

Frame Localisation Optical Projection Tomography

Craig Russell^{1,2}, Pedro P Vallejo Ramirez^{2,✉}, and Eric Rees^{1,✉}

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

²European Bioinformatics Institute, Wellcome Genome Campus, Cambridge CB10 1SD

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

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

17 Sharpe *et. al* proposed OPT (?) using visible light to im-
18 age transparent or translucent mesoscopic samples, with mi-
19 crometer resolution. OPT addresses the scale gap between
20 photographic techniques (for samples typically larger than
21 10 mm), and light microscopy techniques (samples smaller
22 than 1 mm) to image biological samples in the 1 mm to
23 10 mm range.

24 OPT is based on computerised tomography techniques (?) in
25 which a set of projections of a specimen are acquired as the
26 specimen travels through a full rotation, shown in Fig. 1b.
27 Typically, a Radon transform is then used to transform this
28 set of images into a 3D image stack in Cartesian coordinates
29 (X, Y, Z). The Radon transform relies heavily on the assump-
30 tion of circular motion with constant angular steps about a
31 vertical axis. Prior to the Radon transform, an attempt is
32 made to find the centre of rotation (CORs) and correct the
33 image shift (? ? ?); this step is both computationally ex-
34 pensive, error prone and incomplete with regards to all avail-
35 able degrees of freedom. This work presents an improved
36 general reconstruction algorithm that is robust to spatial and
37 angular mechanical drifts during acquisitions, as well as to
38 inconsistent angular steps. The proposed algorithm trian-
39 gulates points between image pairs to extract camera pose using
40 the theoretical framework used in stereoscopic imaging.

41 Stereoscopic imaging

42 When the features or fiducial markers in one view are
43 uniquely identifiable, the stereoscopic imaging of scenes al-
44 lows for the triangulation of individual features in three di-
45 mensional space (known as world points), see Fig. 2 and 80

46 Fig. 1 for the coordinate system which describes this geom-
47 etry. Triangulation requires that each feature is detected in
48 both images of a stereo imaging system and for these de-
49 tections to be correctly associated with one another. This
50 is known as the correspondence problem. Various methods
51 exist to ensure that features are detected from image data
52 and accurately associated between two cameras or views (?)
53 and the properties of scale-independent features and their
54 surrounding pixel environment in one image can thus be
55 matched to a similar feature in a second image.

Coordinates in two adjacent views with a common epi-pole
56 (the vector connecting the O and O' , see Figure 2) are related
57 by the essential matrix (E) for uncalibrated cameras and the
58 fundamental matrix (F) for calibrated cameras. Their prop-
59 erties are described by:

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

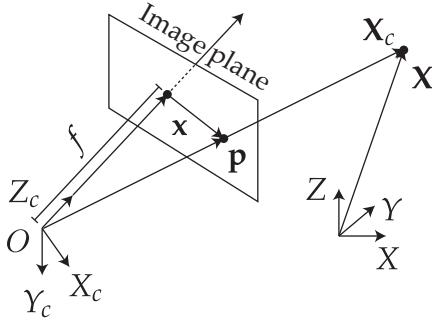
$$\mathbf{E} = \mathbf{K}'^T \mathbf{F} \mathbf{K} \quad (2)$$

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

61 The proposed algorithm (fIOPt)

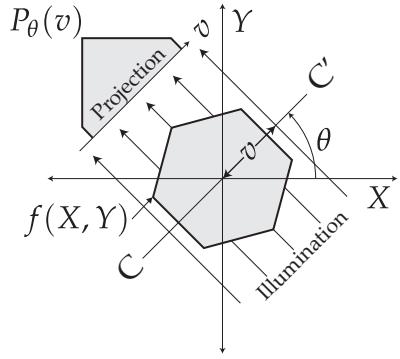
The motion of a rotating sample, as in an OPT acquisition,
62 with a transformation matrix ($[\mathbf{R} \mid \mathbf{T}]$) in view of a fixed
63 camera is analogous to the motion of a camera around the
64 scene with the inverse transformation matrix . During an
ideal OPT acquisition, a marker will appear to follow an el-
65 liptical path in the xy image plane. For the volume recon-
66 struction procedure, there is a fitting step to recover the path
67 of the fiducial marker, which is used to correct the sinogram
68 before applying the inverse Radon transform. This type of
69 reconstruction not only ignores any mechanical jitter of the
70 sample, but also any affine, systematic, mechanical drift (in
71 $X, Y, Z, \theta, \phi, \psi$). This can be rectified by recovering the com-
72 plete non-scaling transformation for every projection. Now,
73 using two adjacent images of a scene (separated by some ro-
74 ation and translation) world points in 3D space may be trian-
75 gulated within the scene given the rotational and translational
76 matrices of the respective camera views.

Once a sufficient amount of fiducial markers are reliably
77 tracked from the first to the second image, either of the fun-
78 damental or essential matrices can be computed. Using the
79 factorisation of one of these matrices, between each adjacent



(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.

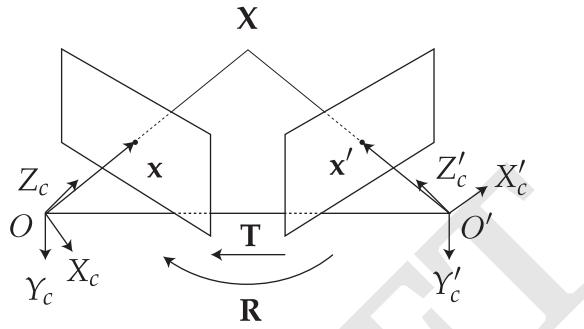


Fig. 2. Epi-polar geometry described for two adjacent views (or cameras of a scene). Coordinates as expressed in Fig. 1a with prime notation ('') denoting the additional right camera view. Transforming from right to left camera-centered coordinates (\mathbf{X}'_c to \mathbf{X}_c) requires a rotation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

Algorithm 1 The proposed flOPT algorithm, where: *localiseFiducials* generates a 2D coordinate list from an image containing punctate fiducials. *findHomography* generates \mathbf{F}_n between a pair of 2D coordinate lists. *chooseBestHomography* is a function that chooses the most likely homography matrix \mathbf{F}_n but minimising the least squares difference between \mathbf{F}_{n-1} and the four candidate $\mathbf{F}_{n,\{0,1,2,3\}}$ matrices. *decomposeHomography* factorises \mathbf{F}_n into affine rotation and translation matrices (\mathbf{R}_n and \mathbf{T}_n). *affineHomogenise* merges the separate affine matrices into a more manageable homogenous matrix. *spatialFilter* finally corrects the image using standard computer tomography filters.

```

 $n \leftarrow 1$ 
 $\mathbf{x} \leftarrow \text{localiseFiducials}(I_0)$ 
while  $n \leq k$  do
     $\mathbf{x}' \leftarrow \text{localiseBeads}(I_n)$ 
     $\mathbf{F}_{n,\{0,1,2,3\}} \leftarrow \text{findHomography}(\mathbf{x}, \mathbf{x}')$ 
     $\mathbf{F}_n \leftarrow \text{chooseBestHomography}(\mathbf{F}_{n,\{0,1,2,3\}}, \mathbf{F}_{n-1})$ 
     $\mathbf{R}_n, \mathbf{T}_n \leftarrow \text{decomposeHomography}(\mathbf{F}_n)$ 
     $[\mathbf{R}_n | \mathbf{T}_n] \leftarrow \text{affineHomogenise}(\mathbf{R}_n, \mathbf{T}_n)$ 
     $I_n^* \leftarrow [\mathbf{R}_n | \mathbf{T}_n]^{-1} \cdot I_n$ 
     $I_{\text{unfiltered}} \leftarrow \sum_{n=0}^k I_n^*$ 
     $n \leftarrow n$ 
end while
 $I_{\text{final}} \leftarrow \text{spatialFilter}(I_{\text{unfiltered}})$ 
return  $I_{\text{final}}$ 
```

view of a rotating scene, the translation and rotational matrices can be recovered.

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

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

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

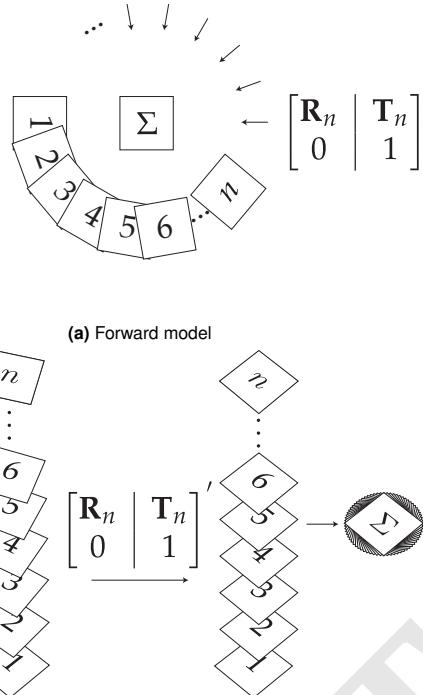


Fig. 3. The simulation of OPT data incorporating rotational and translational offsets, and the proposed reconstruction algorithm. (a): The n projections of the object (Σ), at rotation (\mathbf{R}_1 to \mathbf{R}_n) and translation (\mathbf{T}_1 to \mathbf{T}_n), produces n frames of image data. During 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.

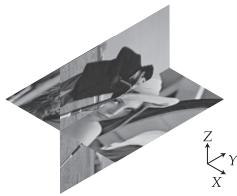


Fig. 4. Ground truth 3D object for reconstruction, based on the Cameraman and Lena testcard images.

likely angle of rotation of the system's imaging properties.

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 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.

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

Where \mathbf{Y} is the vector of observed values and \hat{Y}_i is mean of the i th value of the predicted values. The more challenging case of a sample drifting systematically along the X axis, with a constant velocity, was then considered. This drift produced a helical path of a single fiducial within the sample, see Fig. 7b. In Fig. 7c, the Radon transform fails to produce a recognisable reproduction of the test image with the addition of a slight helicity to the rotation. The proposed algorithm produces an equivalent result to that of a sample rotating without any systematic drift, see Fig. 10b. In Fig. 6c the respective reconstructions from each

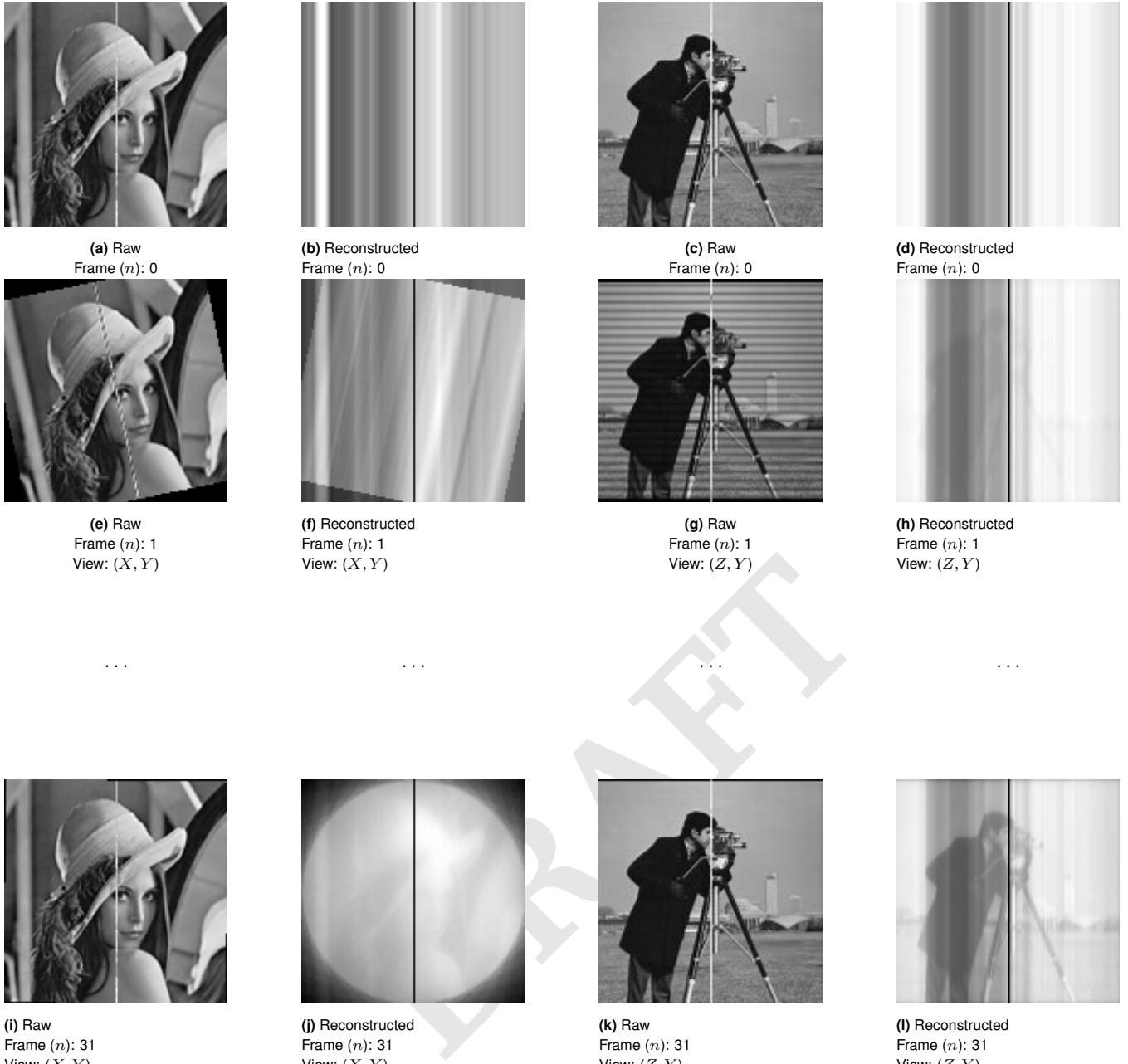


Fig. 5. A 3D test-volume of two orthogonal and different testcard images, from Fig. 4, was used to verify the reconstructive capabilities of the proposed algorithm. The projected image data (b), (f), (j) and (d), (h), (l), were used to iteratively generate reconstructions where the n^{th} reconstruction 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.

algorithm were compared, as before, while the helical shift was incremented. See Fig. 7b for a sinogram of a sample wherein a helical shift has been induced. When using correlation as a metric of reproduction quality, the new algorithm fares slightly worse at zero helicity, with 94 % correlation compared to the Radon transform at 96 %. As expected, the Radon transform rapidly deteriorates once a systematic drift is applied, whereas the new algorithm maintains the quality of the reconstruction, see Fig. 6c.

Recovery of \mathbf{R} and \mathbf{T} using matrix decomposition. To quantitatively verify that the matrix decomposition technique was valid and robust, the accuracy of the reproduction of \mathbf{R} and \mathbf{T} was tested directly. The original \mathbf{R} and \mathbf{T} matrices

were computed and compared to \mathbf{R} and \mathbf{T} generated from matrix decomposition. This absolute difference was computed element-wise in each matrix and then an average for each matrix was taken. Overall, the worst-case scenario produced a percentage error of 2 % (see Fig. 8 for full statistics). The accuracy of the calculated \mathbf{R} and \mathbf{T} deteriorated when adding in additional degrees of combined movement, but with no correlation between the degree of helicity and the error produced. The translation matrix (\mathbf{T}) was consistently more accurately reproduced, which is likely due to it having fewer available degrees of freedom.

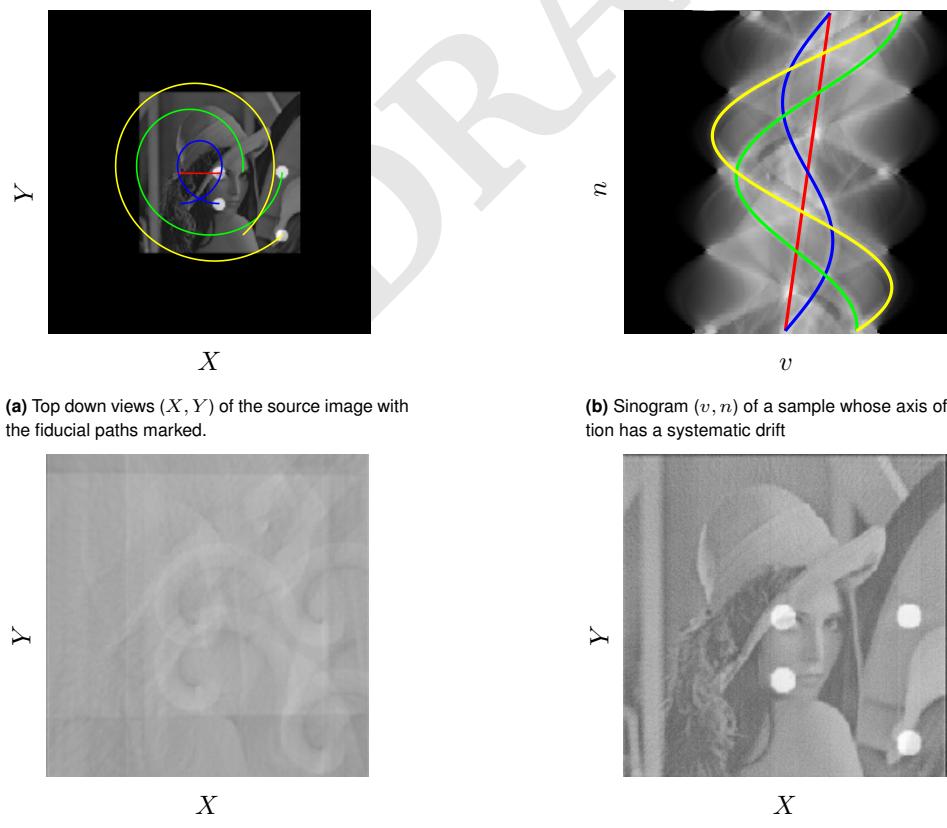
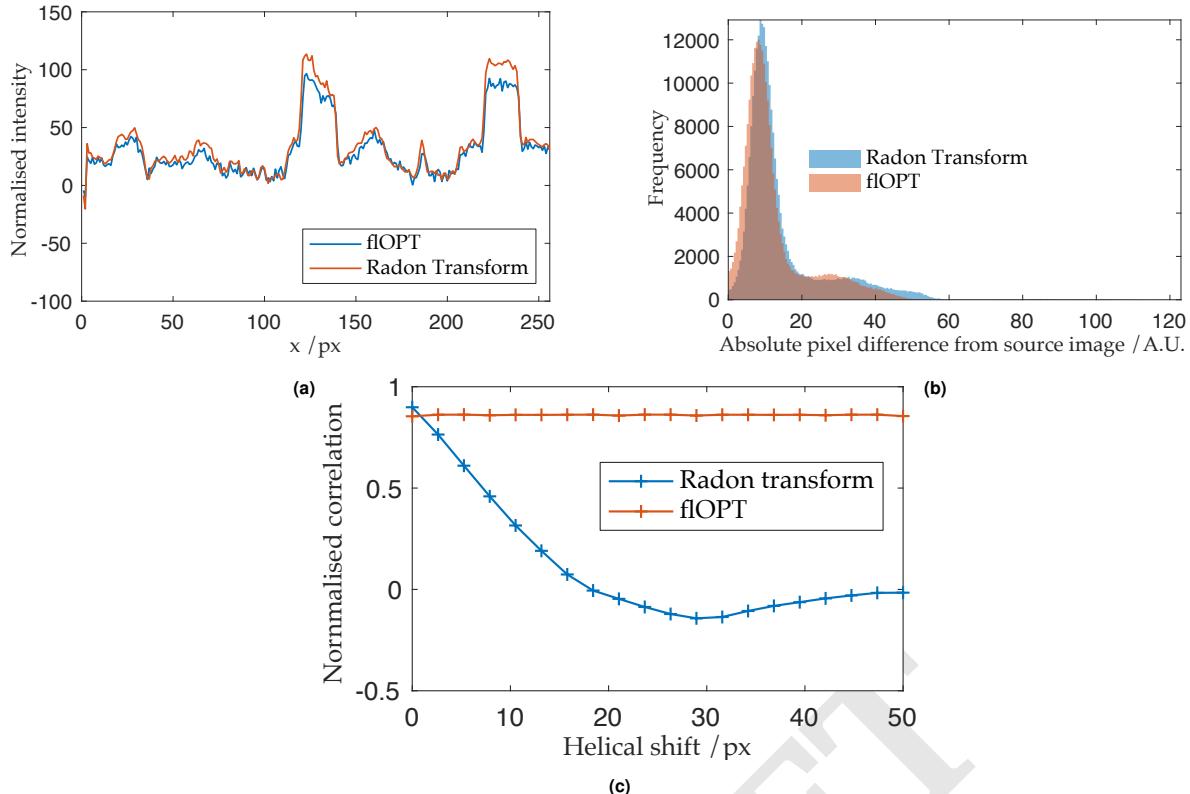


Fig. 7. Comparison of the two reconstructions under sample imaging with a systematic drift, in 3D though represented here in 2D.

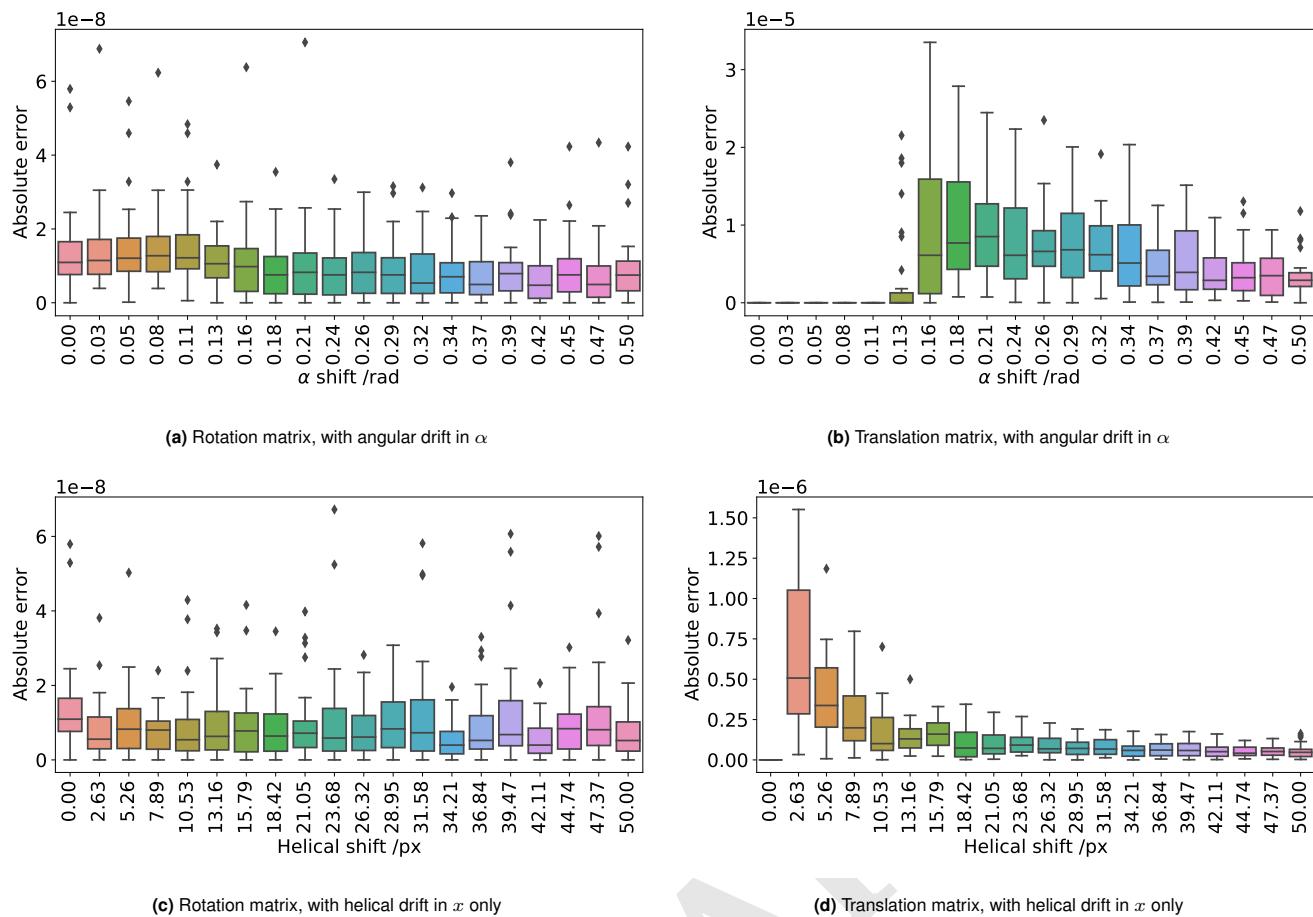


Fig. 8. Box plots demonstrating that the rotational and translations matrices can be recovered accurately from fiducial marker positions. Panels (a) and (b) introduce an angular drift during rotation, to an observer at the detector this would appear as a tilt of the sample towards them, causing precession. Panels (c) and (d) introduce a lateral drift in X causing a helical path to be drawn out. In all cases, the percentage error introduced by the the addition of undesirable additional movements was on the order of <2 %.¹

Discussion

A new algorithm for reconstructing OPT data has been demonstrated. The new algorithm uses multiple fiducial markers to recover the matrix which describes the rotation and translation of the sample. The quality of the reconstructions shows a slight improvement when compared to the standard Radon transform, with a great effect when a systematic drift is introduced. The accuracy of the decomposition of \mathbf{F} into \mathbf{R} and \mathbf{T} was compared to the ground truth matrices. The element-wise absolute difference $\left(\frac{|x-y|}{2(x+y)} \right)$ of each matrix was averaged across the matrix for \mathbf{R} and \mathbf{T} . In the worst-case scenario, a maximum of 2 % average absolute difference was found between ground truth and recovered matrices, suggesting that the technique is robust to various forms of drift in all dimensions and general instability. Such an algorithm could be used to minimise ghosting effects seen in real samples, particularly in samples where slipping is likely to occur, such as in gels or in cheaper OPT systems which tend to be more mechanically unstable and imprecise. In particular the imaging of large mobile gels is set to become more prevalent given the surge of new techniques in Expansion Microscopy (?), whereby fragile expanded samples embedded

in thin lubricious gels.

Future work

The proposed algorithm relies on triangulation between two view points. However, it is possible to use three separate views to reconstruct a scene, one such approach being quaternion tensors (?). Working with tensors is more complex, but a future iteration of the algorithm presented here may benefit from using three views to provide a more accurate transformation matrix. Beyond three views, there is currently no mathematical framework for four or more views. If such tools were to be developed, it may be possible to have the algorithm described above be a non-iterative, single-shot reconstruction from pixels to voxels.

Fiducial markers could also be extracted from the image texture alone, circumventing the need for the additional beads embedded in the sample. To find such correspondences, points with similar local texture are found and matched in between each image using standard algorithms such as SIFT (?) and RANSAC (?). This was attempted in this work, however, the errors introduced into the transformation matrices make this approach currently unviable; and so by requiring

206 bright punctuate fiducial markers the burden of collecting the
207 fiducial coordinates is shifted to well established curve fitting
208 algorithms that are robust to noise.

209 **Code availability**

210 All of the code presented here is FOSS using OpenCV and
211 Python for our simulation, and can be found on GitHub (?).

212 **Disclosures**

213 The authors declare that there are no conflicts of interest re-
214 lated to this article.

215 **Bibliography**

DRAFT

216 **Supplementary Note 1: Supplemental**

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

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

223 Where $P_\theta(v)$ is the Radon transform of $f(X, Y)$ which represents the contrast image of a 2D slice of the specimen. The Radon
 224 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. The back projection step is given by:

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

225 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
 226 during back-projection to avoid spatial frequency oversampling during the object's rotation (see Fig. 10b) a high pass filter is
 227 commonly used to compensate for the perceived blurring. The blurring arises as Q_θ is back-projected (smeared) across the
 228 image plane for each angle of reconstruction; which means that not only does the back-projection contribute at the line it is
 229 intended to (along line C in Fig. 1b), but all other points along the back-projecting ray.

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

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

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

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

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

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

241 The **baseline** is defined as the line joining the optical centres. An **epipole** is the point of intersection of the baseline with
 242 the image plane and there are two epipoles per feature, one for each camera. An **epipolar line** is a line of intersection of the
 243 epipolar 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
 244 point (\mathbf{X}). For different world points, the epipolar plane rotates about the baseline. All epipolar lines intersect the epipole.

245 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
 246 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
 247 epipolar line. The two camera-centered coordinate systems \mathbf{X}'_c and \mathbf{X}_c are related by a rotation, \mathbf{R} and translation, \mathbf{T} (see in
 248 Fig. 2) as follows:

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

B. The Essential matrix.

Taking the scalar product of Eq. (11) with \mathbf{X}'_c , we obtain:

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

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

A vector product can be expressed as a matrix multiplication:

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

where

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

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

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

$$\mathbf{X}'_c \mathbf{T} \mathbf{R} \mathbf{X}_c = 0 \quad (17)$$

where

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

250 \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
 251 position vectors \mathbf{X}_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

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

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

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

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

It can be shown that

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

and

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

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

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

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}') project from camera pixel coordinates to world coordinates. Given these projection matrices, scene structure can be recovered (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 resolved by ensuring that visible points lie in front of the two cameras. As with the essential matrix, the fundamental matrix can be factorised into a skew-symmetric matrix corresponding to translation and a 3×3 non-singular matrix corresponding to rotation.

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

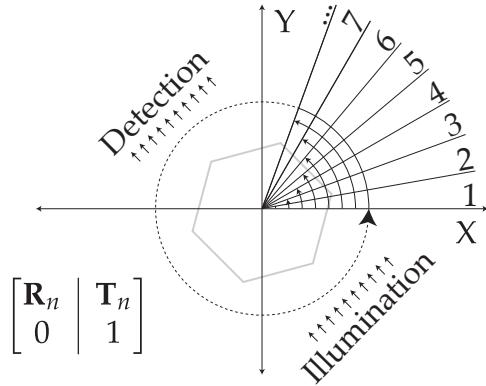
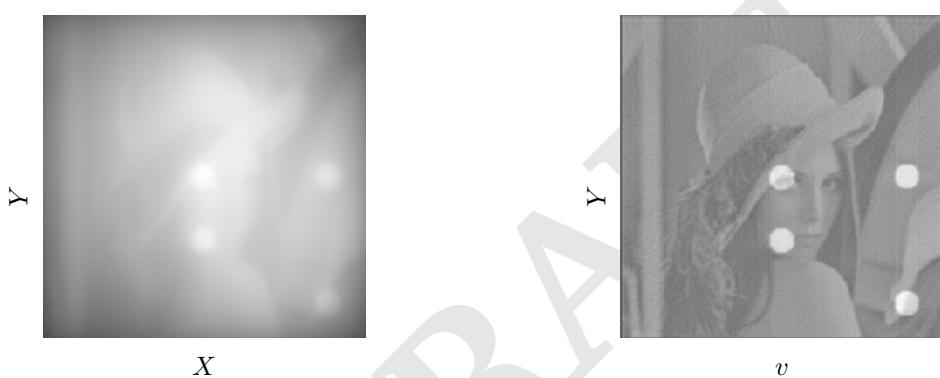


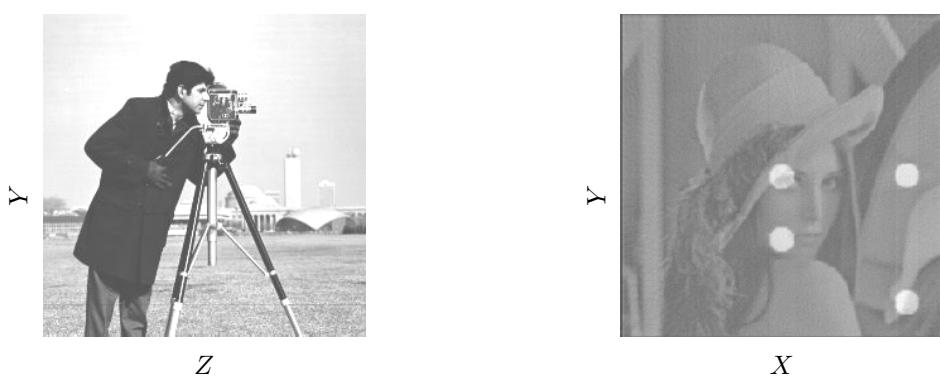
Fig. 9. Principles of the proposed algorithm. Each successive frame of OPT image data will have an associated \mathbf{R} and \mathbf{T} (shown here in augmented form using homogenous coordinates), these matrices can be recovered from comparing the fiducial marker positions in each frame (n) and its successor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



(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 of the Cameraman testcard

(b) Filtered reconstruction of the Lena testcard

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