and accurately associ-
47 ated between two cameras or views (3) and the properties of
48 scale-independent features and their surrounding pixel envi-
49 ronment in one image can thus be matched to a similar feature
50 in a second image.

Coordinates in two adjacent views with a common epi-pole
(see Figure 2) are related by the essential matrix (E) for un-
calibrated cameras and the fundamental matrix (F) for cali-
brated cameras. Their properties are described by:

$$\mathbf{p}'^T \mathbf{E} \mathbf{p} = 0 \quad (1)$$

$$\mathbf{E} = \mathbf{K}'^T \mathbf{F} \mathbf{K} \quad (2)$$

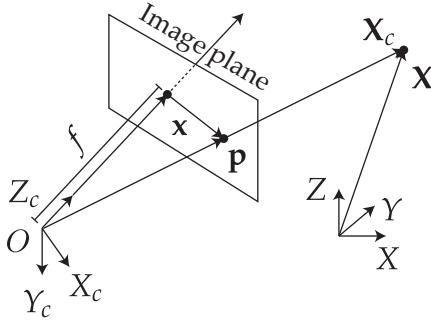
51 Where \mathbf{K} is a matrix that converts image plane coordinates
52 to camera pixel coordinates and where \mathbf{p} refers to a point in
53 the image plane.

The proposed algorithm

54 The motion of a rotating sample, as in an OPT acquisition,
55 with a transformation matrix ($[\mathbf{R} \mid \mathbf{T}]$) in view of a fixed
56 camera is analogous to the motion of a camera with the in-
57 verse transformation matrix around the scene. During an
58 ideal OPT acquisition, a marker will appear to follow an
59 elliptical path in the xy image plane. For the volume re-
60 construction procedure, there is a fitting step to recover the
61 path of the fiducial marker, which is used to correct the sino-
62 gram before applying the inverse Radon transform. This type
63 of reconstruction not only ignores any mechanical jitter of
64 the sample, but also any affine, systematic, mechanical drift
65 (in $X, Y, Z, \theta, \phi, \psi$). Using two adjacent images of a scene,
66 separated by some rotation and translation, world points in
67 3D space may be triangulated within the scene given the ro-
68 tational and translational matrices of the respective camera
69 views.

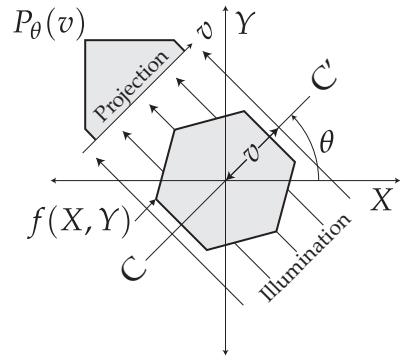
70 Once a sufficient amount of fiducial markers are reliably
71 tracked from the first to the second image, either of the fun-
72 damental or essential matrices can be computed. Using the
73 factorisation of one of these matrices, between each adjacent
74 view of a rotating scene, the translation and rotational mat-
75 rices can be recovered.

76 To reconstruct the image, we compute \mathbf{F} for the current im-
77 age and the first image using 5 or more fiducial markers; hav-
78 ing additional beads helps to remove ambiguity and increase
79 confidence in \mathbf{F} . Once \mathbf{F} is calculated, it is decomposed into
80 \mathbf{R}_n and \mathbf{T}_n between each view n and $n+1$. The image at
81 view $n+1$ is then back projected along the virtual optical



(a) Coordinate system describing a camera with an associated image plane one focal distance f away, imaging an object at point X .

Fig. 1. $\mathbf{X}_c = (X_c, Y_c, Z_c)$ is the camera-centered coordinate point in 3D space. $\mathbf{X} = (X, Y, Z)$ is the world coordinate point in 3D space. $\mathbf{p} = (x, y, f)$ is the ray vector to point of image plane. $\mathbf{x} = (x, y)$ is the image plane coordinates. $\mathbf{w} = (u, v)$ are the pixel coordinates (not shown) corresponding to the point \mathbf{x} . The optical axis travels along the Z_c axis through the image plane.



(b) From an angle θ , an object $f(X, Y)$ and its projection $P_\theta(v)$ are known.

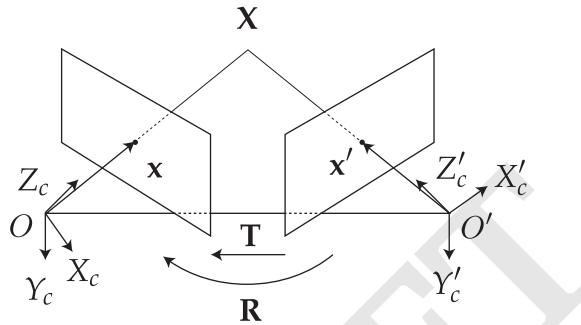


Fig. 2. Epi-polar geometry described for two adjacent views (or cameras of a scene). Coordinates as expressed in Fig. 1a with prime notation ('') denoting the additional right camera view. Transforming from right to left camera-centered coordinates (\mathbf{X}'_c to \mathbf{X}_c) requires a rotation (\mathbf{R}) and a translation (\mathbf{T}).

axis within a virtual volume where the sample will be reconstructed. The size of this back projection and virtual volume is chosen to be suitably large, preventing the loss of important data. The recovered transformation matrices are then matrix inverted and applied to the back projection of the image to realign the rays in the volume to their respective source positions.

In both cases, a decomposed \mathbf{F} matrix will produce four possible transformation pairs (\mathbf{R}, \mathbf{T} ; $\mathbf{R}, -\mathbf{T}$; $-\mathbf{R}, \mathbf{T}$; $-\mathbf{R}, -\mathbf{T}$). Once the transformation matrix between the current view (n) and the first view is calculated, the proceeding transformation matrices are then easily chosen by similarity to the previously collected matrix and general direction of motion. An example of this type of selection would be:

$$\min_{I(n)} \left[I(n) = ([\mathbf{R}_n \mid \mathbf{T}_n] - [\mathbf{R}_{n-1} \mid \mathbf{T}_{n-1}])^2 \right] \quad (3)$$

To find the correct matrix between the $n = 0$ and $n = 1$ orientations, each of the four matrices are compared to an ideal matrix which is composed using *a priori* knowledge of the likely angle of rotation of the system's imaging properties.

the rotation and translation of the image. The reference image is then rotated through 128 angles over 2π radians and projected along the Y axis, then an image slice in (X, Y) is taken to create a single line projection, shown three dimensionally in Fig. 5. This is repeated for each angle, with each line projection stacked to create a sinogram.

In the standard approach for OPT reconstruction, the sinogram undergoes the inverse Radon transform, as shown in Fig. 10a, followed by post-filtering, as shown in Fig. 10b. This step is substituted for the proposed algorithm; in Fig. 6a the two techniques are compared for ideal conditions of smooth, predictable rotation. The proposed algorithm produces a faithful reconstruction on the original image, as shown in Fig. 11. Fig. 6b illustrates the strong overlap of the images produced by the new algorithm and the Radon transform when considering the histogram of the absolute pixel-wise difference between the original source image and the respective reconstructions. The proposed algorithm generates lower deviance from the source image than the Radon transform. The mean square errors (MSE, see Equation Eq. (4)) of the new algorithm and the Radon transform are 15.01 % and 14.84 %, respectively, see Fig. 6b for a histogram of a pixel-wise comparison.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (4)$$

The more challenging case of a sample drifting systemati-

Verification of the proposed algorithm

To verify the validity and quality of the proposed reconstruction algorithm, the image of Lena, superposed with an orthogonal image of Cameraman, is used as a testcard volume. Virtual fiducial beads are dispersed in the volume to track

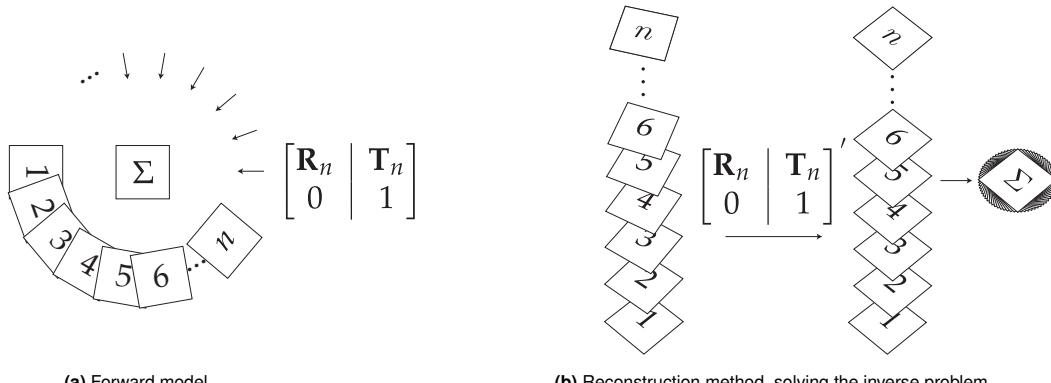


Fig. 3. The simulation of OPT data incorporating rotational and translational offsets, and the proposed reconstruction algorithm. (a): The n projections of the object (Σ), at rotation (\mathbf{R}_1 to \mathbf{R}_n) and translation (\mathbf{T}_1 to \mathbf{T}_n), produces n frames of image data. During the OPT measurement, n projections of the object Σ are observed with rotations \mathbf{R}_1 to \mathbf{R}'_n and corresponding translations \mathbf{T}_1 to \mathbf{T}'_n , where the translations account for imperfect alignment. (b): In the reconstruction algorithm, the rotational and translational matrices are recovered (\mathbf{R}'_1 to \mathbf{R}'_n and \mathbf{T}'_1 to \mathbf{T}'_n) from triangulation of the fiducial markers. These transformation matrices are then used to obtain a contribution to the volumetric reconstruction from each observed frame and the summated reconstruction is assembled from the n frames. The now realigned back projections are summed to produce an unfiltered back projection. The transformation matrices are shown in augmented form using homogenous coordinates.

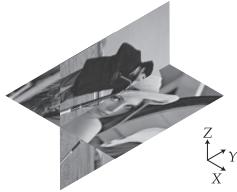


Fig. 4. Ground truth 3D object for reconstruction, based on the Cameraman and Lena testcard images.

123 cally along the X axis, with a constant velocity, was then
124 considered. This drift produced a helical path of a single
125 fiducial within the sample, see Fig. 7b. In Fig. 7c, the Radon
126 transform fails to produce a recognisable reproduction of the
127 test image with the addition of a slight helicity to the rota-
128 tion. The proposed algorithm produces an equivalent result
129 to that of a sample rotating without any systematic drift, see
130 Fig. 10b. In Fig. 6c the respective reconstructions from each
131 algorithm were compared, as before, while the helical shift
132 was incremented. See Fig. 7b for a sinogram of a sample
133 wherein a helical shift has been induced. When using corre-
134 lation as a metric of reproduction quality, the new algorithm
135 fares slightly worse at zero helicity, with 94 % correlation
136 compared to the Radon transform at 96 %. As expected, the
137 Radon transform rapidly deteriorates once a systematic drift
138 is applied, whereas the new algorithm maintains the quality
139 of the reconstruction, see Fig. 6c.

140 **Recovery of \mathbf{R} and \mathbf{T} using matrix decomposition.** To 141 quantitatively verify that the matrix decomposition technique 142 was valid and robust, the accuracy of the reproduction of \mathbf{R} 143 and \mathbf{T} was tested directly. The original \mathbf{R} and \mathbf{T} matrices 144 were computed and compared to \mathbf{R} and \mathbf{T} generated from 145 matrix decomposition. This absolute difference was com- 146 puted element-wise in each matrix and then an average for 147 each matrix was taken. Overall, the worst-case scenario pro- 148 duced a percentage error of 2 % (see Fig. 8 for full statis- 149 tics). The accuracy of the calculated \mathbf{R} and \mathbf{T} deteriorated 150 when adding in additional degrees of combined movement, 151 but with no correlation between the degree of helicity and the 152 error produced. The translation matrix (\mathbf{T}) was consistently 153

154 more accurately reproduced, which is likely due to it having
fewer available degrees of freedom.

Discussion

155 A new algorithm for reconstructing OPT data has been
156 demonstrated. The new algorithm uses multiple fiducial
157 markers to recover the matrix which describes the rotation
158 and translation of the sample. The quality of the reconstruc-
159 tions shows a slight improvement when compared to the stand-
160 ard Radon transform, with a great effect when a systematic
161 drift is introduced. The accuracy of the decomposition of \mathbf{F}
162 into \mathbf{R} and \mathbf{T} was compared to the ground truth matrices. The
163 element-wise absolute difference $\left(\frac{|x-y|}{2(x+y)} \right)$ of each matrix
164 was averaged across the matrix for \mathbf{R} and \mathbf{T} . In the worst-
165 case scenario, a maximum of 2 % average absolute difference
166 was found between ground truth and recovered matrices, sug-
167 gesting that the technique is robust to various forms of drift
168 and general instability. Such an algorithm could be used to
169 minimise ghosting effects seen in real samples, particularly in
170 samples where slipping is likely to occur, such as in gels or
171 in cheaper OPT systems which tend to be more mechanically
172 unstable and imprecise.

Future work

173 The proposed algorithm relies on triangulation between two
174 view points. However, it is possible to use three separate
175 views to reconstruct a scene, one such approach being quater-
176 nion tensors (4). Working with tensors is more complex, but
177 a future iteration of the algorithm presented here may ben-
178 efit from using three views to provide a more accurate trans-
179 formation matrix. Beyond three views, there is currently no
180 mathematical framework for four or more views. If such tools
181 were to be developed, it may be possible to have the algo-
182 rithm described above be a non-iterative, single-shot recon-
183 struction from pixels to voxels.

184 Fiducial markers could also be extracted from the image
185 texture alone, circumventing the need for the additional
186 beads embedded in the sample. To find such corre-



Fig. 5. A 3D test-volume of two orthogonal and different testcard images, from Fig. 4, was used to verify the reconstructive capabilities of the proposed algorithm. The projected image data (b), (h), (j) and (d), (h), (l), were used to iteratively generate reconstructions where the n^{th} reconstruction incorporates all the information from observation 0 to n . The results are unfiltered for clarity of demonstrating the iterative reconstruction, which is applied in Fig. 11.

spondences, points with similar local texture are found and matched in between each image using standard algorithms such as SIFT (5) and RANSAC (6). This was attempted in this work, however, the errors introduced into the transformation matrices make this approach currently inviable.

ACKNOWLEDGEMENTS

3. Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer Science & Business Media. ISBN 978-1-84882-935-0.
4. Hartley Hartley and Richard Hartley. *Multiple View Geometry in Computer Vision / Richard Hartley, Andrew Zisserman*. ISBN 978-0-521-54051-3.
5. D.G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157 vol.2. doi: 10.1109/ICCV.1999.790410.
6. Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. 24(6):381–395. ISSN 0001-0782. doi: 10.1145/358669.358692.

Bibliography

1. James Sharpe, Ulf Ahlgren, Paul Perry, Bill Hill, Allyson Ross, Jacob Hecksher-Sørensen, Richard Baldoek, and Duncan Davidson. Optical Projection Tomography as a Tool for 3D Microscopy and Gene Expression Studies. 296(5567):541–545. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1068206.
2. A. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics. ISBN 978-0-89871-494-4. doi: 10.1137/1.9780898719277.

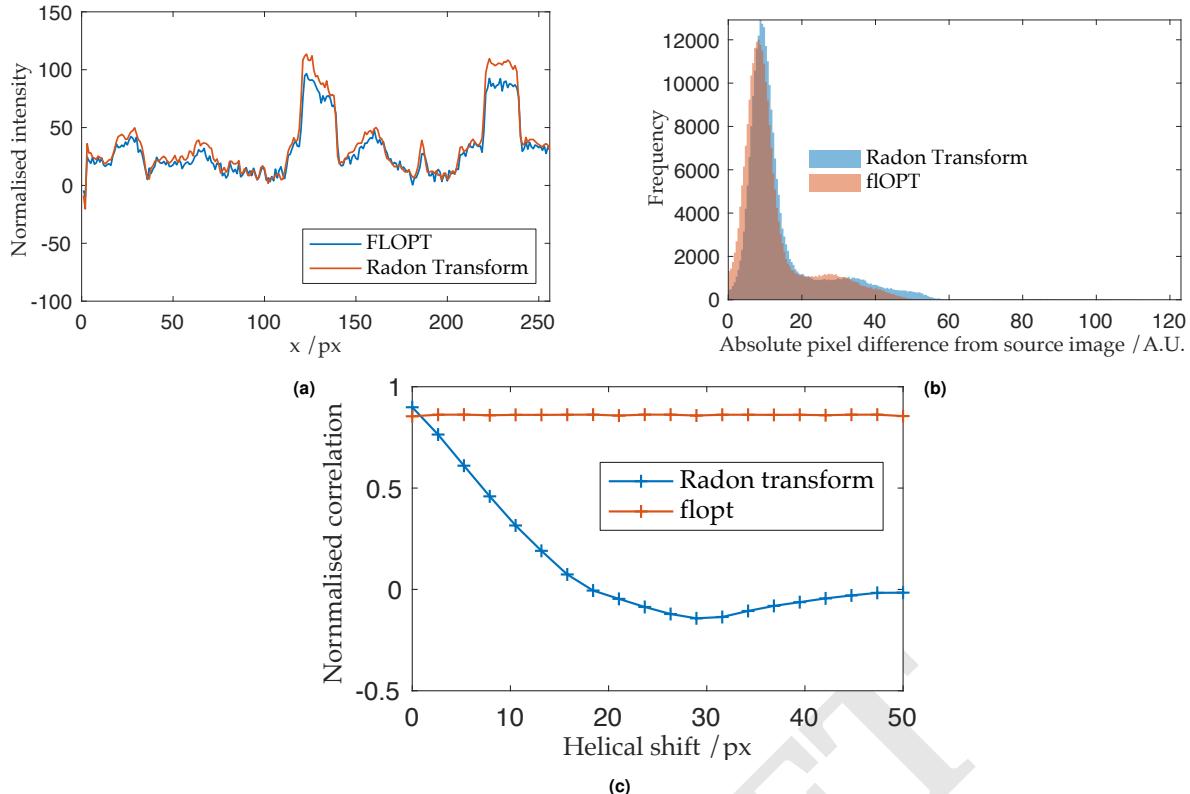


Fig. 6. (a), Line profile comparison of the reconstruction of a reference image computationally rotated, projected and reconstructed using the standard Radon transform and the new proposed algorithm. (b), Histogram of of pixel values compared between reconstructions using the new proposed flOPT algorithm and the Radon transform. The shift of the histogram to towards overall lower deviance from the source image suggests the flOPT algorithm outperforms the Radon transform (c): Comparison of standard and proposed OPT reconstruction algorithms for acquisitions with drift. 2D image correlation of the ground truth and the reconstruction shows that the proposed flOPT algorithm does not degrade with systematic drift, whereas a reconstruction using the standard Radon transform is severely degraded.

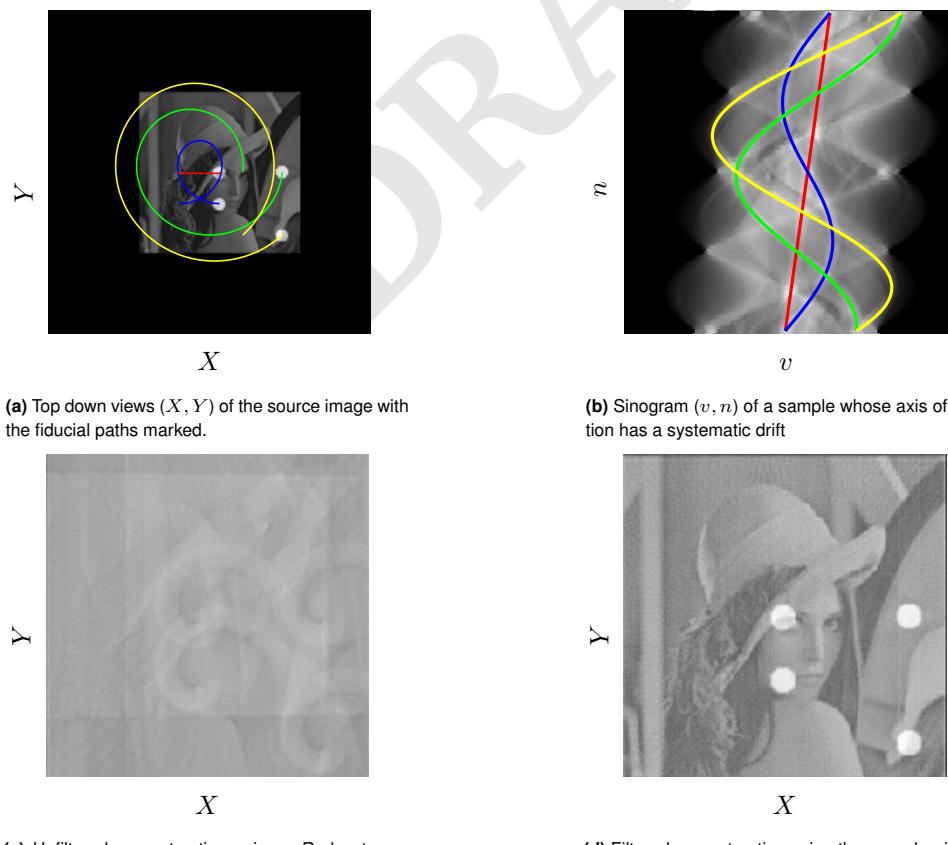


Fig. 7. Comparison of the two reconstructions under sample imaging with a systematic drift, in 3D though represented here in 2D.

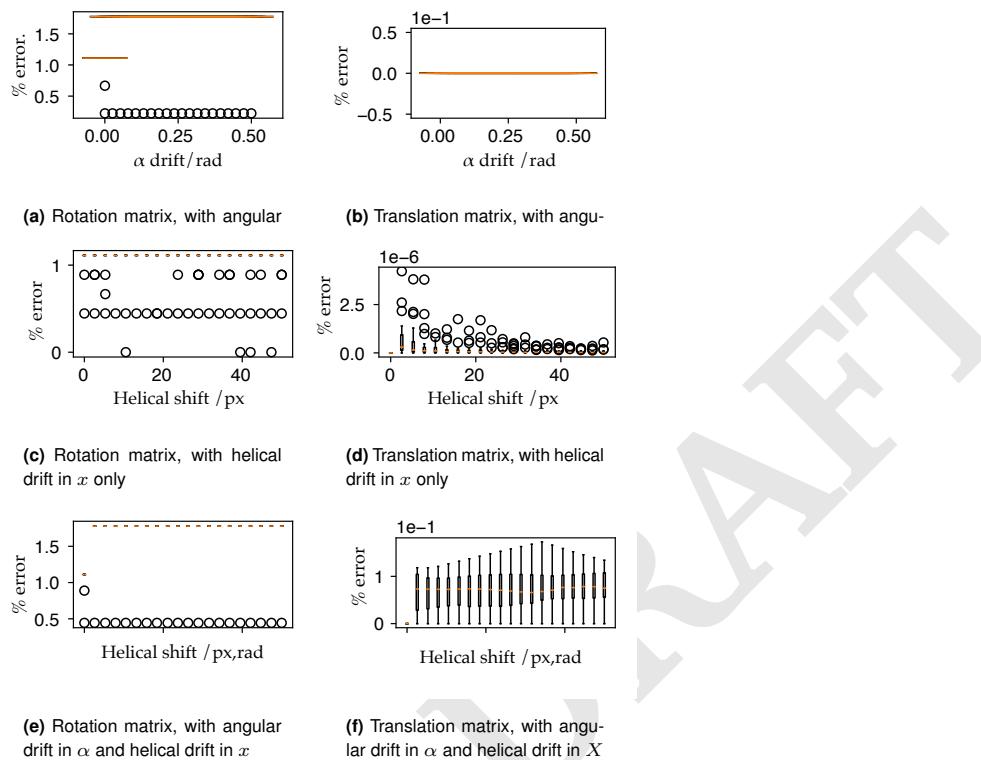


Fig. 8. Box plots demonstrating that the rotational and translations matrices can be recovered accurately from fiducial marker positions. Panels (a) and (b) introduce an angular drift during rotation, to an observer at the detector this would appear as a tip of the sample towards them, causing precession. Panels (c) and (d) introduce a lateral drift in X causing a helical path to be drawn out. Panels (e) and (f) combine the two effects. In all cases, the percentage error introduced by the the addition of undesirable additional movements was on the order of <2 %.

213 Supplementary Note 1: Supplemental

214 **A. Reconstruction.** As the sample is rotated each detector pixel collects an intensity $I(\theta) = I_n e^{-k(\theta)}$ at discrete (n) angles
 215 through a full rotation of the sample; where I_n is the unattenuated radiation intensity from the source to the detector, k is the
 216 attenuation caused by the sample along a detected ray and $I(\theta)$ is the measured intensity, see Fig. 1b. Rays from the sample
 217 to the detector approximate straight lines, and so the rays reaching the detector with a line integrals. A projection is then
 218 the resulting intensity profile at the detector for a rotation angle, and the integral transform that results in $P_\theta(v)$ is the Radon
 219 transform.

The equation of a set of parallel rays from a source passing through the specimen to a point v along the detector is:

$$X \cos(\theta) + Y \sin(\theta) - v = 0 \quad (5)$$

Projecting many such rays through a sample with structure $f(X, Y)$ gives:

$$P_\theta(v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(X, Y) \delta(x \cos(\theta) + y \sin(\theta) - v) dX dY \quad (6)$$

220 Where $P_\theta(v)$ is the Radon transform of $f(X, Y)$ which represents the contrast image of 2D slice of the specimen. The Radon
 221 transform of an image produces a sinogram.

An inverse Radon transform is used to recover the original object from the projection data; which is achieved by taking the Fourier transform of each projection measurement, then reordering the information from the sample into the respective position in Fourier space. This is valid due to the Fourier Slice theorem, which states that the Fourier transform of a parallel projection is equivalent to a 2D slice of the Fourier transform of the original sample.

$$f_{\text{fpb}}(X, Y) = \int_0^\pi Q_\theta(X \cos(\theta) + Y \sin(\theta), \theta) d\theta \quad (7)$$

222 Where Q_θ is the filtered projection data, and $f_{\text{fpb}}(X, Y)$ is the back-projected image. A spatial filtering step is applied during
 223 back-projection to avoid spatial frequency oversampling during the object's rotation (see Fig. 10b) a high pass filter is commonly
 224 used to compensate for the perceived blurring. The blurring arises as Q_θ is back-projected (smeared) across the image plane for
 225 each angle of reconstruction; which means that not only does the back-projection contribute at the line it is intended to (along
 226 line C in Fig. 1a), but all other points along the back-projecting ray.

227 Now, suppose we know the relative positions of the two cameras and their respective intrinsic parameters, such as magnification
 228 and pixel offset. For a single camera and given the camera parameters, we can translate pixel coordinates, $\mathbf{w} = (u, v)$, into the
 229 coplanar image plane coordinates $\mathbf{x} = (x, y)$:

$$u = u_0 + k_u x \quad (8)$$

$$v = v_0 + k_v y \quad (9)$$

230 Knowing the focal length (f) of the imaging system, image plane coordinates may be projected into a ray in 3D. The ray can
 231 be defined by using the point \mathbf{p} in camera-centred coordinates, where it crosses the image plane.

$$\mathbf{p} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (10)$$

232 From the definition of a world point, as observed through an image, we can construct a dual-view model of world points in
 233 space as in Fig. 2. Using a model of a system with two views allows for the triangulation of rays based on image correspondences,
 234 this is an important part of stereo-vision. The most important matching constraint which can be used the *epipolar constraint*, and follows directly from the fact that the rays must intersect in 3D space. Epipolar constraints facilitate the search
 235 for correspondences, they constrain the search to a 1D line in each image. To derive general epipolar constraints, one should
 236 consider the epipolar geometry of two cameras as seen in Fig. 2.

237 The **baseline** is defined as the line joining the optical centres. An **epipole** is the point of intersection of the baseline with the
 238 image plan and there are two epipoles per feature, one for each camera. An **epipolar line** is a line of intersection of the epipolar

240 plane with an image plane. It is the image in one camera of the ray from the other camera's optical centre to the world point
 241 (\mathbf{X}). For different world points, the epipolar plane rotates about the baseline. All epipolar lines intersect the epipole.
 242 The epipolar line constrains the search for correspondence from a region to a line. If a point feature is observed at \mathbf{x} in one
 243 image frame, then its location \mathbf{x}' in the other image frame must lie on the epipolar line. We can derive an expression for the
 244 epipolar line. The two camera-centered coordinate systems $\mathbf{X}'_{\mathbf{c}}$ and $\mathbf{X}_{\mathbf{c}}$ are related by a rotation, \mathbf{R} and translation, \mathbf{T} (see in
 245 Fig. 2) as follows:

$$\mathbf{X}'_{\mathbf{c}} = \mathbf{R}\mathbf{X}_{\mathbf{c}}' + \mathbf{T} \quad (11)$$

B. The Essential matrix. Taking the scalar product of Eq. (11) with $\mathbf{X}'_{\mathbf{c}}$, we obtain:

$$\mathbf{X}'_{\mathbf{c}} \cdot (\mathbf{T} \times \mathbf{X}_{\mathbf{c}}) = \mathbf{X}'_{\mathbf{c}} \cdot (\mathbf{T} \times \mathbf{R}\mathbf{X}_{\mathbf{c}}') \quad (12)$$

$$\mathbf{X}'_{\mathbf{c}} \cdot (\mathbf{T} \times \mathbf{R}\mathbf{X}_{\mathbf{c}}) = 0 \quad (13)$$

A vector product can be expressed as a matrix multiplication:

$$\mathbf{T} \times \mathbf{X}_{\mathbf{c}} = \mathbf{T}_{\times} \mathbf{X}_{\mathbf{c}} \quad (14)$$

where

$$\mathbf{T}_{\times} = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \quad (15)$$

246 So equation Eq. (11) can be rewritten as:

$$\mathbf{X}'_{\mathbf{c}} \cdot (\mathbf{T}_{\times} \mathbf{R}\mathbf{X}_{\mathbf{c}}) = 0 \quad (16)$$

$$\mathbf{X}'_{\mathbf{c}} \cdot \mathbf{T}_{\times} \mathbf{R} \mathbf{X}_{\mathbf{c}} = 0 \quad (17)$$

where

$$\mathbf{E} = \mathbf{T}_{\times} \mathbf{R} \quad (18)$$

247 \mathbf{E} is a 3×3 matrix known as the *essential matrix*. The constraint also holds for rays \mathbf{p} , which are parallel to the camera-centered
 248 position vectors $\mathbf{X}_{\mathbf{c}}$:

$$\mathbf{p}'^T \mathbf{E} \mathbf{p} = 0 \quad (19)$$

This is the epipolar constraint. If a point \mathbf{p} is observed in one image, then its position \mathbf{p}' in the other image must lie on the line defined by Equation Eq. (19). The essential matrix can convert from pixels on the detector to rays \mathbf{p} in the world, assuming a calibrated camera (intrinsic properties are known), and pixel coordinates can then be converted to image plane coordinates using:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (20)$$

We can modify this to derive a relationship between pixel coordinates and rays:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{k_u}{f} & 0 & \frac{u_0}{f} \\ 0 & \frac{k_v}{f} & \frac{v_0}{f} \\ 0 & 0 & \frac{1}{f} \end{bmatrix} \begin{bmatrix} x \\ y \\ f \end{bmatrix} \quad (21)$$

$\tilde{\mathbf{K}}$ is defined as follows:

$$\tilde{\mathbf{K}} = \begin{bmatrix} fk_u & 0 & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (22)$$

then we can write pixel coordinates in homogenous coordinates:

$$\tilde{\mathbf{w}} = \tilde{\mathbf{K}} \mathbf{p} \quad (23)$$

C. The Fundamental matrix.

From Eq. (19) the epipolar constraint becomes

$$\tilde{\mathbf{w}}'^T \tilde{\mathbf{K}}^{-T} E \tilde{\mathbf{K}}^{-1} \tilde{\mathbf{w}} = 0 \quad (24)$$

$$\tilde{\mathbf{w}}'^T F \tilde{\mathbf{w}} = 0 \quad (25)$$

249 The (3×3) matrix \mathbf{F} , is the called the *fundamental matrix*. With intrinsically calibrated cameras, structure can be recovered
250 by triangulation. First, the two projection matrices are obtained via a SVD of the essential matrix. The SVD of the essential
251 matrix is given by:

$$\mathbf{E} = \mathbf{K}'^T \mathbf{F} \mathbf{K} = \mathbf{T}_\times \mathbf{R} = \mathbf{U} \Lambda \mathbf{V}^T \quad (26)$$

It can be shown that

$$\hat{\mathbf{T}}_\times = U \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U^T \quad (27)$$

and

$$\mathbf{R} = U \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} V^T \quad (28)$$

Then, aligning the left camera and world coordinate systems gives the projection matrices:

$$\mathbf{P} = \mathbf{K} [\mathbf{I} \mid \mathbf{0}] \quad (29)$$

and

$$\mathbf{P}' = \mathbf{K}' [\mathbf{R} \mid \mathbf{T}] \quad (30)$$

252 Where $[\mathbf{I} \mid \mathbf{0}]$ is the identity matrix augmented column-wise with a zero matrix, and the two projection matrices (\mathbf{P} and \mathbf{P}')
253 project from camera pixel coordinates to world coordinates. Given these projection matrices, scene structure can be recovered
254 (only up to scale, since only the magnitude of \mathbf{T} ($|\mathbf{T}|$) is unknown) using least squares fitting. Ambiguities in \mathbf{T} and \mathbf{R} are
255 resolved by ensuring that visible points lie in front of the two cameras. As with the essential matrix, the fundamental matrix
256 can be factorised into a skew-symmetric matrix corresponding to translation and a 3×3 non-singular matrix corresponding to
257 rotation.

258 The second approach is less prone to compound errors but relies on precise identification and tracking of fiducial markers.
259 distinction and tracking fiducials. Instead of calculating \mathbf{F} between neighbouring images, \mathbf{F} is calculated between the current
260 projection and the very first projection. \mathbf{F} is then decomposed and the transformation matrix is inverted and applied to the
261 back projected volume. The reoriented back projected volumes are summed and finally filtered to remove the additional spatial
262 frequencies imparted from rotating the sample.

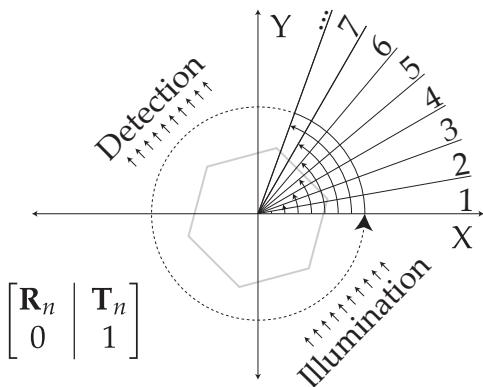
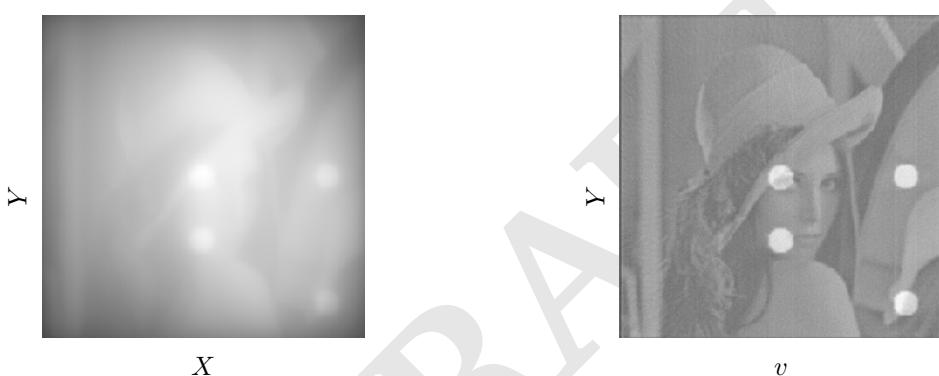


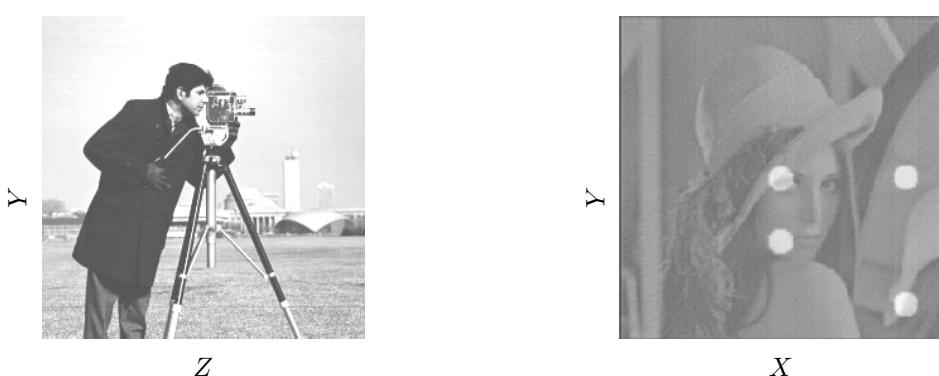
Fig. 9. Principles of the proposed algorithm. Each successive frame of OPT image data will have an associated \mathbf{R} and \mathbf{T} (shown here in augmented form using homogenous coordinates), these matrices can be recovered from comparing the fiducial marker positions in each frame (n) and its successor ($n + 1$).



(a) This is the unfiltered reconstruction of the object using the Radon transform

(b) Ram-lak (Fourier ramp) filter applied to Fig. 10a.

Fig. 10. The result of a tomographic reconstruction (using equally spaced angular steps and no translation between frames) requires Fourier filtering to normalise spatial contrast.



(a) Filtered reconstruction Cameraman testcard

(b) Filtered reconstruction of Lena test-card

Fig. 11. Filtered reconstruction of the ground truth reference image from Fig. 5 using the new proposed algorithm.