SURG70004 Image Guided Intervention

IGI Coursework 2 – Medical Scene Understanding

Due date Friday 26th January 2024 16:00 (via Blackboard)

During this exercise you will investigate the 3D reconstruction of a surgical scene and you will focus on the detection of salient features on the tissue surface and the estimation of stereo feature correspondences. For this purpose, in vivo data from the SCARED dataset [1] will be used with known camera calibration parameters. In addition, you will explore volumetric visualisation methods and estimate volumetric renderings using the 3D scan of a zebrafish.

The questions are provided in black font. Please use blue for your answers in the spaces indicated.

Task 1 – Epipolar Geometry (20%)

Consider a camera which is located at (0,0,0) in an arbitrary coordinate system and it is facing in the direction (0,0,1); it captures an image I1 of the scene. Now consider that the camera position has been translated but its orientation remains the same. In its new position, the camera takes an image I2 of the scene. The camera translation is given by (5,0,0).

(a) Explain why points closer to the camera appear to move faster than those further away.
(b) Sketch the epipolar lines in I1 that correspond to features in I2 and label the location of the epipole in I1.
(c) Explain what is the projective interpretation of $e_r \times x_r$, the cross product of the epipole on the right camera and a point x_r on the image plane of the right camera.
(d) Describe how the structure of the scene determines the epipolar geometry.

Answer here (expand as necessary):

(a) In this pure translation context, assume that our camera translates by t and the image point x is normalised as $\mathrm{x} = (\mathrm{x}, \mathrm{y}, 1)^{\mathrm{T}}$. Attaching the world frame to the camera centre of the 1st view, the camera matrix for the views are $\mathrm{P} = \mathrm{K}[\mathrm{I}|0]$ and $\mathrm{P}' = \mathrm{K}'[\mathrm{R}|\mathrm{t}]$. Then from $\mathrm{x} = \mathrm{PX} = \mathrm{K}[\mathrm{I} \mid 0]\mathrm{X}$, the space point's (inhomogeneous) coordinates are $(X, Y, Z)^{\mathrm{T}} = ZK^{-1}x$, where Z is the depth of the point X (the distance of X from the camera centre measured along the principal axis of the first camera). It then follows from $\mathrm{x}' = \mathrm{P}'\mathrm{X} = \mathrm{K}[\mathrm{I} \mid \mathrm{t}]\mathrm{X}$ that the mapping from an image point x to an image point x′ is

$$\mathrm{x}' = \mathrm{x} + \mathrm{Kt}/\mathrm{Z}. \qquad \textit{Equation 1}$$

The motion Equation 1 shows that the image point "starts" at x and then moves along the line defined by x and the epipole $\mathrm{e} = \mathrm{e}' = \mathrm{Kt}$. The extent of the motion depends on the magnitude of the translation t (fixed in this context) and the inverse depth Z, so that points closer (smaller Z) to the camera appear to move faster than those at greater depth.

(b) In this case, the translation is parallel to the image plane. The baseline intersects the image plane at infinity. Consequently the epipoles are at infinity, and epipolar lines are parallel. The sketch is shown in Figure 1.

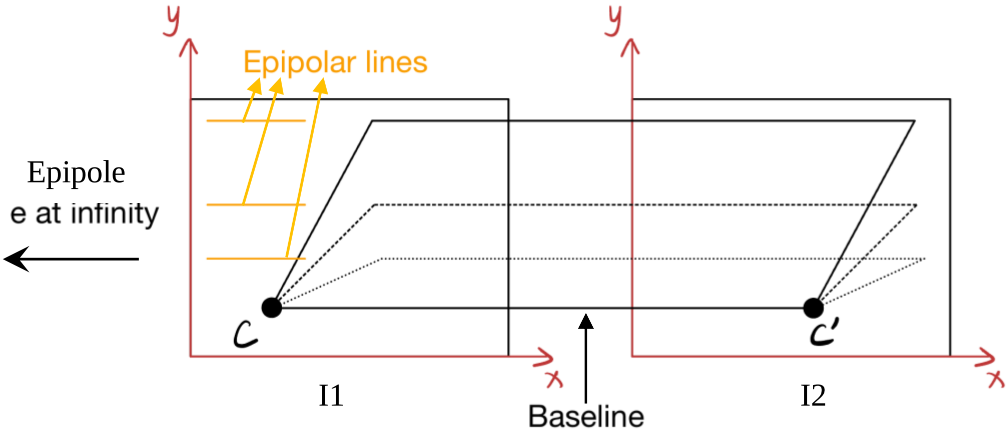

Figure 1: Sketch of epipolars lines and epipole

(c) The projective interpretation of $e_r \times x_r$ is the epipolar line l′ (right camera) passing through point x_r and the epipole e_r. In geometric derivation, given the point x_r, the epipolar line l′ passing through x_r and the epipole e_r can be written as $l' = e_r \times x_r$.

(d) The epipolar geometry is the intrinsic projective geometry between two views. It is independent of scene structure, and only depends on the cameras' internal parameters and relative pose.

Task 2 – Instrument Tracking in the OR (15%)

An endovascular procedure is performed under live X-ray.

(a) Propose a tracking technology that would be suitable to track the position of catheters within the aorta. Justify your choice.
(b) Sketch and label the different components (equipment, localisation systems, etc.) required in the setup.
(c) Select and justify your choice of reference coordinate frame. Write out the series of frame transformations required to transform all other coordinate frames to the reference.
(d) List the challenges of using a 3D preoperative model to guide an endovascular intervention.

Answer here (expand as necessary):

(a) Considering that the endovascular procedure is performed under live X-ray, a suitable catheter tracking technology is **fluoroscopy**. Here's the justification for this choice:

Real-time Anatomical Imaging: Fluoroscopy provides real-time X-ray imaging of the aorta, allowing continuous visualisation of the catheter's position.

Established and Widely Used: It is well-established, widely used and enables dynamic guidance throughout the procedure.

Non-invasiveness: It is non-invasive that typically doesn't require additional markers on the catheter. This minimises the intrusion into the vascular system, simplifies the setup, and reduces the risk of introducing foreign objects.

Compatibility: Fluoroscopic imaging is intrinsic to X-ray systems commonly used in endovascular procedures, making it a seamless and readily available tracking technology.

Additionally, the integration of fluoroscopy with other technologies (e.g. electromagnetic tracking) can further enhance spatial information.

(b) Figure 2 shows the sketch of Fluoroscopy catheter tracking system. The Key Components of X-Ray Fluoroscopic System: X-ray generator with X-ray tube, Spectral shaping filters, Collimator, Anti-scatter grid, Image Receptor — Flat Panel Detector (FPD) and Monitor.

X-ray tubes expose radiation either in a continuous or pulsed manner. A collimator limits the extent of the X-ray field. The filtration elements allow choosing between low and high dose modes during the procedure. Anti-scatter grids are used when the procedure is performed in high-scattered areas. The image receptor can be either an x-ray image intensifier or a Flat Panel Detector (FPD). In this proposed system, FPD is selected for its heightened sensitivity to X-rays, presenting the potential to decrease patient radiation exposure. As an optional choice, the electromagnetic tracking system could be set near the patient table to provide additional spatial information. Essentially, it includes equipment such as electromagnetic field generator, sensors on the catheter. In addition, a contrast agent (e.g. iodine-based) could be injected to enhance visibility of blood vessels in X-ray.

The Fluoroscopy System Configuration is based on angiographic systems. It employs a "C-arm" geometry to enable easy patient access as fluoroscopy guides selective arterial catheter placement. This system includes advanced features like digital subtraction and road mapping.

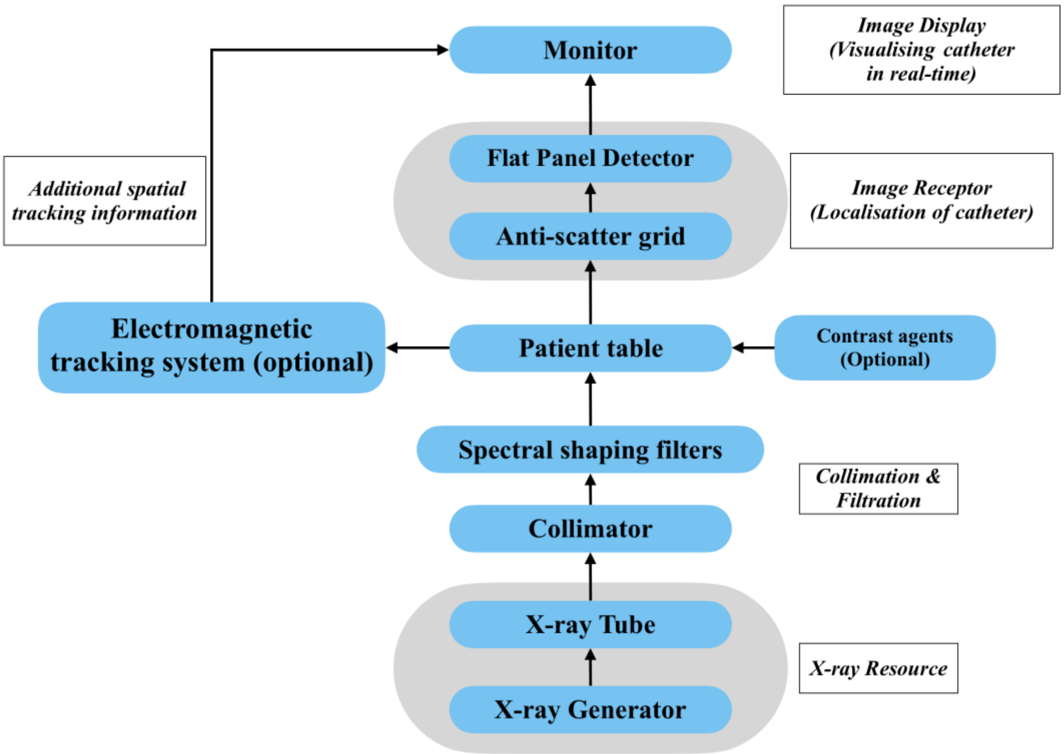

Figure 2: The sketch of Fluoroscopy Catheter Tracking System

(c) Choice of Reference Coordinate Frame:

The Patient Anatomical Frame (PAF) should be chosen as the reference coordinate frame. The justification for this lies in its direct correlation with the clinical objectives of the procedure. By using the patient's anatomy as a reference, the tracking system can provide the most relevant and accurate information regarding the catheter's position relative to the targeted anatomical structures, such as the aorta. This choice ensures that all movements and positions of the catheter are directly mapped and understood in the context of the patient's body, which is crucial for the precision and safety of the endovascular procedure.

Series of Frame Transformations:

- From Fluoroscopy Image Frame (FI) to Patient Anatomical Frame (PA):

$$p_{PA} = {}^{PA}F_{FI} p_{FI}$$

Where $${}^{PA}F_{FI} = \begin{bmatrix} {}^{PA}R_{FI} & {}^{PA}t_{FI} \\ 0 & 1 \end{bmatrix}.$$

- From Electromagnetic Tracking Frame (ET) to Patient Anatomical Frame (PA): If an electromagnetic tracking system is used

$$p_{PA} = {}^{PA}F_{ET} p_{ET}$$

Where $${}^{PA}F_{ET} = \begin{bmatrix} {}^{PA}R_{ET} & {}^{PA}t_{ET} \\ 0 & 1 \end{bmatrix}.$$

- From Catheter Coordinate Frame (CC) to Electromagnetic Tracking Frame (ET) to Patient Anatomical Frame (PA):

  First, the transformation from the CC to ET, ${}^{ET}F_{CC}$, is applied. This accounts for the position and orientation of the catheter within the electromagnetic field.

$$p_{ET} = {}^{ET}F_{CC}p_{CC}$$

  Where
$${}^{ET}F_{CC} = \begin{bmatrix} {}^{ET}R_{CC} & {}^{ET}t_{CC} \\ 0 & 1 \end{bmatrix}.$$

  Then, the point in the ET is transformed to the PA using the previously defined transformation ${}^{PA}F_{ET}$.

$$p_{PA} = {}^{PA}F_{ET}p_{ET} = {}^{PA}F_{ET}\,{}^{ET}F_{CC}p_{CC}$$

These series of transformations are critical to accurately map the catheter's movements and positions to the patient's anatomical frame.

(d) Challenges are listed as following.

a. Accuracy of the Model: A key challenge is ensuring the accuracy of the 3D preoperative model in representing the patient's current anatomy. Anatomical structures can vary between individuals and may change over time due to disease progression, prior interventions, or natural physiological changes.

b. Image Registration: Aligning the 3D preoperative model with the patient's actual anatomy during the procedure can be challenging. It requires precise calibration and may be complicated by patient movement, respiratory motion, or changes in the body position during the procedure.

c. Limited Real-time Information: A preoperative model is static and does not account for real-time physiological changes, such as blood flow dynamics, heartbeat, and vessel movement. Adapting the static model to reflect these dynamic changes in real-time is a significant challenge.

d. Data Integration Complexity: Integrating data from the 3D model with intraoperative imaging (like fluoroscopy) and other tracking technologies requires sophisticated software and hardware. Ensuring seamless and accurate integration for real-time guidance is technically challenging.

e. Cost: Developing, maintaining, and using advanced 3D modelling technology can be expensive.

f. Radiation Exposure: The use of additional imaging modalities, such as preoperative CT scans, may contribute to increased radiation exposure for the patient and medical staff.

Task 3 – 3D Scene Reconstruction (40%)

The following are example frames from a video sequence.

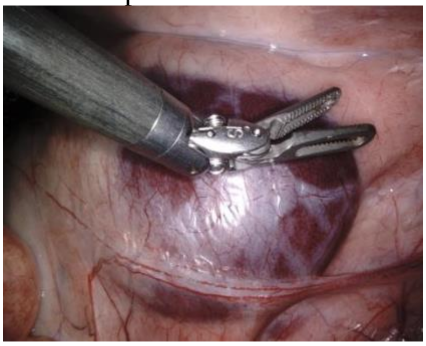
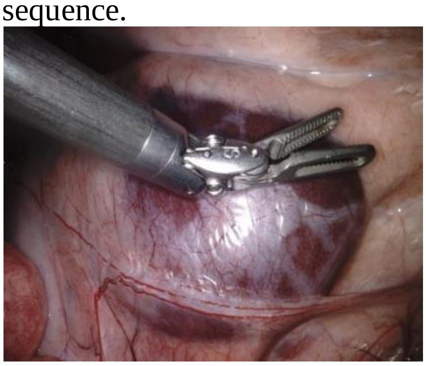

Figure: Stereo images included in the dataset. (Left) Left camera frame (Right) Right camera frame.

Use a programming environment of your choice and the provided stereo images to:

a. Detect salient features and plot the detected features on the provided pair of frames.
b. Find corresponding features between the two frames. Create a composite image (e.g. centred overlay image) from the two frames and demonstrate the matched points.
c. Use the matched features to estimate the fundamental matrix between the two images.
d. Find the correctly matched points that meet the epipolar constraint.
e. Estimate the diameter of the shaft of the tool shown in the stereo images. (hint: you can establish the disparity map between the left and right camera images or you can apply 3D surface reconstruction)

For each of the above tasks, provide a brief explanation about the approach you followed to estimate your results.

Answer here (expand as necessary):

a. The RGB stereo images (original) first need to be transferred into grey image. Then, to detect salient features in a pair of stereo images, MATLAB functions ‘detectSURFFeatures’ has been used. These features are typically corners, edges, or distinct blobs in the image. Finally, the top 50 features (points) with strongest metrics are selected and plotted on the original images, shown in Figure 3.

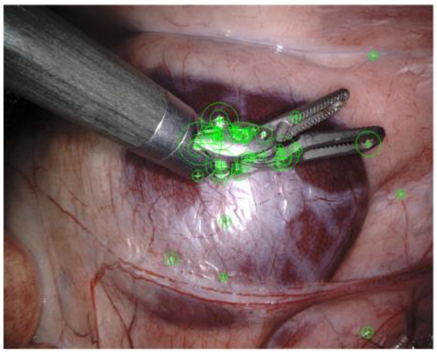
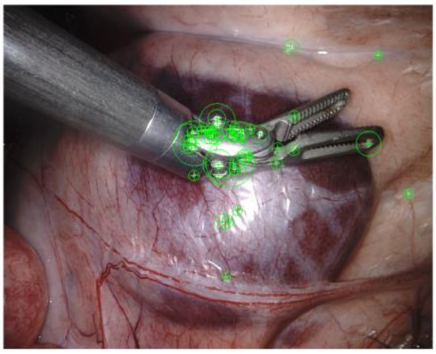

Figure 3: Salient features on the stereo images. (Left) Left camera frame (Right) Right camera frame.

b. After detecting the features, to find corresponding features, feature descriptors and a matching process have been used. The ‘extractFeatures’ function is used to obtain descriptors for each feature point, and ‘matchFeatures’ finds the best matches between the two sets of descriptors. The result is visualised by creating a composite image showing lines connecting the matched features across the two images, shown in Figure 4. The composite image is created by function ‘showMatchedFeatures’.

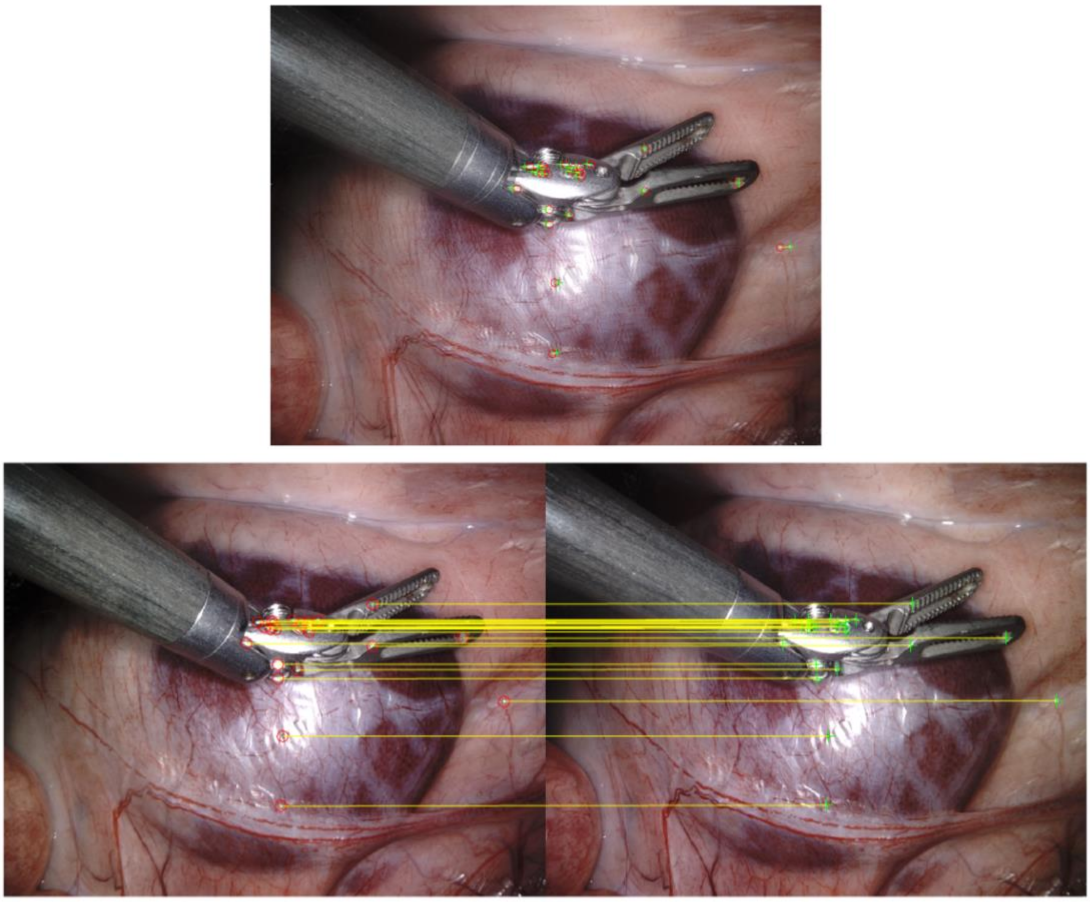

Figure 4: Composite image in two styles. (Top) Centred Overlay Image (Bottom) Side-by-Side Montage.

c. The fundamental matrix relates corresponding points between stereo images in terms of their epipolar geometry. The fundamental matrix is estimated using the ‘estimateFundamentalMatrix’ function in MATLAB, which takes matched points from the stereo images and computes the matrix. This matrix is vital for tasks like rectification and 3D reconstruction. The estimated fundamental matrix is shown below.

$$\begin{bmatrix} 7.6946e-7 & -1.377e-4 & 0.0510 \\ 1.372e-4 & 3.2218e-6 & -0.1005 \\ -0.0516 & 0.0972 & 0.9876 \end{bmatrix}$$

d. The fundamental matrix helps in identifying and filtering out incorrect matches. Points that adhere to the epipolar constraint imposed by the fundamental matrix are likely to be correct matches. MATLAB's estimateFundamentalMatrix function identifies the inlier points, which sets the elements of the vector to true when the corresponding point was used to compute the fundamental matrix. By using these true values in the matched features, the correctly matched points will be extracted, shown in Figure 5.

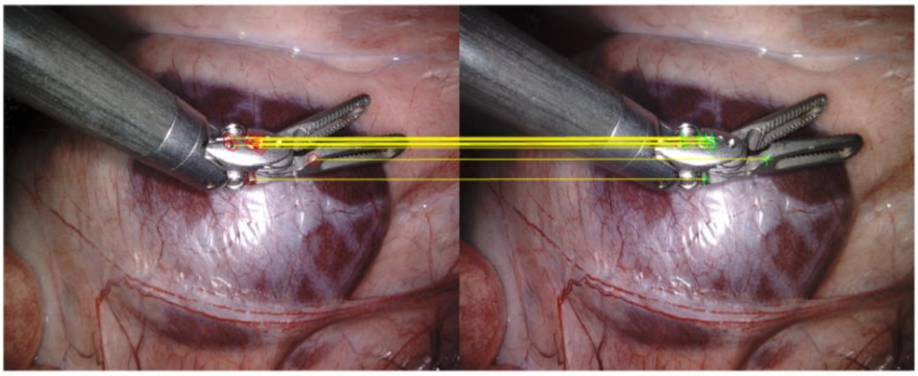

Figure 5: Composite image with correctly matched points in Montage style.

e. Begin by setting the stereo camera parameters, rotation, and translation matrices from provided data. Utilise ‘rectifyStereoImages’ to rectify images, followed by creating a disparity map with disparitySGM. The input parameters, crucial for accurate 3-D coordinates, include 'DisparityRange' [104,128] and 'UniquenessThreshold' of 15. Figure 6 displays this disparity map, highlighting the tool's shaft in red.

   Geometric analysis involves overlaying boundary and perpendicular lines on the map, allowing the estimation of the shaft's diameter by comparing points along these lines that fall within the highlighted red regions.

   Finally, reconstruct the 3D scene with ‘reconstructScene’, using point pairs from the geometric analysis. This process involves calculating the distance between point pairs derived from the geometric analysis. The findings, illustrated in Figure 7, reveal that the average diameter, as calculated from the 3D data (x, y, z), is 8.16mm. However, when calculated using only the x and y data, the diameter measures at 7.53mm. Given that the shaft resembles a cylinder and the Z-depth has minimal impact in this context, the x and y 3-D world point coordinates are deemed more precise. Therefore, the shaft's diameter is established at 7.53mm.

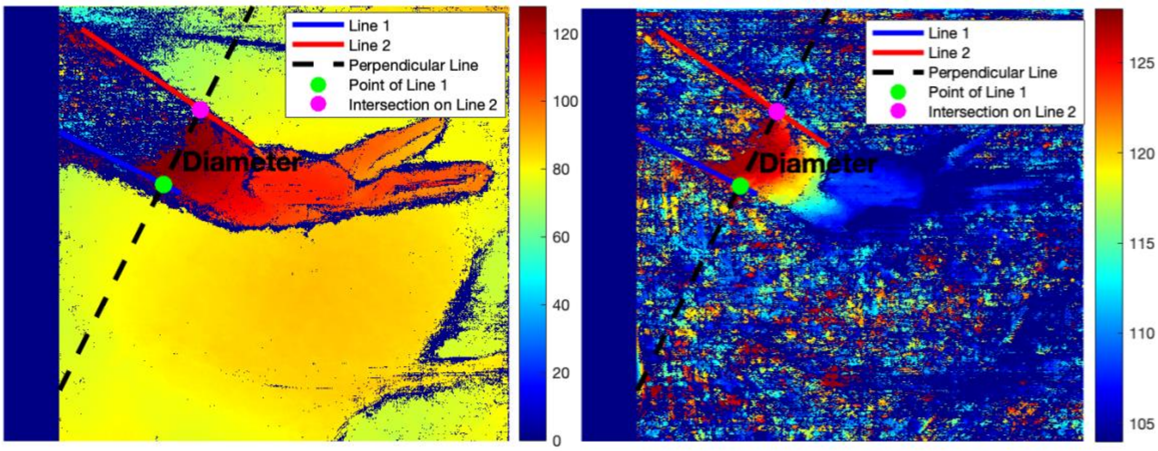

Figure 6: Disparity map with geometry analysis. (Left)DisparityRange [0,128]

(Right) DisparityRange [104,128].

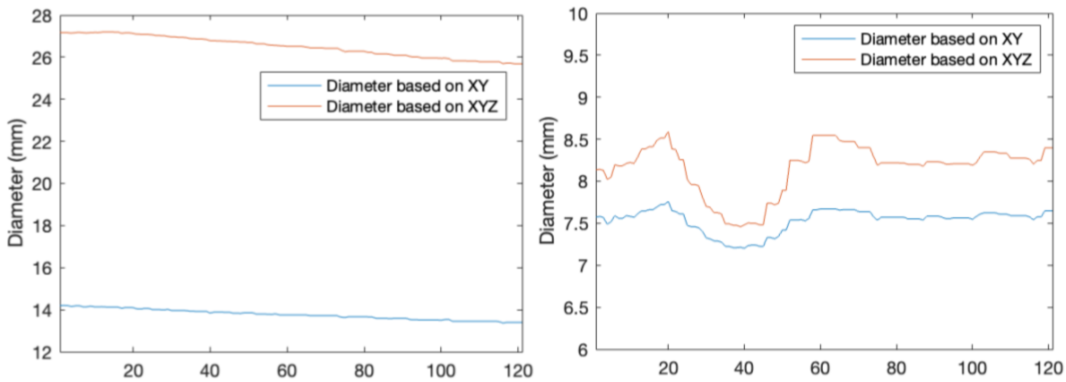

Figure 7: Shaft diameter calculated from the 3D data (x, y, z) and (x,y) along boundary lines within highlighted red regions. (Left)DisparityRange [0,128] (Right) DisparityRange [104,128].

Task 4 - Visualisation (25%)

In the provided dataset, you can find the optical projection tomography volumetric data of a zebrafish.

a. Draw a set of transfer functions that can be used to create a rendering of a CT dataset of the abdomen. Assume that the signal in CT units from bone is >80, that from liver is 25-50 and that the signal from fat is 10-25. The tissues appear semi-realistic, i.e. bone should be white and opaque, liver should be dark red and opaque, fat should be white and semi-transparent.

b. In the provided coursework datasets, you can find the optical projection tomography volumetric data of a zebrafish. Implement the MIP method and apply it to project the volumetric data along the 3 orthogonal directions (x, y and z-axis). Display each projection plane of your rendering.

Answer here (expand as necessary):

a. Table 1 shows the colour and opaque setup. Figure 8 shows the set of transfer functions.

Table 1: Colour and opaque

	Red	Green	Blue	Opacity	CT units	Comment
Bone	1	1	1	1	>80	white, opaque
Liver	0.5	0	0	1	25-50	dark red, opaque
Fat	1	1	1	0.5	10-25	white, semi-transparent

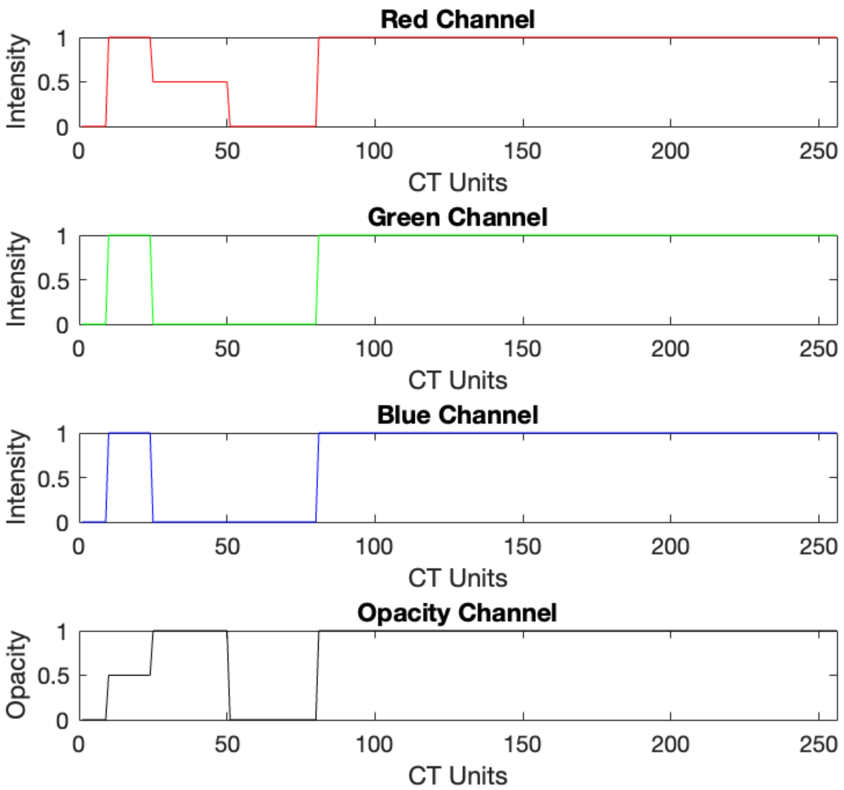

Figure 8: Drawing of transfer functions

b. Figure 9 shows the projection along the 3 orthogonal directions

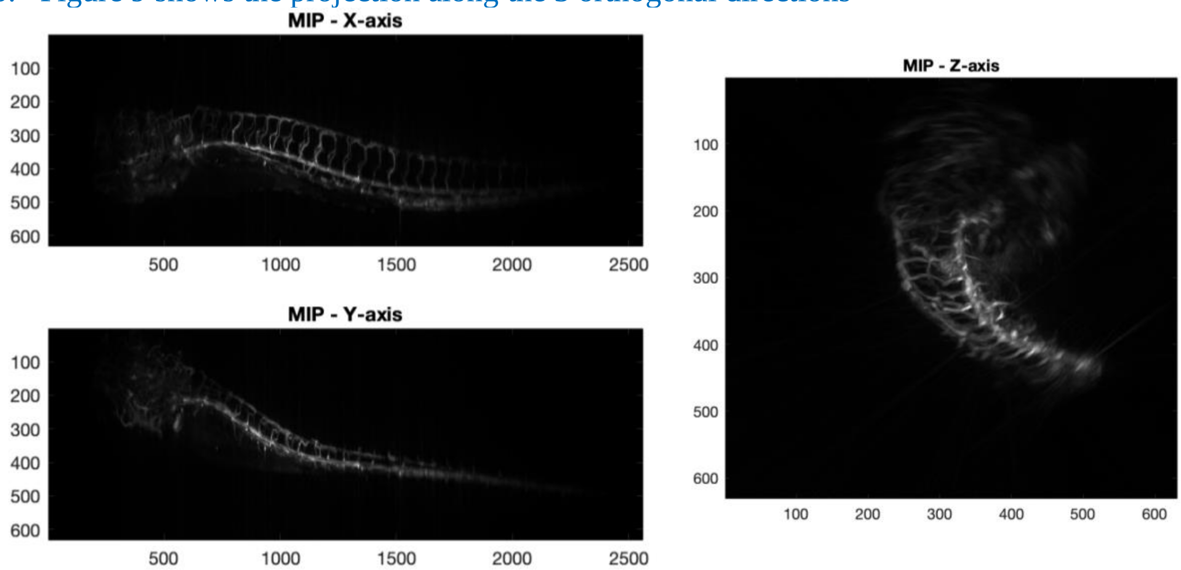

Figure 9: Projection based on MIP method

References

[1] http://hamlyn.doc.ic.ac.uk/vision/