Stereo X-ray Tomography

Zhenduo Shang and Thomas Blumensath

***Abstract*—X-ray tomography is a powerful volumetric imaging technique, but detailed three dimensional (3D) imaging requires the acquisition of a large number of individual X-ray images, which is time consuming. For applications where spatial information needs to be collected quickly, for example, when studying dynamic processes, standard X-ray tomography is therefore not applicable. Inspired by stereo vision, in this paper, we develop X-ray imaging methods that work with two X-ray projection images. In this setting, without the use of additional strong prior information, we no longer have enough information to fully recover the 3D tomographic images. However, up to a point, we are nevertheless able to extract spatial locations of point and line features. From stereo vision, it is well known that, for a known imaging geometry, once the same point is identified in two images taken from different directions, then the point's location in 3D space is exactly specified. The challenge is the matching of points between images. As X-ray transmission images are fundamentally different from the surface reflection images used in standard computer vision, we here develop a different feature identification and matching approach. In fact, once point like features are identified, if there are limited points in the image, then they can often be matched exactly. In fact, by utilising a third observation from an appropriate direction, matching becomes unique. Once matched, point locations in 3D space are easily computed using geometric considerations. Linear features, with clear end points, can be located using a similar approach.**

***Index Terms*—feature detection, X-ray Computed Tomography, stereo matching**

I. Introduction

X Ray Computed Tomography (XCT) is a powerful volumetric imaging technique, widely applied in medical and industrial applications to reveal the internal structure of an object by collecting thousands of projection images from different views around the object. The achievable image quality depends on the number of acquisitions. With the help of advanced algorithms and using regularisation functions such as Total Variation (TV) to enforce certain image smoothness properties [1], or material heterogeneity [2], or learned, data-driven regularisations [3], [4], [5], [6], [7], a significant reduction in the number of measurements is possible, without a significant reduction in image quality. In fact, a range of recent papers has shown that even single projection images are sufficient in certain settings to identify one known image out of a small set of possible reconstructions [8], though this only works by imposing very strong prior constraints, which is only possible if the objects are extremely predictable in their 3D shape.

Our work takes a different approach. Instead of trying to reconstruct full tomographic images from limited observations, which require very strong prior information, we only aim at the recovery of the 3D location of point and line like features. Our approach is thus different from standard limited view tomography reconstruction. It has the advantage of only requiring two projection images, but it has the disadvantage that it only allows the spatial mapping of point and line features. However, this is achieved without the strong constraints that are imposed in full image reconstruction from limited observations and thus also works for general objects, as long as these include line and point features. 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 image at each time-step.

There are several approaches where such an approach might be of interest. For example, in medical imaging one might envisage a system that tracks medical instruments inside a patient during an operation, whilst in the study of manufacturing processes, one can study the carbon fibre ply deformations during the carbon fibre layup and curing process. Similar applications are possible whenever the object includes natural markers or in cases where it is possible to include artificial judicial markers as can be done in digital volume correlation experiments or in the study of powder movements.

In these applications, using a setup with two or more X-ray sources and detectors, spatial mapping of these features then becomes possible at the speed of the detector frame rate which is orders of magnitudes faster than full computed tomography. Our approach is inspired by computer stereo vision systems and we thus assume an imaging setup with two (or a few) X-ray source/detector pairs as shown in Fig. 1. This allows us to take two (or a small number of) projection images at the same time, but, without significantly restricting the volumetric images we are inspecting. With such a setup, we show that we are able to identify point and line like features and estimate their location in 3D space.

With visible light, for non-transparent objects, the amount of light measured at any point on a camera's image plane is assumed to be due to the reflection of light from a single point in the 3D scene. Thus, if we take two images of a scene from slightly different directions and if we are able to match points that come from the same point in the 3D scene between the two images, then a depth map for each image can be computed from simple geometric principles, placing each point in each image to a unique point in 3D space. The main issue in stereo vision is thus one of matching points between the two images of a stereo pair [9].

Unlike stereo vision with visible light, in X-ray imaging, individual points on the imaging plane represent X-ray attenuation values along the entire X-ray path, that is, along the straight line from the X-ray source to the pixel on the X-ray detector [10]. Therefore, individual points on one imaging plane no longer correspond to a single point in 3D space, but to an entire line. For a full tomographic reconstruction, point

Zhenduo S. and Thomas B. are with the University of Southampton, Southampton, United Kingdom.

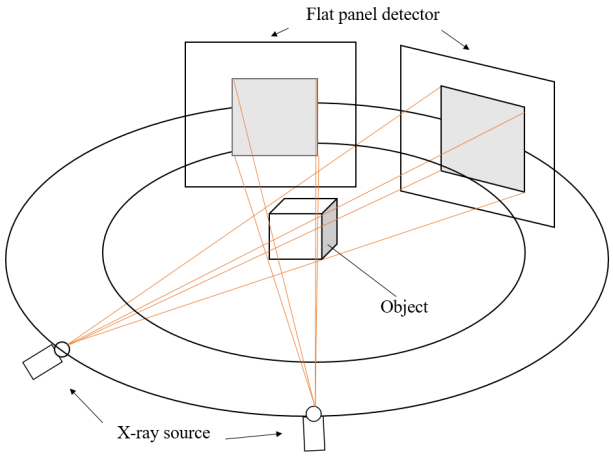

Fig. 1: For stereo X-ray tomographic imaging with two views, two X-ray projection images are taken from an object from two different directions.

matching based approaches are thus not possible. However, if there are point features in 3D space, such as the end of a linear feature, the corner of an object or an object that is of a similar size to the resolution of the system, then, this feature can be mapped in 3D space even from as few as two tomographic projections. To achieve this, we need to 1) identify all point like features in the two X-ray views and 2) match these features between the two views. Once found and matched, mapping the features into 3D space uses the same geometric considerations as stereo vision. This approach can also be extended to line like features, by treating each line as a string of point like features, though here, matching can become more difficult. Whilst there has been significant work on stereo vision over the years, we find that the methods used currently in stereo vision to solve both, the point identification and the point matching issues do not work with X-ray images. To achieve these goals, we instead develop new approaches that use deep neural networks, one that detects relevant features in the individual 2D projection images and one that matches these features and places them into 3D space.

A. Proposed method

Whilst there are many existing feature detection methods [11], [12], [13], [14], [15], these have generally been developed for standard imaging applications and we found they do not work well on X-ray transmission images. In this paper, we thus propose a novel, deep learning based feature detector that can be trained on specific point and line like features. The particular challenge here is for the method to ignore the other image information that clutters X-ray projection images, but that is not related to the features of interest. The proposed framework is shown in Fig. 2. After extracting the features in the 2D X-ray projection images, we then match the features by using the filtered back-projected (FBP) [16], [17] method to generate a 3D volumetric image of the extracted features. The back-projection of the features from the two images that are due to the same point in space will overlap in 3D space. If there are few features and if these features are randomly

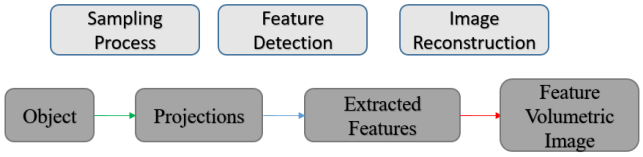

Fig. 2: The proposed feature reconstruction framework. The sampling process, shown with the green arrow, represents the imaging process which generates a few projection images. Our approach then implements a feature detection step (blue arrow) and, once features are detected, a 3D mapping step (red arrow).

located in space, then it is likely that this intersection is unique, which will then lead to a unique match, though the more features we have, the more likely will it be that there will be an intersection of the lines in 3D space from more than one point in each image. In this case, an exact match is not possible from two observations, though by adding a third observation, this issue can be overcome, though adding additional views does not allow us to identify planar features using stereo vision ideas. Another issue that can arise if there are many features in 3D space is that two features are aligned such that they both are measured in a single point on one of the imaging planes. To increase the robustness to the exact localisation of features on the two imaging planes, instead of simply looking at intersections in 3D space after back-projection, we use a simple deep neural network to process the back-projected volumetric image to generate a 3D image containing the point like features.

B. Contributions

The main contributions of this paper are:

1) We design a convolutional neural network to identify point and line features in individual X-ray projection images that work even if the images show additional complex object features.

2) We design a deep neural network that uses pairs of feature images to map the features to their spatial location in 3D. This network in effect matches the features and uses geometric information to compute spatial locations.

3) We evaluate the methods on two datasets, a synthetic dataset and a real X-ray imaging data-set.

II. Methodology

A. Feature Detection

There are many feature detection methods that have been developed for 2D images and common methods include the Speeded Up Robust Feature (SURF) method [11], [12], Scale Invariant Feature Transform (SIFT) method [13], Features from Accelerated Segment Test (FAST) method [14] and Harris corner detection method[15]. However, these methods search for features commonly found in photographic images. They have not been designed to detect point and line like features in tomographic data and are thus not suitable for our task. Not only are they often unable to detect point and line features, they are also prone to detect other image structure in the tomographic background, which is not of interest here. Therefore, we use a binary classification neural network to

detect features of interest. The feature detection problem can be summarized as finding parameters θ such that

$$\mathbf{y}_{mask} = f\left(\mathbf{x}_{raw}; \theta\right) \tag{1}$$

is a mapping from a raw 2D X-ray projection images $\mathbf{x}_{raw} \in \mathbb{R}^{m \times n}$ to a binary feature mask $\mathbf{y}_{mask} \in \mathbb{B}^{m \times n}$. The parameters are learned from training a data-set $\left\{\left(\mathbf{x}^i_{raw}, \mathbf{y}^i_{mask}\right)\right\}_{i=1}^{N}$ which comprises N sample pairs. The function in (1) is here implemented using a standard U-net architecture as described in [18], where the method is implemented as a classification network as in the original paper.

B. Mapping Features into 3D space

Once we have identified the features in the 2D projections, we could try and match points in projection pairs in the same way in which this is done in stereo vision. This is a two stage process with the difficult step being the matching of points in one projection to those in the other projection. In stereo vision, this matching is typically done by looking at the similarity of the image around a point and matching the points if their neighbourhoods are similar. This matching is further constrained by the fact that a point in one image can only match with points along a line in the other image, a constraint known as the epipolar constraint of the stereo camera model. In fact, if there are few points that need to be matched between two images, this constraint often means that there is a unique match as long as there is no more than one point feature on each epipolar plane in space. This process can thus also be used for matching points in our setting. Note that, for a trinocular X-ray stereo system, that is, for a system with 3 X-ray sources and detectors, where none of the epipolar planes of each of the three detector pairs is parallel, unique matching becomes increasingly likely, as now, not only do matching points lie on one epipolar plane, but on three epipolar planes [19]. To match linear features, we can start by matching the endpoints of a linear feature and then track the line from its endpoints. Once matched, point features can then be mapped into 3D space using knowledge of the camera geometry.

The similarity of the geometric depth estimation problem is shown in Fig 3. The left panel shows the principle of visible-light stereo matching and depth estimation. The right panel shows the geometry of stereo X-ray tomographic imaging with two projections, which can also calculate the spatial information of a point in space from matched points in the two projections.

To increase the robustness of this approach to errors in the localisation of point features, we here instead use a learned feature matching and 3D mapping method. We train the 3D U-net as described in [20] to process the back-projected volumetric image of the detected feature maps to generate a 3D image containing the point and line features only. Formally, if $B_L(\cdot)$ and $B_R(\cdot)$ are the filtered backprojection operators [21] for the two projection images, then we train a mapping

$$\mathbf{y}_{vol} = g\left(B_L(\widehat{\mathbf{y}}^L_{mask}) + B_R(\widehat{\mathbf{y}}^R_{mask}); \theta\right) \tag{2}$$

from two extracted feature maps $\widehat{\mathbf{y}}^L_{mask}$ and $\widehat{\mathbf{y}}^R_{mask}$ to a 3D tomographic volume $\mathbf{y}_{vol} \in \mathbb{R}^{m \times n \times o}$. The function $g(\cdot)$ in (2) is then the 3D U-net as described in [20].

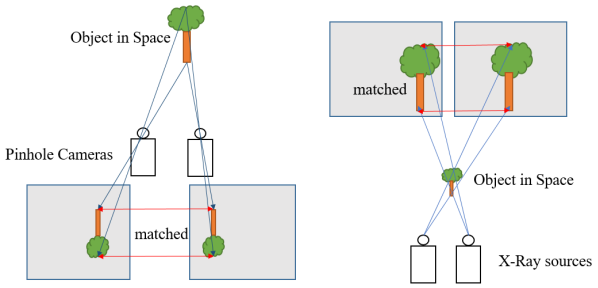

Fig. 3: The geometry of visible-light stereo matching of point features using a pinhole camera model and X-ray projection point matching.

C. Proposed Framework

Our proposed approach is summarised in Fig. 4. We employ a standard 2D U-net [18] as a feature detector $f(\cdot; \theta)$, and a similar standard 3D U-net [20] to mode the mapping $g(\cdot, \cdot; \theta)$ to extracted projections to 3D feature maps. Both networks act as multi-label classification networks so we use a sigmoid output non-linearity together with the binary cross-entropy loss function.

III. Datasets

To train and test our approach, we use a synthetic and a real XCT dataset. The synthetic dataset uses randomly located point and line features superimposed over randomly placed polyhedra and spheres, each with varying attenuation. The real XCT data was generated during a proof of concept experiment to study the consolidation during carbon fibre tape layup, where thin copper wires were embedded in a carbon fibre tape to act as fiducial markers.

A. The Synthetic Dataset

Our synthetic dataset consists of 100 3D images, generated by randomly generating 10 shapes (either polyhedra or ellipsoids), each with a random orientation, random dimensions and random attenuation values. To generate the volume, the attenuation of overlapping shapes is added to generate the attenuation in the overlapping region. The shapes are restricted to lie within the cylinder that is covered by the X-ray cone-beam by removing the parts of each object that lie outside the cylinder. Three random points and one random line feature are then added. Point and line features are generated with 10 different attenuation values for each volume so that we have 1000 different volumes. Each volume is then projected twice using a cone-beam geometry generated using the Astra Toolbox [22], where the two projection directions are rotated by 90 degrees relative to each other. A 3D rendering of an example is shown in Fig. 5.

The original 3D volumes have $512 * 512 * 512$ voxels whilst the projections have $1024 * 1024$ pixels which can generate 144 overlapping blocks of $256 * 256$ pixels from each projection. We show three small blocks of randomly selected 2D training data pairs in Fig. 6. To make the synthetic data more realistic, for each image we have different intensities for the point and

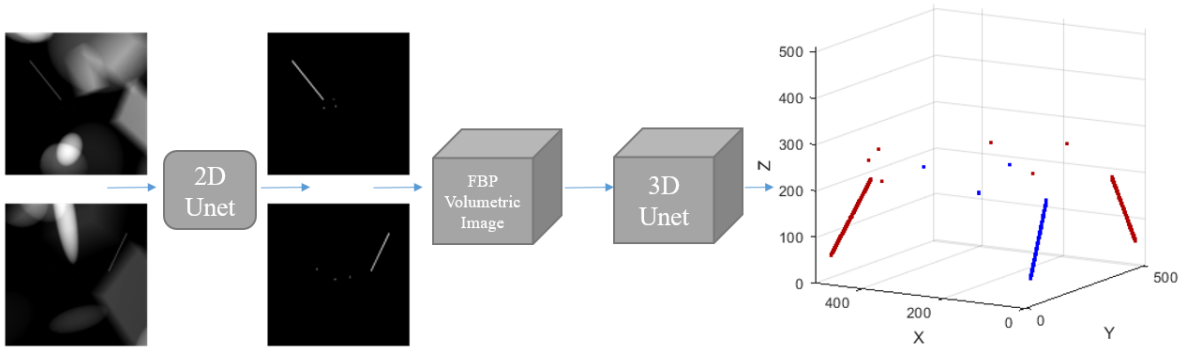

Fig. 4: Overview of the proposed framework. The input is a pair of X-ray projection images which contain the X-ray projection of objects as well as the projection of line and point like features. Each projection is fed independently into the same 2D U-net to compute two different feature maps, where the background is removed. The two feature maps are then back-projected into a 3D volume using the FDK algorithm, which is then further processed using a 3D U-net to generate the 3D spatial feature maps (Then blue line and points are the 3D feature locations and the red is their projections on YZ/XZ planes). Note that for the 2D and 3D U-net in our framework, the input size is $256 * 256$ for 2D U-net, $512 * 512 * 512$ for 3D U-net, and the binary cross entropy loss function for the classification problem, different from the standard 2D/3D U-net.

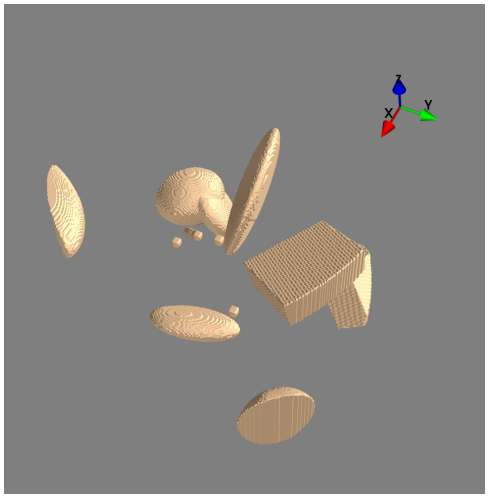

Fig. 5: A rendering of an example of the 3D volumetric data showing polyhedra and ellipsoids as well as point and line features.

line features, gray values for point and line features are blurred with a Gaussian blur. All attenuation values are drawn from a uniform random distribution with values between 0 and 1.

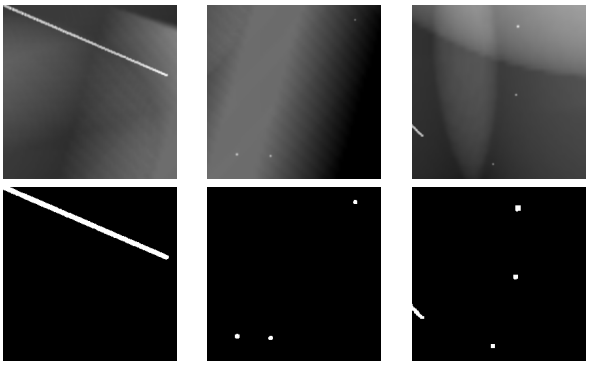

Fig. 6: Three sets of training samples (small blocks) for the feature detection method. The top images show the input data and the bottom images show the feature masks, which are used as the ground truth for feature detection.

B. The Carbon Fibre XCT dataset

The real data was generated for a carbon fibre layup and consolidation experiment, where carbon fibre tape was deposited over a layup tool. Whilst the exact geometry and experimental setup used has not yet been published, the general approach is similar to that described in [23], but with a much more complex geometry. For the current paper, the crucial aspect is that each of the two X-ray views were collected at roughly 90 degree angles, but that each view contained several parts of the mechanism. We then collected x-ray projections at 60 different time points during the consolidation process. 25 μm thick copper wires were embedded into the carbon-fibre tape and their location hand-annotated in the projection images to provide ground truth for training and testing. The experiment was conducted with a bespoke test rig placed in a Nikon XTH225 X-ray tomography system. Images were acquired with a magnification of 8, giving a field of view of about 25mm so that each detector pixel is about 25 micrometers squared. To allow fast imaging and to reduce noise, the data from 2000 by 2000 pixel detector was binned into 1000 by 1000 pixels. The images were pre-processed by converting the measured X-ray intensity into attenuation values [10].

Two different projection images are shown on the left in Fig. 7, where we also show the annotated locations of three linear features on the right. As the copper wires produced in the original projections are extremely faint, we also generated augmented data-sets, where we changed the attenuation value along these features, by either doubling or halving their values.

IV. Experimental evaluation

For our stereo X-ray tomography approach, we introduced two new steps, a feature detection and a 3D feature mapping step. Each of these will be evaluated independently here.

A. General training and evaluation approach

The 2D classification network for feature detection is trained using the projections as inputs and the binary images showing

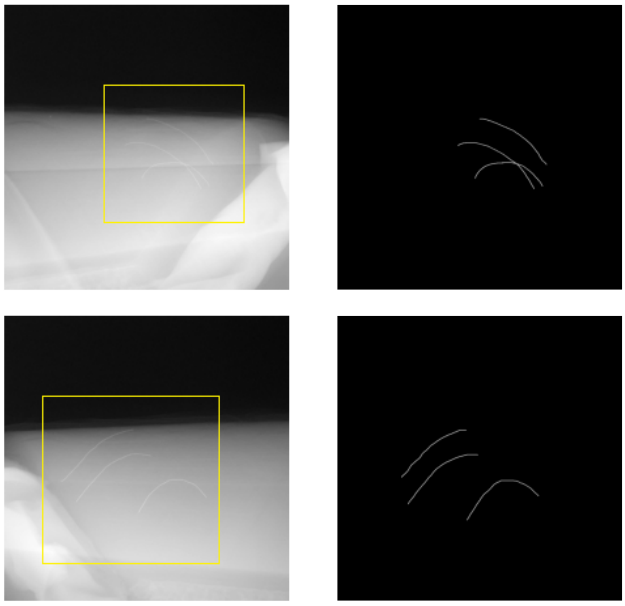

Fig. 7: A pair of x-ray stereo projection images from a carbon fibre layup experiment. Two projections are shown on the left and hand annotated images that show the feature location are shown on the right. We limited the evaluation of the confusion matrix into the yellow box to reduce the occurrences of non-feature pixels to better present evaluation results.

point locations as output. As for the 3D volumetric reconstruction experiment, based on the Eq. 2, the inputs are the filtered back-projection volume images generated from the feature maps detected with the feature detection network. Both networks are implemented using TensorFlow 2.0 and optimised using an Nvidia Titan XP graphics card. We use the Adam optimiser with a learning rate of 10^{-4} run for 100 epochs.

B. Feature Detection

Our first feature detection experiment uses the synthetic data-set. We split the data into test and train sets so that we use 95 of the volumes for training and the projections from the remaining 5 volumes for testing. Note that for each of these volumes, we have 10 different feature intensities for each volume and two projections from each, so that we have 1900 projections for training and 100 independently generated projections for testing.

Once trained, the performance of the method on 3 of the test samples (small blocks) is shown in Fig. 8, where we show predicted feature locations on the left, ground truth locations in the centre and the original projections on the right. To evaluate the performance numerically, we show the normalised confusion matrix in Table. I. We also show the ROC curve (receiver operating characteristic curve) in Fig. 9. Our test performance curve almost reaches the effect of the perfect model, indicated from the area under the curve (AUC) is close to 1. Note that we did not adjust the cost function to compensate for the fact that we have many more pixels without features than we have pixels with features. As expected, this here leads to a larger true negative rate compared to the true positive rate, something that could be addressed by taking the fraction of positive and negative samples into account in the cost function if required.

TPR = 93.7 %	FNR = 6.7 %
FPR = 0.1 %	TNR = 99.9 %

TABLE I: Confusion matrix on synthetic test data showing the percentage of pixels correctly (TPR) and incorrectly (FPR) identified as "a feature" vs pixels correctly (TNR) and incorrectly (FNR) identified as "not a feature".

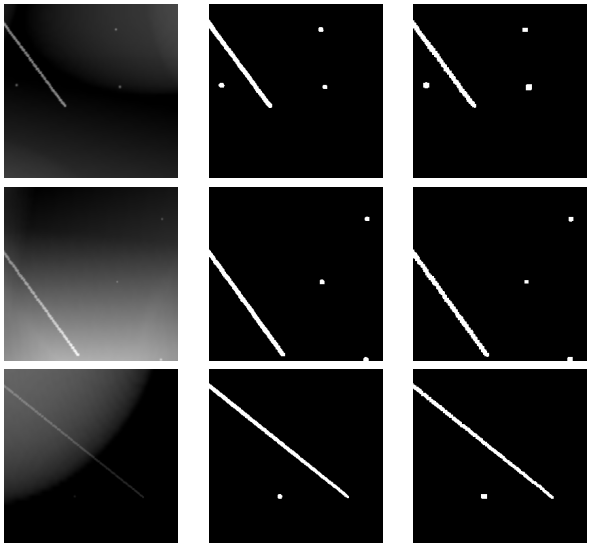

Fig. 8: Three test samples with features at different gray-levels. The first column shows the 2D projection, the second column is the estimated feature locations and the last column is the ground truth.

As our real X-ray projection images have 1000×1000 pixels, but relatively few training examples, we train the network on smaller image blocks. From the 120 projection images, we thus generate 11520 overlapping blocks of size 256×256 as our training set. We split the data into test and train sets by taking one projection direction as the training set (i.e. 60 full size images) and the test set as the other projection images. Example results are shown in Fig. 10, where we again show predicted feature locations (left), ground truth (middle) as well as projections (right). The confusion matrix is shown in Table.

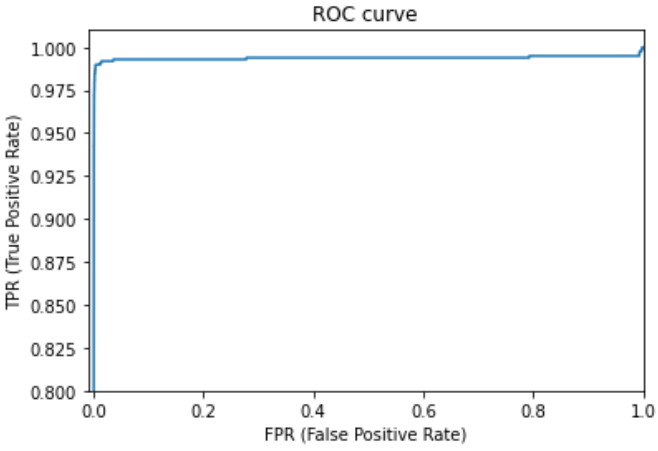

Fig. 9: ROC curve for our test data for the synthetic data-set. The AUC is 0.9937.

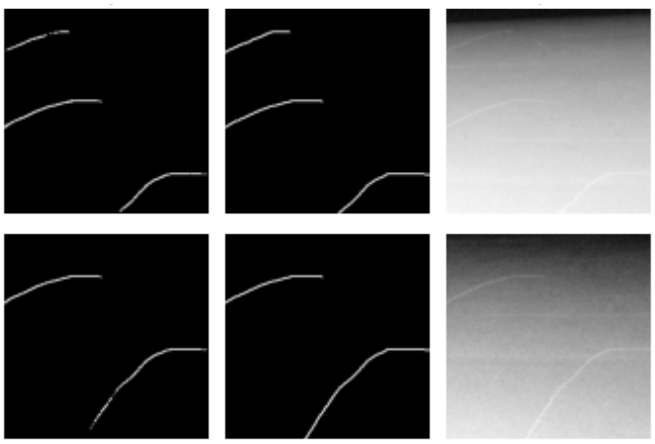

Fig. 10: Results on two test samples. The first column is the predicted feature location, the second column is the ground truth feature location, and the last column is the original projection image.

TPR = 93.2	FNR = 6.8%
FPR = 0.2 %	TNR = 99.8%

TABLE II: Confusion matrix on real test data showing the percentage of pixels correctly (TPR) and incorrectly (FPR) identified as "a feature" vs pixels correctly (TNR) and incorrectly (FNR) identified as "not a feature".

II and the ROC blue curve in Fig. 12.

To further explore the influence of the contrast of the features relative to the background attenuation in the projection images, we compare the performance of the model by training using modified data-sets that had features that were half or 1.5 times attenuating. Example images with differently strong features are shown in Fig. 11. Here we use 0.5 times to represent weak intensity, 1 times for normal intensity and 1.5 times for strong intensity. Classification performance is shown in Fig. 12, where we show the ROC curves for different feature intensities. As long as the features are not too weak, our model correctly detects the features.

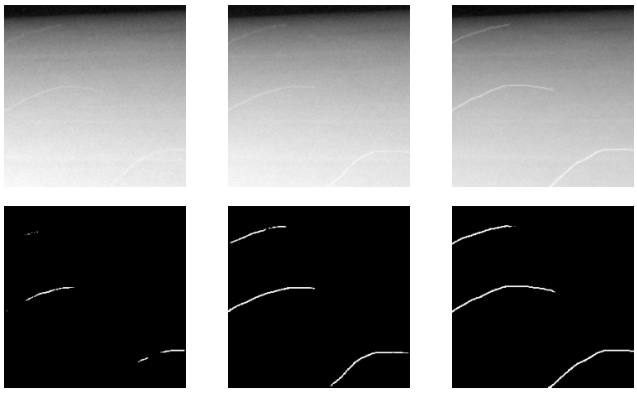

Fig. 11: A comparison of the model's performance with features of different levels of contrast. From left to right, we show features that are 0.5 times, 1 times and 1.5 times as strongly attenuating relative to the features in the original data. The top row shows the same block from one projection image but with different feature attenuation and the bottom row shows the detected features.

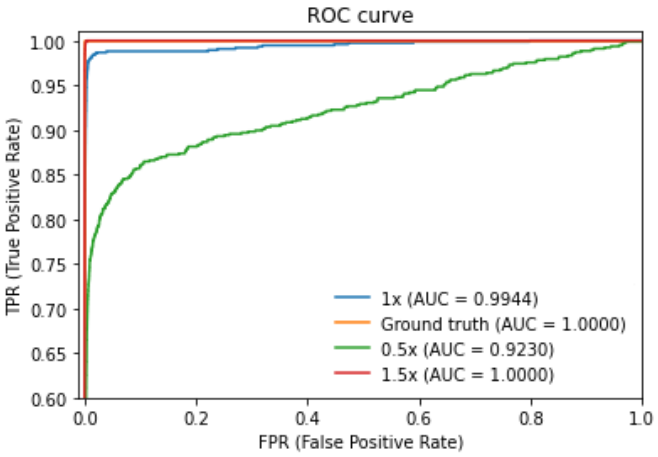

Fig. 12: ROC curve for the real data but with varying feature attenuation. The red curve shows features that are 1.5 times stronger than those in the original data, the blue curve is the original feature strength and the green curve is the performance if features have half the attenuation. The stronger features lead to perfect recovery, whilst the AUC for the less strong features drops to 0.9230.

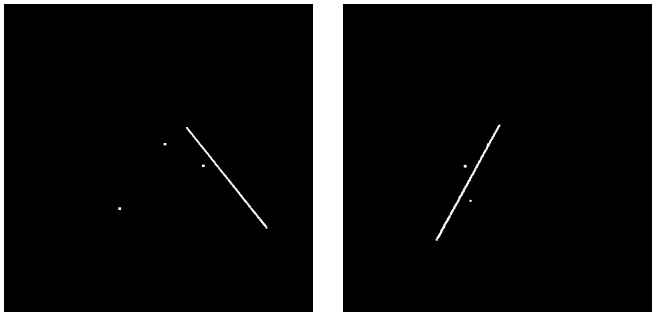

Fig. 13: A pair of projections from a synthetic test sample. The first view (left) shows all three point features as well as the line feature, whilst in the second view (right), a point feature overlaps with the line feature and is thus not visible.

C. 3D Feature mapping

To evaluate our approach to feature matching and 3D location estimation, we use the synthetic data-set only, as we do not have the ground truth for the real data. We train the 3D U-net on the estimated feature maps from the training data-set, using the ground truth 3D location of feature as the target. We then evaluate the method using the test data.

Fig. 13 shows an example where the geometry does not allow us to find unique 3D mappings, as in the second projection image, a point feature has been occluded by the line feature, so that the point feature from the first view can only be match with its corresponded point feature in the other view if we assume that two features overlap in one image. We find that even for this difficult case, the neural network can handle this uncertainty, with the rendering of the true and estimated 3D feature locations for this example shown in Fig. 14. We list 10 of our test samples with their average confusion matrix at Table. III and numerically evaluate the performance of this approach compare the absolute difference between the centre position of point features and line feature end point centers for the 10 test samples. We found that the average absolute difference is lower than 1.5 pixels.

To demonstrate the improved performance of our neural

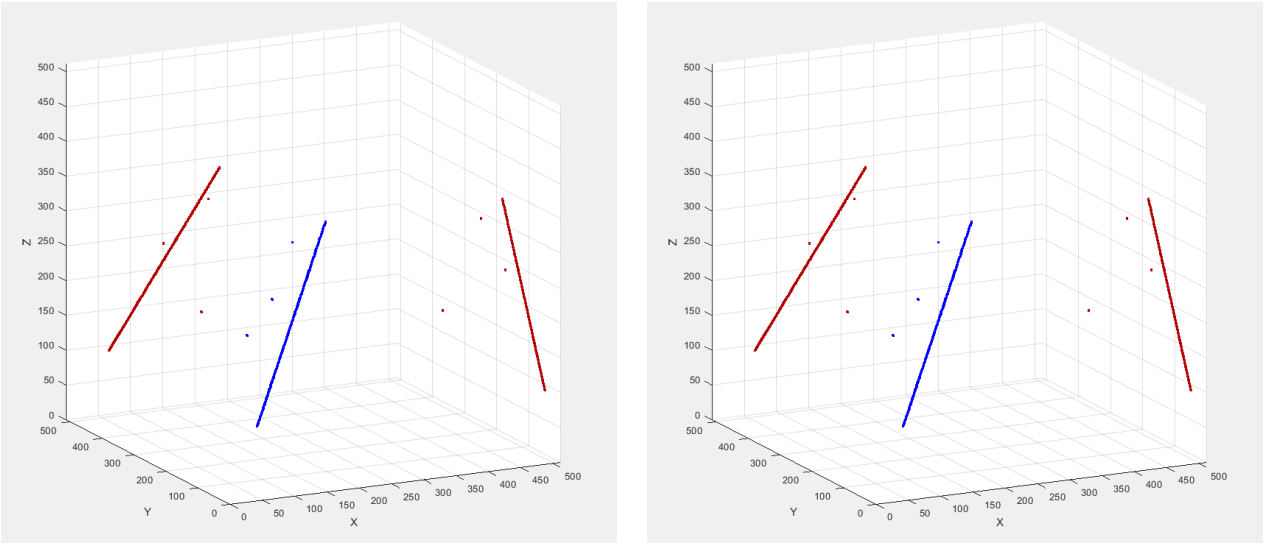

Fig. 14: Rendering of the 3D location of the features reconstructed from the estimated feature positions in the projections as shown in Fig. 13 (left), showing that the method still reconstructs the original feature locations (right), even though they are not all visible in both projection views.

TPR = 84.3 %	FNR = 15.7 %
FPR = 99.998 %	TNR = 0.002 %

TABLE III: Confusion matrix for 3D mapping result showing the percentage of pixels correctly (TPR) and incorrectly (FPR) identified in 3D space up to voxel resolution as "a feature" vs pixels correctly (TNR) and incorrectly (FNR) identified as "not a feature".

network approach, we also match the features of Fig. 13 and map them into 3D space via the epipolar geometric information as a reference. In this epipolar geometric method, the intersection point in 3D space of the back-projection generated by two different single views is the position of the feature point in 3D space, which is mathematically verifiable and accurate. However, considering the voxel and pixel size in 3D space and 2D projections in the digital image, there will inevitably be errors when selecting the center of the feature point. As shown in Fig 15, the absolute pixels error for the three points feature and two points feature at the end of line features is 1. Comparing Fig 15 with Fig. 14, our neural network method performs as well as the epipolar geometric method but crucially also works in some of the cases where the epipolar method fails due to point occlusion.

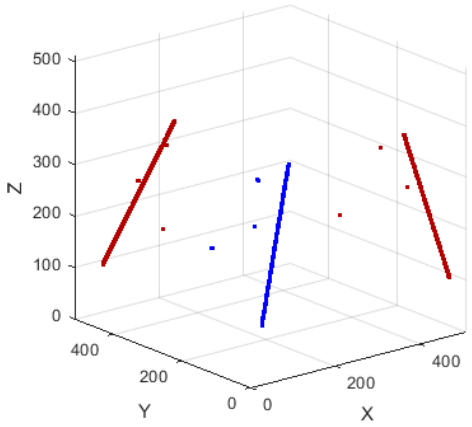

Fig. 15: The feature mapped from two feature detected projections into 3D space by epipolar geometric method with a replaced projection on second view due to a point feature is occluded by the line feature.

V. Discussion and Conclusions

In this paper, we introduce the concept of stereo X-ray tomography, where we use two (or possibly several) X-ray projection images to estimate spatial locations of features in 3D space. Whilst we are not able to reconstruct arbitrary objects in 3D without additional prior information, we show that it is possible to reconstruct the location of the point and line features in 3D. This can have many applications in tomographic imaging, especially when we are unable to collect full tomographic projections, which is useful when mapping dynamic processes that are too fast for full tomographic acquisition. For these imaging systems, setups with two (or few) X-ray source and detector systems can be envisaged that inspect an object at roughly 90-degree angles. In this setting, the main challenge is the identification and matching of points in the individual projections. Methods used in stereo vision, which typically rely on feature matching methods that match entire pixel neighbourhoods, do not work in transmission tomography. Instead, we propose the use of a learned feature detector together with a feature matching method that exploits epipolar geometry constraints. We have shown the robustness of the feature detection method and could demonstrate that, for problems with few features where unique matching is possible, a simple 3D U-net can map back-projected feature maps into 3D locations. It should be stressed that for two projection images (binocular Stereo CT), feature matching can become a challenge when using the epipolar constraint if we have larger numbers of features as matching is not uniquely possible if two features lie on the same epipolar plane in 3D. In this case, matching is however much more likely if we have 3

projections (triocular Stereo CT), where none of the epipolar planes are parallel.

VI. Acknowledgement

We would like to thank Pedro Galvez Hernandez from the University of Bristol for collecting the real data, https://research-information.bris.ac.uk/en/persons/pedro-galvez-hernandez and Nvidia for the provision of an Nvidia Titan GPU.

References

[1] J. Song, Q. H. Liu, G. A. Johnson, and C. T. Badea, "Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-ct," *Medical physics*, vol. 34, no. 11, pp. 4476–4483, 2007.

[2] H. LI, S. KAIRA, J. MERTENS, N. CHAWLA, and Y. JIAO, "Accurate stochastic reconstruction of heterogeneous microstructures by limited x-ray tomographic projections," *Journal of Microscopy*, vol. 264, no. 3, pp. 339–350, 2016. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/jmi.12449

[3] H. Yuan, J. Jia, and Z. Zhu, "Sipid: A deep learning framework for sinogram interpolation and image denoising in low-dose ct reconstruction," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1521–1524.

[4] H. Lee, J. Lee, H. Kim, B. Cho, and S. Cho, "Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 109–119, 2018.

[5] D. M. Pelt, O. Roche i Morgó, C. Maughan Jones, A. Olivo, and C. K. Hagen, "Cycloidal ct with cnn-based sinogram completion and in-scan generation of training data," *Scientific Reports*, vol. 12, no. 1, pp. 1–13, 2022.

[6] P. Ernst, S. Chatterjee, G. Rose, O. Speck, and A. Nürnberger, "Sinogram upsampling using primal-dual unet for undersampled ct and radial mri reconstruction," *arXiv preprint arXiv:2112.13443*, 2021.

[7] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.

[8] X. Li, S. Wang, P. Chen, and L. Wang, "3-d inspection method for industrial product assembly based on single x-ray projections," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021.

[9] B. Horn, B. Klaus, and P. Horn, *Robot vision*. MIT press, 1986.

[10] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*. SIAM, 2001.

[11] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.

[12] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[13] G. Lowe, "Sift-the scale invariant feature transform," *Int. J*, vol. 2, no. 91-110, p. 2, 2004.

[14] D. G. Viswanathan, "Features from accelerated segment test (fast)," in *Proceedings of the 10th workshop on image analysis for multimedia interactive services, London, UK*, 2009, pp. 6–8.

[15] K. G. Derpanis, "The harris corner detector," *York University*, vol. 2, 2004.

[16] Y. Sagara, A. K. Hara, W. Pavlicek, A. C. Silva, R. G. Paden, and Q. Wu, "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*, vol. 195, no. 3, pp. 713–719, 2010.

[17] E. J. Hoffman, S.-C. Huang, and M. E. Phelps, "Quantitation in positron emission computed tomography: 1. effect of object size." *Journal of computer assisted tomography*, vol. 3, no. 3, pp. 299–308, 1979.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[19] N. Ayache and F. Lustman, "Trinocular stereo vision for robotics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 1, pp. 73–85, 1991.

[20] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 424–432.

[21] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Josa a*, vol. 1, no. 6, pp. 612–619, 1984.

[22] W. Van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The astra toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.

[23] R. Arquier, I. Iliopoulos, G. Régnier, and G. Miquelard-Garnier, "Consolidation of continuous-carbon-fiber-reinforced paek composites: a review," *Materials Today Communications*, vol. 32, p. 104036, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352492822008893