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

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

We present a tomographic reconstruction algorithm, which is applied to Optical Projection Tomography (OPT) [1] images, that is robust to mechanical jitter and systematic angular and spatial drift. OPT relies on precise mechanical rotation and is less mechanically stable than large-scale CT scanning systems, leading to reconstruction artefacts. The algorithm uses multiple (5+) tracked fiducial beads to recover the sample pose and the image rays are then back-projected at each orientation. The quality of the image reconstruction using the proposed algorithm shows an improvement when compared to the Radon transform. Moreover, when adding a systematic spatial and angular mechanical drift, the reconstruction shows a significant improvement over the Radon transform.

Sharpe *et. al* proposed Optical Projection Tomography (OPT) (1) using visible light to image transparent or translucent mesoscopic samples, with micrometer resolution. OPT addresses the scale gap between photographic techniques (for samples typically larger than $10\,\mathrm{mm}$), and light microscopy techniques (samples smaller than $1\,\mathrm{mm}$) to image biological samples in the $1\,\mathrm{mm}$ to $10\,\mathrm{mm}$ range.
OPT is based on computerised tomography techniques (2) in which a set of projections of a specimen are imaged as the specimen travels through a full rotation. Typically, a Radon transform is then used to transform this set of images into a 3D image stack in Cartesian coordinates (X,Y,Z). The Radon transform relies heavily on the assumption of circular motion with constant angular steps about a vertical axis. This work presents an improved reconstruction algorithm that is robust to spatial and angular mechanical drifts during acquisitions, as well as to inconsistent angular steps. The proposed algorithm triangulates points between image pairs to extract camera pose using the theoretical framework used in stereoscopic imaging.

Stereoscopic imaging

When the features or fiducial markers in one view are uniquely identifiable, the stereoscopic imaging of scenes allows for the triangulation of individual features in three dimensional space (known as world points), see Figure 2 for the coordinate system which describes this geometry. Triangulation requires that each feature is detected in both images of a stereo imaging system and for these detections to be correctly associated with one another. This is known as the correspondence problem. Various methods exist to ensure that features are detected from image data and accurately associated between two cameras or views (3) and the properties of scale-independent features and their surrounding pixel environment in one image can thus be matched to a similar feature in a second image.
Coordinates in two adjacent views with a common epi-pole (see Figure 2) are related by the essential matrix ($\mathbf{E}$) for uncalibrated cameras and the fundamental matrix ($\mathbf{F}$) for calibrated cameras. Their properties are described by:

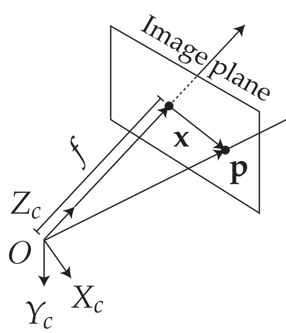
$$\mathbf{p'}^T\mathbf{E}\mathbf{p} = 0 \tag{1}$$

$$\mathbf{E} = \mathbf{K'}^T\mathbf{F}\mathbf{K} \tag{2}$$

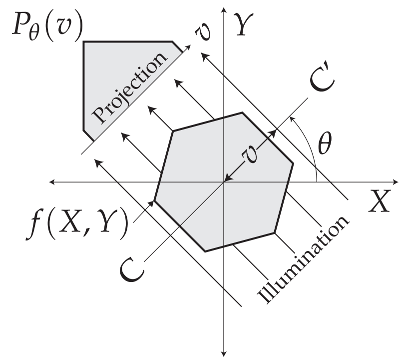
Where $\mathbf{K}$ is a matrix that converts image plane coordinates to camera pixel coordinates and where $\mathbf{p}$ refers to a point in the image plane.

The proposed algorithm

The motion of a rotating sample, as in an OPT acquisition, with a transformation matrix ($\left[\mathbf{R} \mid \mathbf{T}\right]$) in view of a fixed camera is analogous to the motion of a camera with the inverse transformation matrix around the scene. During an ideal OPT acquisition, a marker will appear to follow an elliptical path in the xy image plane. For the volume reconstruction procedure, there is a fitting step to recover the path of the fiducial marker, which is used to correct the sinogram before applying the inverse Radon transform. This type of reconstruction not only ignores any mechanical jitter of the sample, but also any affine, systematic, mechanical drift (in X,Y,Z,θ,ϕ,ψ). Using two adjacent images of a scene, separated by some rotation and translation, world points in 3D space may be triangulated within the scene given the rotational and translational matrices of the respective camera views.
Once a sufficient amount of fiducial markers are reliably tracked from the first to the second image, either of the fundamental or essential matrices can be computed. Using the factorisation of one of these matrices, between each adjacent 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



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



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

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.

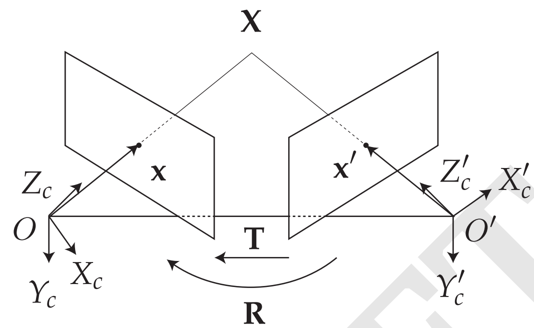


Fig. 2. Epi-polar geometry described for two adjacent views (or cameras of a scene). Coordinates as expressed in Fig. 1a with prime notation ($'$) denoting the additional right camera view. Transforming from right to left camera-centered coordinates ($\mathbf{X_c'}$ to $\mathbf{X_c}$) requires a rotation ($\mathbf{R}$) 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:

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

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.

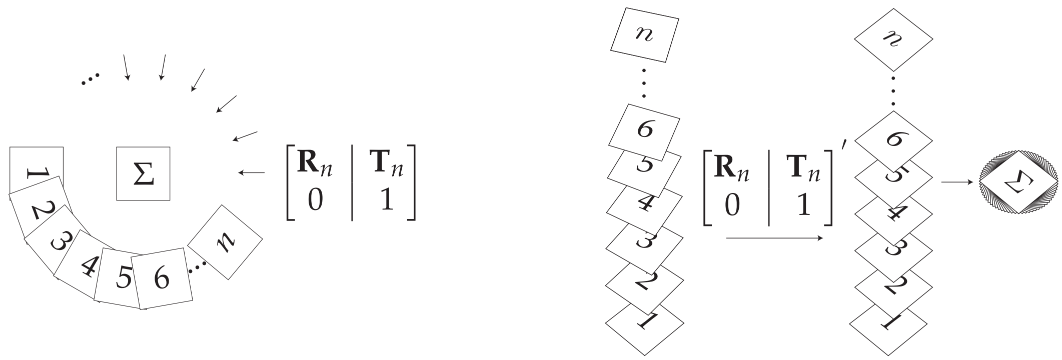
Verification of the proposed algorithm

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.

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

The more challenging case of a sample drifting systemati-



(a) Forward model

(b) Reconstruction method, solving the inverse problem.

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.

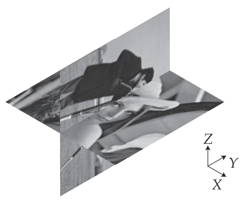


Fig. 4. Ground truth 3D object for reconstruction, based on the Cameraman and Lena testcard images.

cally 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 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 R and 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.

Discussion

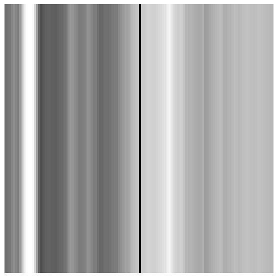
A new algorithm for reconstructing OPT data has been demonstrated. The new algorithm uses multiple fiducial markers to recover the matrix which describes the rotation and translation of the sample. The quality of the reconstructions shows a slight improvement when compared to the standard Radon transform, 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 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.

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 (4). 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.

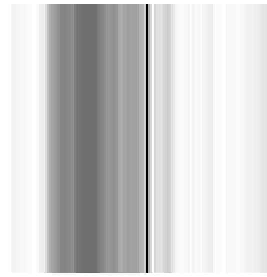
Fiducial markers could also be extracted from the image texture alone, circumventing the need for for the additional beads embedded in the sample. To find such corre-

(a) Raw
Frame (n): 0



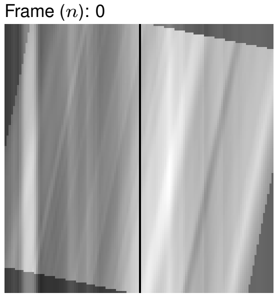
(b) Reconstructed
Frame (n): 0

(c) Raw
Frame (n): 0

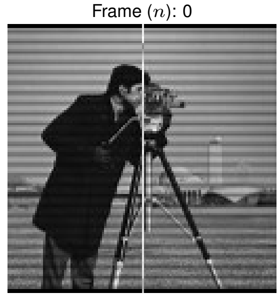


(d) Reconstructed
Frame (n): 0

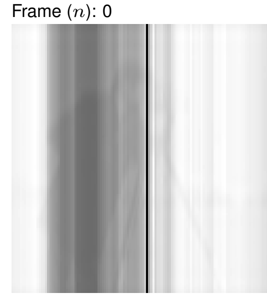
(e) Raw
Frame (n): 1
View: (X, Y)



(f) Reconstructed
Frame (n): 1
View: (X, Y)



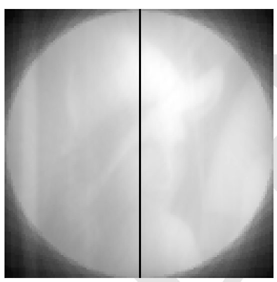
(g) Raw
Frame (n): 1
View: (Z, Y)



(h) Reconstructed
Frame (n): 1
View: (Z, Y)

. . . .

(i) Raw
Frame (n): 31
View: (X, Y)



(j) Reconstructed
Frame (n): 31
View: (X, Y)

(k) Raw
Frame (n): 31
View: (Z, Y)

(l) Reconstructed
Frame (n): 31
View: (Z, Y)

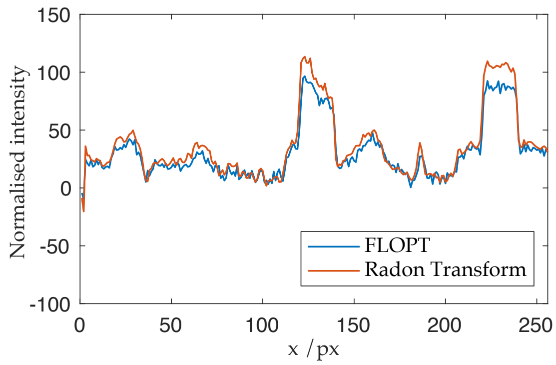
Fig. 5. A 3D test-volume of two orthogonal and different testcard images, from Fig. 4, was used to verify the reconstructive capabilities of the proposed algorithm. The projected image data (b), (h), (j) and (d), (h), (l), were used to iteratively generate reconstructions where the n^{th} reconstruction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

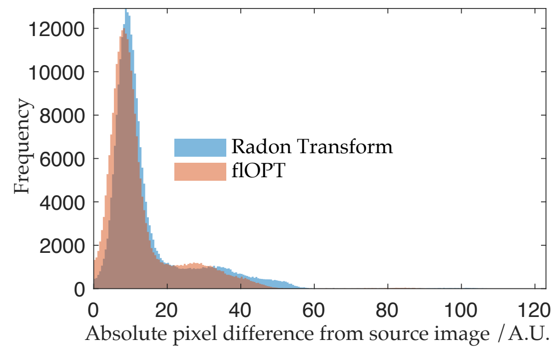
ACKNOWLEDGEMENTS

Bibliography

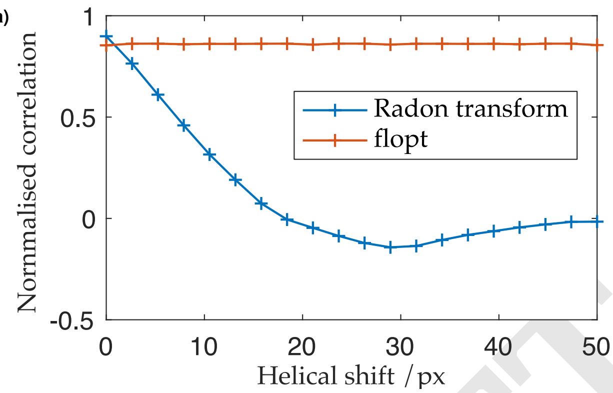
1. James Sharpe, Ulf Ahlgren, Paul Perry, Bill Hill, Allyson Ross, Jacob Hecksher-Sørensen, Richard Baldock, and Duncan Davidson. Optical Projection Tomography as a Tool for 3D Microscopy and Gene Expression Studies. 296(5567):541–545. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1068206.
2. A. Kak 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.
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.



(a)

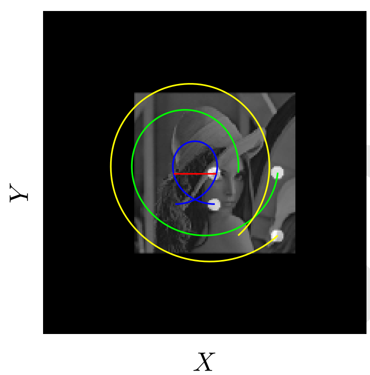


(b)

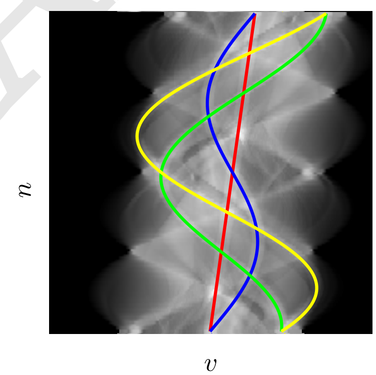


(c)

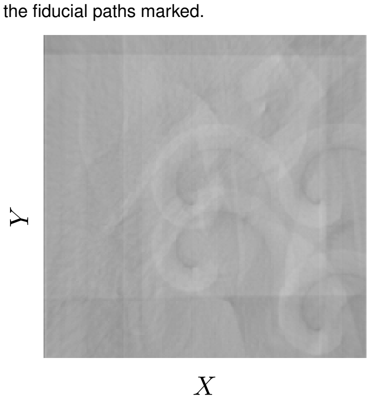
Fig. 6. (a), Line profile comparison of the reconstruction of a reference image computationally rotated, projected and reconstructed using the standard Radon transform and the new proposed algorithm. (b), Histogram of of pixel values compared between reconstructions using the new proposed flOPT algorithm 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.



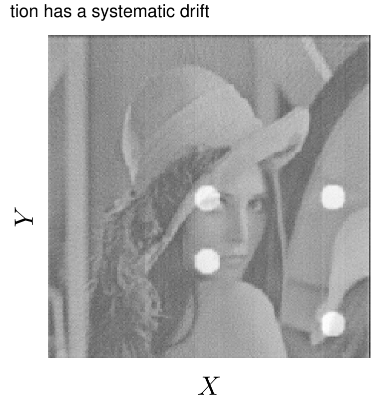
(a) Top down views (X, Y) of the source image with the fiducial paths marked.



(b) Sinogram (v, n) of a sample whose axis of rotation has a systematic drift

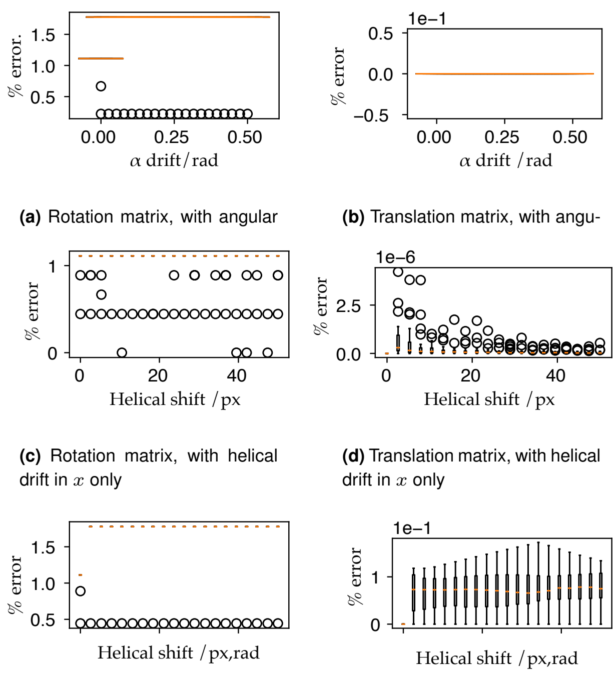


(c) Unfiltered reconstruction using a Radon transform.



(d) Filtered reconstruction using the new algorithm.

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



(a) Rotation matrix, with angular

(b) Translation matrix, with angu-

(c) Rotation matrix, with helical drift in x only

(d) Translation matrix, with helical drift in x only

(e) Rotation matrix, with angular drift in α and helical drift in x

(f) Translation matrix, with angular drift in α and helical drift in X

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 $<2\,\%$.

Supplementary Note 1: Supplemental

A. Reconstruction. As the sample is rotated each detector pixel collects an intensity $I(\theta) = I_n e^{-k(\theta)}$ at discrete (n) angles through a full rotation of the sample; where I_n is the unattenuated radiation intensity from the source to the detector, k is the attenuation caused by the sample along a detected ray an $I(n)$ is the measured intensity, see Fig. 1b. Rays from the sample to the detector approximate straight lines, and so the the rays reaching the detector with a line integrals. A projection 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 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:

$$X\cos(\theta) + Y\sin(\theta) - v = 0 \tag{5}$$

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

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

Where $P_\theta(v)$ is the Radon transform of $f(X,Y)$ which represents the contrast image of Two Dimensional (2D) slice of the specimen. The Radon 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.

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

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 back-projection to avoid spatial frequency oversampling during the object's rotation (see Fig. 10b) a high pass filter is commonly used to compensate for the perceived blurring. The blurring arises as Q_θ is back-projected (smeared) across the image plane for each angle of reconstruction; which means that not only does the back-projection contribute at the line it is intended to (along line C in Fig. 1a), but all other points along the back-projecting ray.
Now, suppose we know the relative positions of the two cameras and their respective intrinsic parameters, such as magnification and pixel offset. For a single camera and given the camera parameters, we can translate pixel coordinates, $\mathbf{w} = (u,v)$, into the coplanar image plane coordinates $\mathbf{x} = (x,y)$:

$$u = u_0 + k_u x \tag{8}$$
$$v = v_0 + k_v y \tag{9}$$

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

$$\mathbf{p} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{10}$$

From the definition of a world point, as observed through an image, we can construct a dual-view model of world points in space as in Fig. 2. Using a model of a system with two views allows for the triangulation of rays based on image correspondences, 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 for correspondences, they constrain the search to a 1D line in each image. To derive general epipolar constraints, one should consider the epipolar geometry of two cameras as seen in Fig. 2
The **baseline** is defined as the line joining the optical centres. An **epipole** is the point of intersection of the baseline with the image plan and there are two epipoles per feature, one for each camera. An **epipolar line** is a line of intersection of the 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 point ($\mathbf{X}$). For different world points, the epipolar plane rotates about the baseline. All epipolar lines intersect the epipole. 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 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 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 Fig. 2) as follows:

$$\mathbf{X_c'} = \mathbf{R}\mathbf{X_c'} + \mathbf{T} \tag{11}$$

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

$$\mathbf{X_c'} \cdot (\mathbf{T} \times \mathbf{X_c}) = \mathbf{X_c'} \cdot (\mathbf{T} \times \mathbf{R}\mathbf{X_c'}) \tag{12}$$

$$\mathbf{X_c'} \cdot (\mathbf{T} \times \mathbf{R}\mathbf{X_c}) = 0 \tag{13}$$

A vector product can be expressed as a matrix multiplication:

$$\mathbf{T} \times \mathbf{X_c} = \mathbf{T}_\times \mathbf{X_c} \tag{14}$$

where

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

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

$$\mathbf{X_c'} \cdot (\mathbf{T}_\times \mathbf{R}\mathbf{X_c}) = 0 \tag{16}$$

$$\mathbf{X_c'}\mathbf{T}\mathbf{E}\mathbf{X_c} = 0 \tag{17}$$

where

$$\mathbf{E} = \mathbf{T}_\times \mathbf{R} \tag{18}$$

$\mathbf{E}$ 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 position vectors $\mathbf{X_c}$:

$$\mathbf{p'}^T \mathbf{E}\mathbf{p} = 0 \tag{19}$$

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:

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

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

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

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

$$\widetilde{\mathbf{K}} = \begin{bmatrix} fk_u & 0 & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{22}$$

then we can write pixel coordinates in homogenous coordinates:

$$\widetilde{\mathbf{w}} = \widetilde{\mathbf{K}}\mathbf{p} \tag{23}$$

C. The Fundamental matrix.

From Eq. (19) the epipolar constraint becomes

$$\widetilde{\mathbf{w}'}^T \widetilde{\mathbf{K}}^{-T} E \widetilde{\mathbf{K}}^{-1} \widetilde{\mathbf{w}} = 0 \tag{24}$$

$$\widetilde{\mathbf{w}'}^T F \widetilde{\mathbf{w}} = 0 \tag{25}$$

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 Singular Value Decomposition (SVD) of the essential matrix. The SVD of the essential matrix is given by:

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

It can be shown that

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

and

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

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

$$\mathbf{P} = \mathbf{K} \left[\mathbf{I} \mid \mathbf{0} \right] \tag{29}$$

and

$$\mathbf{P}' = \mathbf{K}' \left[\mathbf{R} \mid \mathbf{T} \right] \tag{30}$$

Where $\left[\mathbf{I} \mid \mathbf{0} \right]$ 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.

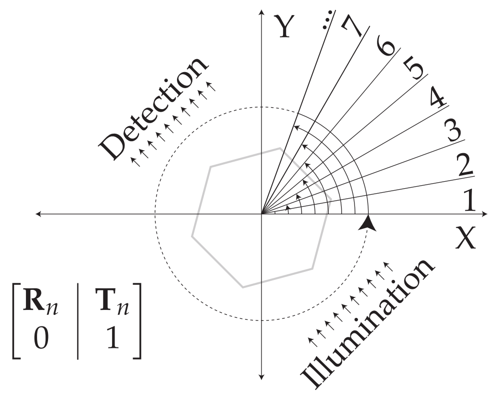
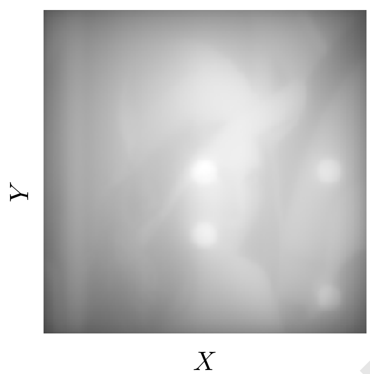
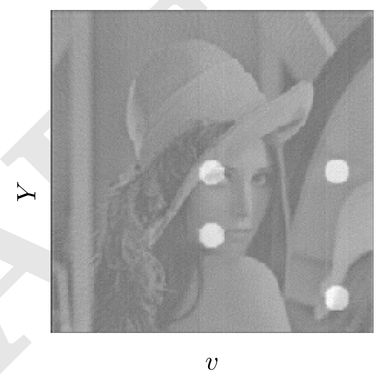


Fig. 9. Principles of the proposed algorithm. Each successive frame of OPT image data will have an associated $\mathbf{R}$ and $\mathbf{T}$ (shown here in augmented form using homogenous coordinates), these matrices can be recovered from comparing the fiducial marker positions in each frame (n) and its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

(b) Filtered reconstruction of Lena testcard

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