

Imaging on the Edge: Mapping Object Corners and Edges with Stereo X-ray Tomography

Zhenduo Shang ¹, Thomas Blumensath ^{2*}

¹ Affiliation 1; zs4n17@soton.ac.uk

² Affiliation 2; thomas.blumensath@soton.ac.uk

* Correspondence: thomas.blumensath@soton.ac.uk

Abstract: X-ray computed tomography (XCT) is a powerful tool for volumetric imaging, where three-dimensional (3D) images are generated from a large number of individual X-ray projection images. Collecting the required number of low noise projection images is however time-consuming and so the technique is not currently applicable when spatial information needs to be collected with high temporal resolution, such as in the study of dynamic processes. In our previous work, inspired by stereo vision, we developed stereo X-ray imaging methods that operate with only two X-ray projection images. Previously, we have shown how this allowed us to map point and line fiducial markers into 3D space at significantly faster temporal resolutions. In this paper, we make two further contributions. Firstly, instead of utilising internal fiducial markers, we demonstrate the applicability of the method to the 3D mapping of sharp object corners, a problem of interest in measuring the deformation of manufactured components under different loads. Furthermore, we demonstrate how the approach can be applied to real stereo X-ray data, even in settings where we do not have the annotated real training data that was required for the training of our previous Machine Learning approach. This is achieved by substituting the real data with a relatively simple synthetic training dataset designed to mimic key aspects of the real data.

Keywords: feature detection; 3D mapping; X-ray image reconstruction

1. Introduction

X-ray Computed Tomography (XCT) is an established volumetric imaging technique used throughout medical, scientific and industrial applications. However, in order to generate volumetric images of high spatial resolution and with limited artefacts, the method requires a very large number of individual X-ray measurements to be collected from around the object. Whilst the use of advanced algorithms, such as those based on machine learning or regularised optimisation (e.g using Total Variation (TV) constraints) offers the ability to reduce the number of required measurements somewhat [1–8], a significant reduction in the number of measurements is still not possible for most objects without sacrificing image quality.

We recently demonstrated a different approach [9]. Instead of trying to reconstruct full tomographic images from limited observations, which requires very strong prior information, we instead developed a novel stereo X-ray imaging approach that only recovered the 3D location of simple features, such as points and lines. The advantage of our new method is that it only requires two (stereo) projection images, but works without the strong constraints that are imposed in full image reconstruction from limited measurements. Our approach thus worked for general objects, as long as these included simple linear fiducial

Received:

Revised:

Accepted:

Published:

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Tomography* **2025**, *1*, 0. <https://doi.org/>

Copyright: © 2025 by the authors.

Submitted to *Tomography* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

markers. This approach is of particular interest in time-sensitive applications, where the internal structure of an object changes rapidly, and where we might only be able to take a single pair of images at each time step.

In this paper, we extend the above approach from fiducial markers to another commonly found set of linear features, namely the corners and edges of objects. The location of these point and line features is often more complex to identify. There are several applications where this might be of interest, but where fiducial markers cannot be embedded into the object. Of particular interest to us are applications where we want to study a manufactured component with sharp corners and edges whilst they are undergoing rapid deformations.

As in our previous work, we assume an imaging setup with two or more X-ray sources and detectors, providing an X-ray stereo vision system. Using ideas from computer stereo vision, spatial mapping of object corners and edges then becomes possible at the speed of the detector frame rate, which is orders of magnitude faster than full computed tomography data acquisition.

In contrast to visible light imaging, where the light registered at a specific location on a camera's image sensor is commonly associated with light reflecting from a single 3D point on the surface of an opaque object, X-ray intensity measured on an X-ray detector contain X-ray attenuation information from an entire line through the object [10]. Whilst the main challenge in traditional stereo vision lies in accurately aligning points between the two images forming a stereo pair [11], for X-ray stereo imaging, not only does this matching step become more difficult, an additional challenge is the identification of the location of distinct 3D points in the 2D projected images [9].

1.1. Our method

Our previous work [9] has identified an approach that 1) identifies all point-like features in the two X-ray views, and 2) matches these features between the views. Once identified and matched, mapping the features into 3D space then employed the same mathematical theory of projective geometry used in traditional stereo vision. For limit computational complexity and allow for efficient scaling to realistically sized X-ray projections, we use a block-based deep learning approach to identify the projected locations of 3D features in 2D space (that is, we identify the features in the projection images). These identified locations are then mapped into 3D space using the filtered back-projection (FBP) method [12,13]. Potential 3D feature locations can then be identified as those points where individual back-projected features overlap between the two views in 3D space. Whilst this is unique with a high probability if the features are sparse and randomly distributed in space, to further enhance the robustness against the exact localisation of features on the two imaging planes, we here employ a second deep neural network to process the back-projected volumetric image to identify feature locations.

1.2. Contributions

Our previous approach is here extended to the mapping of corner and edge features. Whilst these features are conceptually similar to point and line fiducial markers, the difference is that they are more difficult to identify in X-ray projection images. Whilst using a fiducial marker with a higher X-ray attenuation value will produce projection images with discontinuous image intensities at the fiducial locations, this is not true for edges and corners, where the image intensity in the projected images changes smoothly at the feature locations. Therefore, detecting and matching in this case can pose greater difficulty. Not only do we demonstrate in this paper that our previous approach still works in this setting, we here make a second key contribution. In our previous paper, we used a machine

learning method for feature identification that required training. In real applications, we do however seldom have sufficient amounts of real training data to train a model for a specific imaging task. To overcome this challenge, we here demonstrate that model training can also be done on simplified simulated data matched specifically to a given real imaging setting. Whilst the trained model in this case would not generalise well to other objects, this approach allows more efficient training for a more limited set of object features of interest.

2. Methodology

Our methodology uses three key steps: feature detection in each of the two 2D projection images, matching features, and mapping features' locations in 3D.

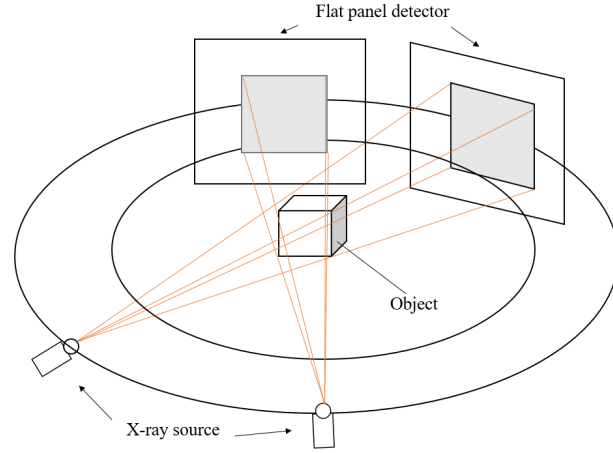


Figure 1. For stereo X-ray tomographic imaging with two views, two X-ray projection images are taken of an object from two different viewing directions.

The approach utilises a similar point identification and mapping process to our previous work [9]. We assume a stereo X-ray tomography system as shown in 1. Features in each of the stereo images are identified and mapped into 3D space, where the back-projected volume is used to identify 3D feature location. We summarise our proposed approach in Fig. 2.

2.1. Feature Detection

We formulate the feature detection problem as a standard binary classification problem, using a deep neural network that, for each pixel, estimates the probability that this pixel comes from a feature. A neural network implements a parametrised map

$$\mathbf{y}_{mask} = f(\mathbf{x}_{raw}; \theta), \quad (1)$$

where θ are the model parameters, $\mathbf{x}_{raw} \in \mathbb{R}^{m \times n}$ is the X-ray projection images encoding the spacial distribution of measured X-ray attenuation on one of the imaging planes and $\mathbf{y}_{mask} \in \mathbb{B}^{m \times n}$ is a pixel wise class probability map that can be thresholded to estimate feature locations. Parameters θ adapted using stochastic gradient optimisation to minimise the empirical risk over a training data-set $\{(\mathbf{x}_{raw}^i, \mathbf{y}_{mask}^i)\}_{i=1}^N$ which comprises N image pairs. The function $f(\cdot)$ in Eq. (1) is realised using the standard U-net architecture described in [14], but implemented and trained as a classification network(i.e. using a sigmoid activation function and a binary cross-entropy loss).

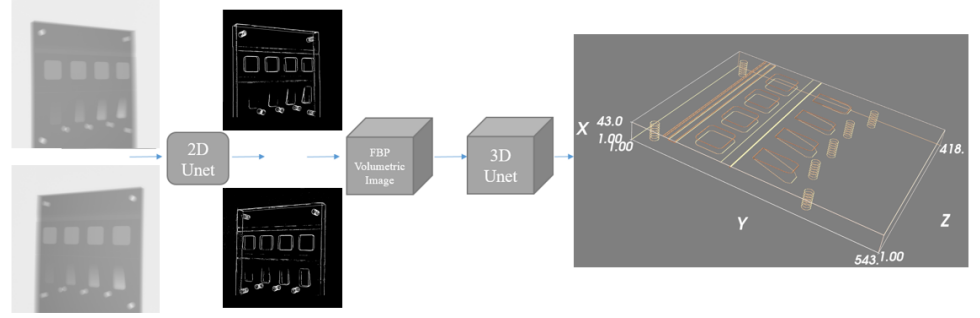


Figure 2. Overview of the proposed framework. The input is a pair of X-ray projection images. Each projection is fed independently into the same 2D U-net to compute two different feature maps, where the background is removed, leaving estimates of the line and point feature locations. Utilising scan geometry knowledge, the two feature maps are then back-projected into a 3D volume using the FDK algorithm, with the back-projected volume further processed using a 3D U-net to generate the 3D spatial feature maps.

2.2. Feature matching and 3D mapping

As in [9], to derive a robust feature matching and 3D mapping approach, we use standard filtered back-projection methods to map the identified feature locations into 3D space. This is followed by a feature location identification step, where a deep neural network is applied to the 3D image to identify feature locations. Formally, if $B_L(\cdot)$ and $B_R(\cdot)$ are the filtered back-projection operators [15] for the left and right projection images¹, then we train a mapping $g(\cdot)$ that maps the two extracted feature maps $\hat{\mathbf{y}}_{mask}^L$ and $\hat{\mathbf{y}}_{mask}^R$ to a 3D tomographic volume $\mathbf{y}_{vol} \in \mathbb{R}^{m \times n \times o}$.

$$\mathbf{y}_{vol} = g\left(B_L(\hat{\mathbf{y}}_{mask}^L) + B_R(\hat{\mathbf{y}}_{mask}^R); \alpha\right) \quad (2)$$

Here, $\hat{\mathbf{y}}_{mask}^R$ and $\hat{\mathbf{y}}_{mask}^L$ are the estimated 2D feature maps estimated from the left and right X-ray images and $\mathbf{y}_{vol}$ is the estimated volumetric image encoding the probability that each voxel contains one of the features. The function $g(\cdot)$ in Eq. (2) is again parameterised by trainable parameters α . We here use the 3D U-net described in [16] to implement this function.

3. Dataset

We demonstrate the method's ability by mapping the edge and line features of a test phantom manufactured from homopolymer acetal. This phantom was imaged previously in unrelated work [17] but is a useful example here as it contains a range of different linear edge features. We show a photo of the object together with its original design drawing and a 3D rendering in Fig. 3.

The object was originally scanned over a range of angles, though here we utilise only two projections taken at approximately 60° . Scanning was performed on a Nikon XTH225 X-ray micro-tomography system. The original X-ray intensity images are shown in Fig. 6. The 2000×2000 detector had a pixel size of 0.2mm and the object was scanned with a source-to-detector distance of about 923mm, and a source object distance of about 290mm.

To train the feature detection 2D U-net model, we generated synthetic training datasets, generating 12 3D images of size $256 \times 256 \times 256$. Each image contained several simple 3D shapes with straight or rounded corners that approximately matched the shapes expected in the real X-ray images. We then positioned these shapes within a large rectangular prism,

¹ Note, extensions to settings with three or more projections follow the same ideas

assigning low attenuation to the shapes and higher attenuation to the prism. Gaussian noise was added to the background. From these 12 3D images we generate 24 1024×1024 2D projection images at random object orientations using the Astra Toolbox [18]. Each 2D projection was partitioned into 144 overlapping blocks of size 256×256 , providing 3456 samples for training. We show three randomly selected 2D training data pairs in Fig. 4, where we show the projection images (top) together with the projected ground truth binary images identifying the locations of the corner and edge features (bottom).

To train the 3D U-net to map the detected features in the projection images into 3D space, we generate a 3D edge map from its CAD drawing, which only contains object edge features. To generate a diverse set of images, the same edge feature map was rotated with 1° intervals around an axis parallel to the longest object side. This generated 360 3D images with edge features, each of size $544 \times 64 \times 544$ voxels and with a voxel size of $0.24 \times 0.24 \times 0.24 \text{ mm}^3$. Each of these 3D blocks was then projected to generate pairs of 2D projection images, where projections were collected at $\pm 30^\circ$. These ideal 2D feature maps were then back-projected into 3D images using the FDK algorithm to generate simulated back-projected feature maps. When training our 3D U-net, we can thus use the simulated back-projected feature maps as network inputs, with the original clean edge feature maps as desired outputs. Example data is shown in Fig. 5.

3.1. Calibration for stereo X-ray imaging system

Whilst we had nominal values for the main system parameters, at the time of scanning, the system was not fully calibrated so that the nominal values given above might have significant errors. Thus, with only two real projections, by the epipolar constraint method [19] with manually selected matching points from two real projections, the relative camera matrix between two cameras can be calculated. We here use the first view as the reference

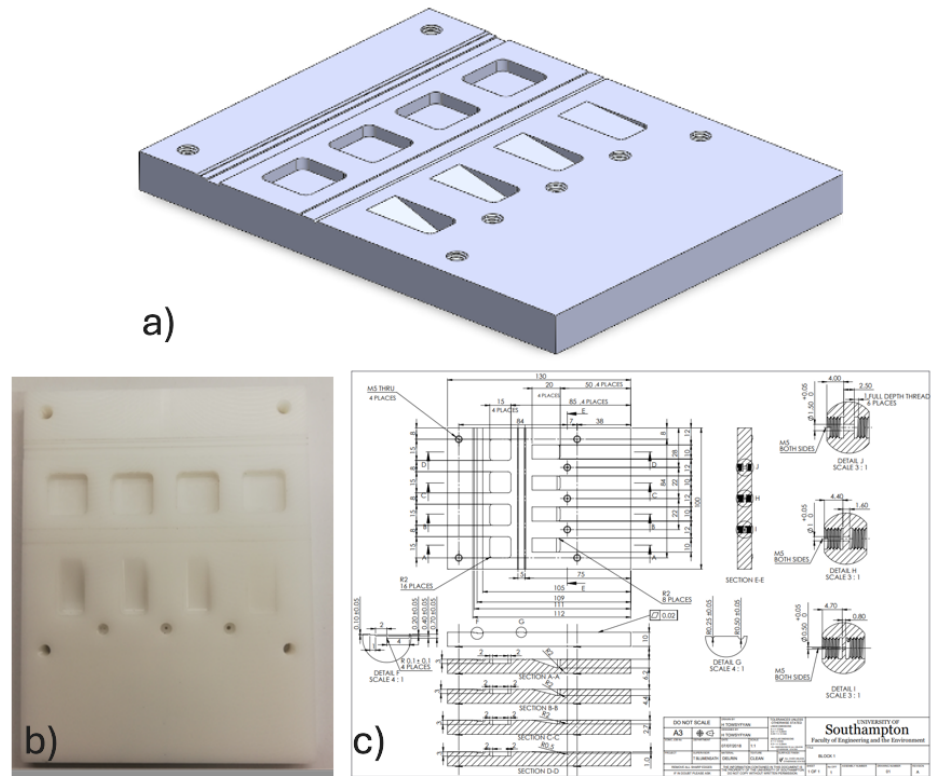


Figure 3. a) A 3D rendering of the 3D block; b) a photo of the workpiece and c) its CAD drawing.

coordinate. Thus, the first camera matrix can be denoted as Eq. (3), where P_1 is the projection matrix of the first view, K is the intrinsic matrix, and $\mathbf{I}$ is the identity matrix. The second camera matrix is denoted as Eq. (4), where R is the relative rotation matrix and t is the normalised translation matrix between two views.

$$P_1 = K[I|0] \quad (3)$$

$$P_2 = K[R|t] \quad (4)$$

To further refine the calibration, the simulated 3D edge feature map was projected using the two estimated camera matrices and compared visually with the two real projections (see Fig 7). We define the world coordinates based on the rotation centre of the stereo X-ray imaging system under the Astra Toolbox. Here we add a tiny pitch, roll and yaw to the centre point of the 3D block outline features to control its pose to make its forward projections under the stereo X-ray geometry mostly overlap with the real projections, while the two views angle are -29° and $+32^\circ$. The comparison of the real outline features and simulated projections is shown in Fig. 7, demonstrating a good overlap of the simulated features and the real data after calibration.

4. Results

Due to the fact that our system calibration for the available data did not use an optimised and calibrated calibration phantom, errors in the estimated 3D feature location are dominated by calibration errors as well as deviations of the manufactured object's geometry from the ground truth CAD data (for example, subtractive manufacturing led to a slight warping of the workpiece). Numerical values that try to quantify the precision of locating features in 3D space are thus dominated by errors in the calibration and assumed ground truth feature locations. Our method is also the first method developed specifically for the 3D mapping of linear X-ray features from stereo X-ray projections so that there is no

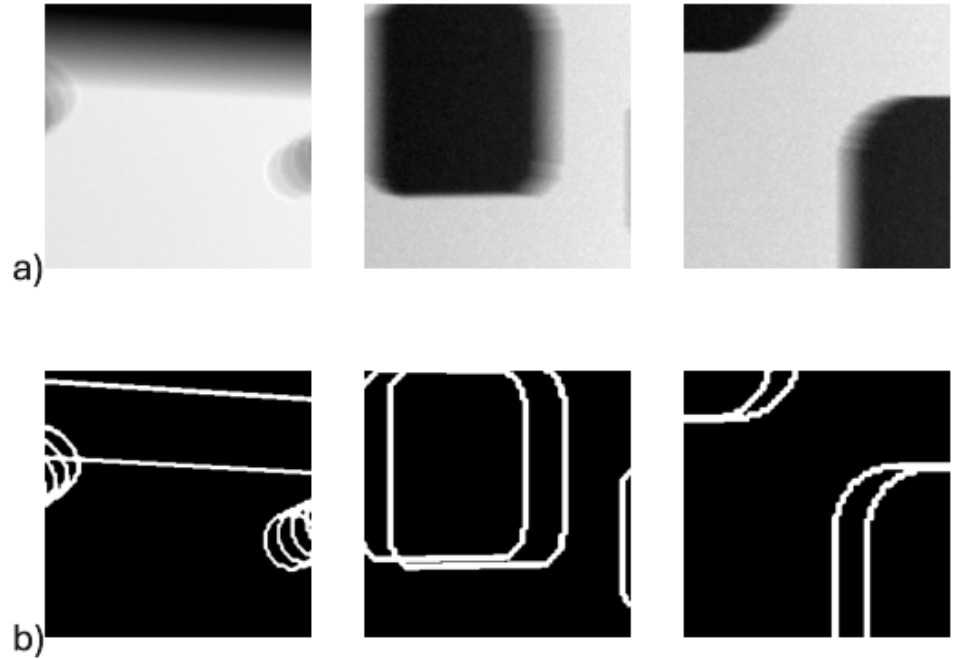


Figure 4. Panel a) shows three projections from the set of training samples used to train the feature detection network whilst panel b) shows the projected ground truth edge feature maps.

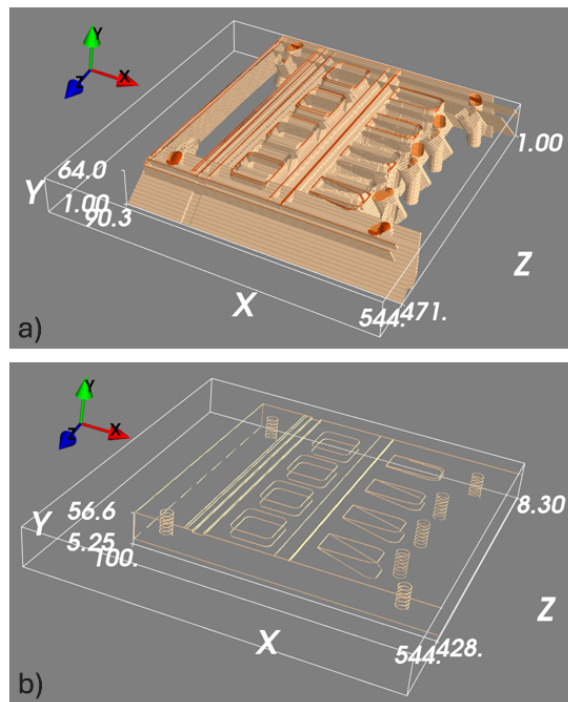


Figure 5. A set of 3D feature mapping training samples. Panel a) is a thresholded 3D rendering of the back-projected volume that is used as the input to the machine learning model, and panel b) is the ground truth we are trying to predict.

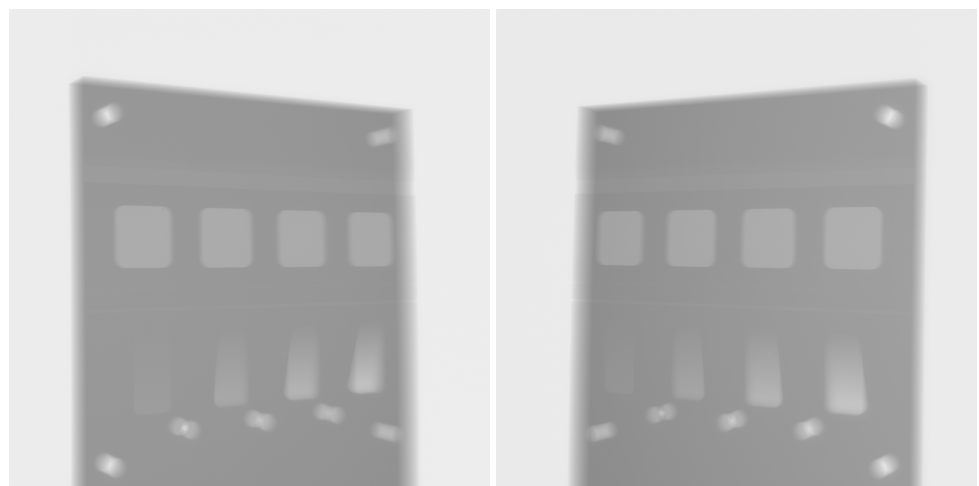


Figure 6. Intensity projection images of the physical test sample, collected by Nikon XTH225 X-ray tomography system with a 60° relative rotation.

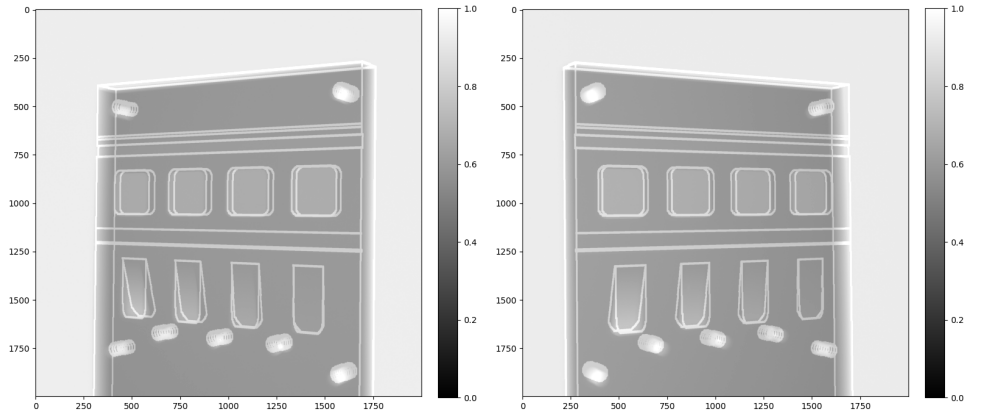


Figure 7. The overlap between simulated edge map projections and the real projections. We measured the geometric error at the object corners and the six screw holes positions, finding a 5 pixels error on average due to the geometric calibration procedure.

comparable state-of-the-art method available that we could use for relative performance comparisons. Whilst presenting some numerical results, we primarily limit our evaluation to visual inspection, which is able to show that a) features are located accurately in the projections and the correct matching is performed in 3D.

We evaluate feature detection and the 3D feature mapping steps independently.

4.1. General training and evaluation approach

Both networks were trained as classification networks for feature detection using the projections (2D network) or the filtered back-projection of the 2D projections (for the 3D network) as inputs and the binary images showing 2D projected or 3D point locations as output. Both networks were implemented using TensorFlow 2.x and optimised using an Nvidia GTX4070ti graphics card for 2D network and NVIDIA A100 graphics cards on the University of Southampton IridisX cluster for the 3D network. We use the Adam optimiser with the synthetic training data, a learning rate of 10^{-4} and 100 epochs. The loss function was the binary cross-entropy.

4.2. Feature Detection

After training the 2D feature detection U-net using the synthetic data described above as the training set, we used the real projection images from the physical phantom to test the method. The real test images were processed by converting the measured X-ray intensity into attenuation values before cropping the images into 338 overlapping blocks of size 256×256 . We then applied the 2D U-net to all test sample blocks from both projections. Probabilities were averaged over all blocks that contained a particular image pixel before thresholding to produce a full size feature map. To evaluate the performance of the method, we use the simulated projections after calibration as our ground truth and visually compare the CAD data as approximate ground truth to the estimated 2D feature maps (See Fig. 8).

4.3. 3D Feature mapping

We then used the estimated 2D features of the real projections from the feature detection model and, for comparison, the two simulated (CAD-based) feature projections in order to generate two back-projected volumes using the same calibrated projection geometry. Both back-projected volumes will go through the 3D model trained for 3D spatial position estimation.

The back-projected volumes generated by the two simulated projections are shown in Fig. 9 together with the 3D edge-feature map estimated with the 3D U-net model. This

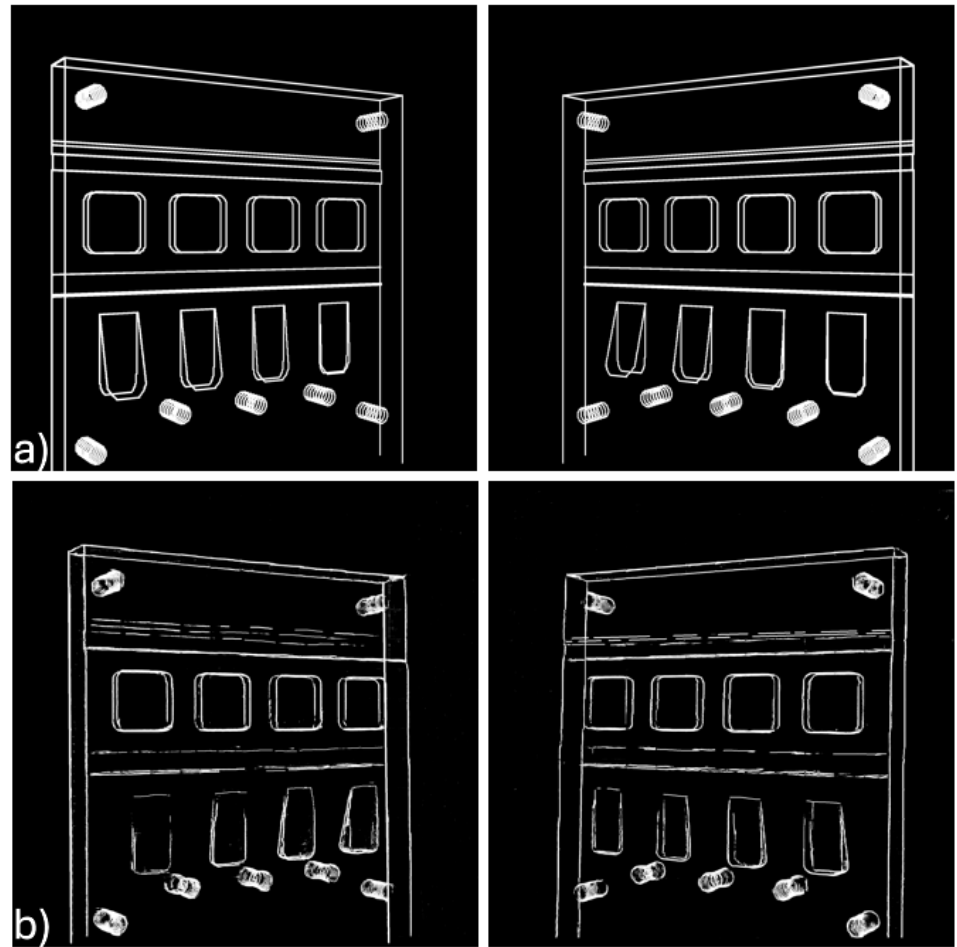


Figure 8. Comparison between the simulated ground truth 2D feature maps (derived from the CAD data of the object) (a) and the 2D feature maps estimated from the real data (b).

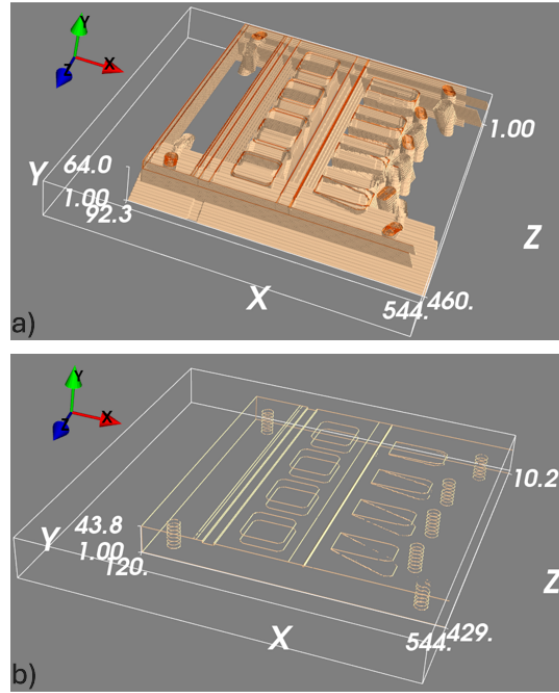


Figure 9. Visualisation of the operation of the 3D U-net. Panel a) shows the back-projected volume generated from two simulated projections of the edge feature maps, whilst panel b) is the 3D edge feature map estimated from the top image using the 3D U-net.

can be compared to the back-projected volume generated by the two 2D feature maps estimated from the real data shown in Fig. 10, where we also show the 3D edge-feature map estimated with the 3D U-net model. To numerically evaluate the error between these two estimations and the ground truth, we compare their corner and screw holes positions to those for the simulated phantom. As seen in Fig. 11, the errors from corners and screw holes are in the range of between 1 to 7 voxels, or 0.24mm to 1.68mm . Considering the size of the 3D block and the inaccuracy from calibration, these are relatively small errors, which are assumed to be mainly due to the ad hoc post-scan calibration used here. Better calibration results could be obtained using a dedicated calibration object with a fixed stereo X-ray imaging setup.

A numerical evaluation of the results was based on the location of features in a nominal geometry that was also used for system calibration and should theoretically be aligned with the true object location. Distance errors in feature location between 0.24mm to 1.68mm were found which were likely dominated by errors in the calibration process, though a detailed analysis of the dimensional accuracy of the approach will be left to a future study where a more controlled system calibration approach can be used.

It is important to contrast our approach to related methods such as limited-angle tomography [20]. In limited-angle tomography, a full 3D object reconstruction is computed from a few projective measurements using advanced iterative or deep learning based image reconstruction algorithms. This is different from our approach in that our stereo XCT approach only reconstructs the 3D location of point and line features (such as the edges and corners of objects as done in the current paper). Whilst stereo XCT thus only recovers limited 3D information compared to limited angle XCT, limited angle CT has the disadvantage that it requires extremely strong prior knowledge about the imaged object, which is then exploited in the iterative or deep learning based reconstruction. If we had no prior information on object geometry and density profiles, then limited-angle

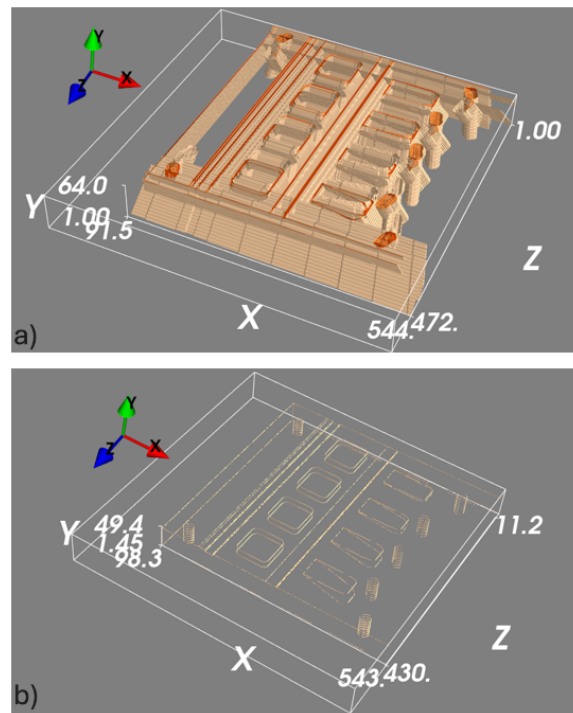


Figure 10. a) Back-projected feature maps and b) estimated 3D features for the real data.

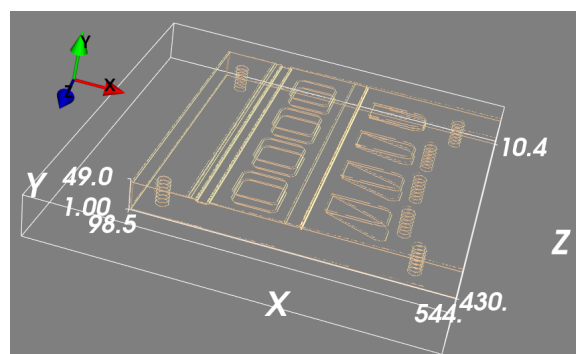


Figure 11. The positions error in 3D space between the 3D outline estimation by simulated projections and real outline features.

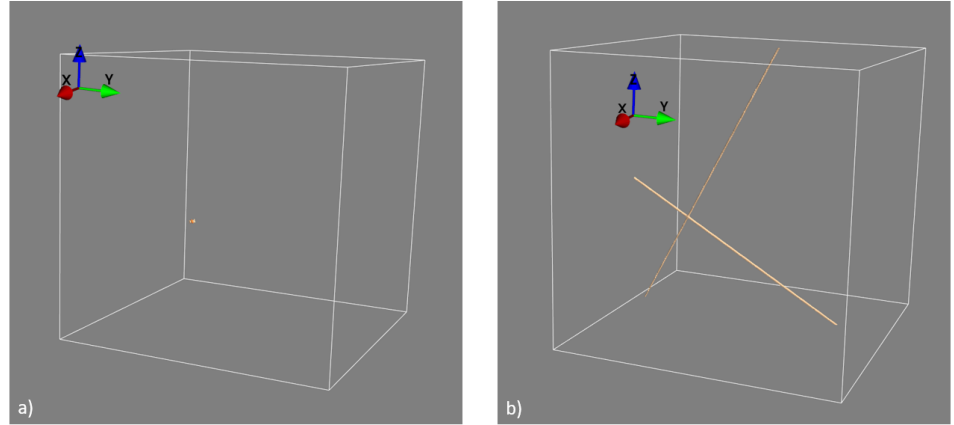


Figure 12. Differences between stereo X-ray tomography and limited-view tomography. a): Stereo XCT reconstruction of a point-like object. b) FBP reconstruction of the same object, the object's information is smeared out along the X-ray paths that intersect the object.

reconstruction does lead to images with significant artefacts. For example, filtered back-projections computed from a pair of projection images would lead to an image without nearly any discernible information. Our stereo XCT method is on the other hand still able to identify object edges and corners, even if we know nothing about the object's shape and density profile. In fact, edges and corners can always be mapped if we are able to locate and match the edges in the 2D projections without the need to impose any additional prior object knowledge. This is demonstrated in an example where we image a point-like object using two X-ray projection views measured from two viewpoints that are roughly 60 degrees apart. The left panel in Fig. 12 shows the recovery of that point when using our stereo XCT technique, whilst the right panel uses the prior information agnostic filtered back-projection reconstruction, which reconstructs the point as two lines in 3D space. If the X-ray views in addition had contained additional objects without clear edges or corners, then the stereo XCT reconstruction would not have been effected, whilst the limited angle reconstruction would be further contaminated by the additional X-ray attenuation estimates from the additional object, but again smeared out along the X-ray paths of the two views.

Our approach also differs substantially from X-ray based methods proposed in medical imaging that track objects during surgery. Here, a single projection image is often used in which an object, such as a medical instrument or tumour is identified and then tracked in a single 2D X-ray projection image [21]. Whilst some of these methods use deep learning based methods to also estimate the location of the object in 3D space, as this information is not measured in a single X-ray projection, this estimate needs to rely again on prior information such as knowledge of typical human anatomy or preoperative full CT scans.

5. Discussion and conclusions

In this paper, we extended the stereo X-ray tomography framework from our previous work to the estimation of the 3D location of the corners and edges of 3D objects. Using two deep neural networks, trained using simulated data, we could extract point and line feature locations from the two real projection images and, using the calibrated camera matrices of the system, project these back into 3D space. To identify feature location in 3D space, a further 3D neural network was used, again trained on simulated data.

Considering that the main objective of this work is to extend our previous study by applying the stereo X-ray tomography system to more complex and realistic scenarios, we aim to reconstruct 3D features using only a pair of uncalibrated projection images and the object's initial blueprint. While the geometry estimated through calibration may introduce

certain errors, which in turn affect the accuracy of feature localisation and consequently impact the final 3D mapping, the resulting errors appear to be within a manageable range. From a qualitative perspective, the system successfully achieves its intended purpose and produces reliable results, identifying 2D feature locations with few omissions and false positives whilst the 3D method clearly matches the correct features to locate these into 3D.

A potential limitation that warrants further investigation in future work is the generalizability of the approach to different feature geometries and object complexities. Although our training data was generated using simple geometric patterns placed in various configurations, there are inherent differences in appearance and physical characteristics compared to real experimental data. This introduces a possible risk that visual similarities may not fully capture underlying structural differences, potentially affecting generalisation performance. We plan to address this limitation in future work by further enhancing the diversity and realism of the training data. In this work, we roughly categorise features into point features (such as highly attenuating points and corners) and line features (including highly attenuating linear features and object edges). However, in real-world applications, features might often be more difficult to estimate. Point-like and line-like features can potentially vary in their width (for example, a rounded object corner leads to a slightly smeared out linear feature in the projections). Furthermore, the strength (and thus visibility) of these features in the projection image will vary with object contrast. For a method trained for a specific application, it is thus crucial to match these aspects in the simulated data to the expected feature range in the real data. If instead, the aim is to develop a deep learning based model that works over a much wider range of objects, the greater variation of feature width and contrasts expected in this more general set of applications will need to be taken into account. As always in deep learning based methods, matching the statistics of these properties in the training data to those in the expected real data will minimise the expected average error.

We were here also limited in that we only had suitable data for a single object. The effect of geometric object complexity, imaging noise and possibly contrast for multi-material objects was thus not evaluated, though the influence of noise on feature detection was already studied in our previous work with fiducial markers and it is to be expected that increased noise will lead to a decreased performance in 2D feature detectability. For the detection of edge features, the 'sharpness', i.e. radius of an object's edge, will likely also play a key role in the exact localisation of the edge in both 2D and then ultimately 3D. These are issues we hope to investigate in future work.

Author Contributions: Conceptualization, Z.S. and T.B.; methodology, Z.S. and T.B.; software, Z.S.; validation, Z.S.; formal analysis, Z.S. and T.B.; investigation, Z.S. and T.B.; resources, Z.S.; data curation, Z.S. and T.B.; writing—original draft preparation, Z.S.; writing—review and editing, Z.S. and T.B.; visualization, Z.S.; supervision, T.B.; project administration, T.B.; funding acquisition, Not applicable. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors confirm that the data supporting the findings of this study are available within the article.

Acknowledgments: We would like to thank NVIDIA for the provision of an NVIDIA GTX4070ti GPU and the authors acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

XCT	X-ray Computed Tomography
3D	Three-dimensional
FBP	Filtered back-projection
TV	Total Variation

References

1. Song, J.; Liu, Q.H.; Johnson, G.A.; Badea, C.T. Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT. *Medical physics* **2007**, *34*, 4476–4483.
2. LI, H.; KAIRA, S.; MERTENS, J.; CHAWLA, N.; JIAO, Y. Accurate stochastic reconstruction of heterogeneous microstructures by limited x-ray tomographic projections. *Journal of Microscopy* **2016**, *264*, 339–350. <https://doi.org/10.1111/jmi.12449>.
3. Yuan, H.; Jia, J.; Zhu, Z. SIPID: A deep learning framework for sinogram interpolation and image denoising in low-dose CT reconstruction. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018, pp. 1521–1524.
4. Lee, H.; Lee, J.; Kim, H.; Cho, B.; Cho, S. Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences* **2018**, *3*, 109–119.
5. Pelt, D.M.; Roche i Morgó, O.; Maughan Jones, C.; Olivo, A.; Hagen, C.K. Cycloidal CT with CNN-based sinogram completion and in-scan generation of training data. *Scientific Reports* **2022**, *12*, 1–13.
6. Ernst, P.; Chatterjee, S.; Rose, G.; Speck, O.; Nürnberger, A. Sinogram upsampling using Primal-Dual UNet for undersampled CT and radial MRI reconstruction. *arXiv preprint arXiv:2112.13443* **2021**.
7. Adler, J.; Öktem, O. Learned primal-dual reconstruction. *IEEE transactions on medical imaging* **2018**, *37*, 1322–1332.
8. Li, X.; Wang, S.; Chen, P.; Wang, L. 3-D inspection method for industrial product assembly based on single X-ray projections. *IEEE Transactions on Instrumentation and Measurement* **2021**, *70*, 1–14.
9. Shang, Z.; Blumensath, T. Stereo X-ray Tomography. *IEEE Transactions on Nuclear Science* **2023**.
10. Kak, A.C.; Slaney, M. *Principles of computerized tomographic imaging*; SIAM, 2001.
11. Horn, B.; Klaus, B.; Horn, P. *Robot vision*; MIT press, 1986.
12. Sagara, Y.; Hara, A.K.; Pavlicek, W.; Silva, A.C.; Paden, R.G.; Wu, Q. Abdominal CT: comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 patients. *American Journal of Roentgenology* **2010**, *195*, 713–719.
13. Hoffman, E.J.; Huang, S.C.; Phelps, M.E. Quantitation in positron emission computed tomography: 1. Effect of object size. *Journal of computer assisted tomography* **1979**, *3*, 299–308.
14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.
15. Feldkamp, L.A.; Davis, L.C.; Kress, J.W. Practical cone-beam algorithm. *Josa a* **1984**, *1*, 612–619.
16. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In Proceedings of the International conference on medical image computing and computer-assisted intervention. Springer, 2016, pp. 424–432.
17. Deyhle, H.; Towsyfy, H.; Biguri, A.; Mavrogordato, M.; Boardman, R.; Blumensath, T. Spatial resolution of a laboratory based X-Ray cone-beam laminography scanning system for various trajectories. *NDT & E International* **2020**, *111*, 102222. <https://doi.org/10.1016/j.ndteint.2020.102222>.
18. Van Aarle, W.; Palenstijn, W.J.; De Beenhouwer, J.; Altantzis, T.; Bals, S.; Batenburg, K.J.; Sijbers, J. The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy* **2015**, *157*, 35–47.
19. Ullman, S. The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **1979**, *203*, 405–426.
20. Pelt, D.M.; Batenburg, K.J.; Sethian, J.A. Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks. *Journal of Imaging* **2018**, *4*, 128.
21. Liu, X.; Geng, L.S.; Huang, D.; Cai, J.; Yang, R. Deep learning-based target tracking with X-ray images for radiotherapy: a narrative review. *Quant. Imaging Med. Surg.* **2024**, *14*, 2671–2692.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

370
371
372