Project 2: 3D Reconstruction

Hayleigh Sanders (RIN#661195735)

1. Introduction

The following report investigates 3D reconstruction, the calculation of the fundamental matrix, and image rectification.

2. Theories for computing F

2.1 Individual camera calibration

Theory

In a stereo system, performing calibration on both cameras separately will yield W_l, R_l, and T_l for the left camera and W_r, R_r, T_r for the right camera. The relative rotation matrix R and relative translation matrix T can be found with the following relationship between the extrinsic parameter matrices of each camera:

$$R = R_l R_r^t \quad (1)$$

$$T = T_l - RT_r \quad (2)$$

A stereo system must follow the epipolar constraint. Given a 2D point in the left image, suppose a 3D line goes through this point to the camera center. The projection of this 3D line onto the right image is the epipolar line and the corresponding 2D point in the right image must lie on this line. The relationship between two corresponding 2D points in the right and left camera frames, P_l and P_r, are constrained by the relative orientation of the stereo system:

$$P_l = RP_r + T \quad (3)$$

P_l, P_r, and T must follow the coplanar constraint because P_l, P_r, and their corresponding 3D point in the object frame P must all lie on the epipolar plane. Therefore P_l and T are related by:

$$\left(T \times P_l\right)^t \left(P_l - T\right) = 0 \quad (4)$$

Since (3) can be rearranged to $P_l - T = RP_r$, (4) becomes:

$$\left(T \times P_l\right)^t RP_r = 0 \quad (5)$$

Because a vector product can be expressed as a multiplication by a rank-deficient matrix, T can be rearranged to S, which is given as:

$$S = \begin{pmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{pmatrix} \quad (6)$$

So (6) can be substituted into (5) to give:

$$\left(T \times P_l\right) = SP_l \qquad (7)$$

Finally, the coordinates of two corresponding 2D points in a stereo system are related by the relative orientation of the system:

$$P_l{}^t E P_r = 0 \quad (8)$$

Where E is the essential matrix, which establishes a link between the epipolar constraint and relative orientation. E is given by:

$$E = S^t R \qquad (9)$$

The fundamental matrix relates the intrinsic parameters of each camera and the relative orientation of the stereo system. It can be expressed as a product of the intrinsic matrix of each camera and the essential matrix:

$$F = W_l{}^{-t} E W_r{}^{-1} \qquad (10)$$

Substituting (6) into (9), (10) can be expanded to:

$$F = W_l{}^{-t} \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} R W_r{}^{-1} \qquad (11)$$

The relative rotation and translation matrices R, T can be computed from R_r, R_l, T_r, and T_l using equations (1) and (2). Therefore, from (11) it is clear that the fundamental matrix can be computed directly from W_l, R_l, and T_l for the left camera and W_r, R_r, T_r for the right camera.

Results

Each camera was calibrated using the linear method to produce a parameter matrix P, from which the values for matrices W, R and T can be extracted.

The left camera had the following values:

$$W_l = \begin{bmatrix} 1.61\cdot 10^3 & 0 & 1.05\cdot 10^3 \\ 0 & 1.56\cdot 10^3 & 6.33\cdot 10^2 \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

$$R_l = \begin{bmatrix} 0.5 & -.86 & -.008 \\ .027 & .036 & -1.03 \\ -.83 & -.49 & .22 \end{bmatrix} \quad (13)$$

$$T_l = \begin{bmatrix} -9.56 & 19.4 & -121.47 \end{bmatrix} \quad (14)$$

The right camera had the following values:

$$W_r = \begin{bmatrix} 1.48\cdot 10^3 & 0 & 9.81\cdot 10^3 \\ 0 & 1.44\cdot 10^3 & 6.8\cdot 10^2 \\ 0 & 0 & 1 \end{bmatrix} \quad (15)$$

$$R_r = \begin{bmatrix} -.83 & .55 & -.003 \\ .038 & .057 & 1.01 \\ .53 & .799 & -.269 \end{bmatrix} \quad (16)$$

$$T_r = \begin{bmatrix} 22.2 & -13.4 & 109.07 \end{bmatrix} \quad (17)$$

Using equation (1), (2), and (11), the fundamental matrix was calculated to be:

$$F = \begin{bmatrix} 5.19\cdot 10^{-6} & 1.54\cdot 10^{-5} & -3.82\cdot 10^{-2} \\ -1.51\cdot 10^{-6} & 3.54\cdot 10^{-6} & -4.08\cdot 10^{-2} \\ 1.35\cdot 10^{-2} & 2.23\cdot 10^{-2} & 1.54\cdot 10^{1} \end{bmatrix} \quad (18)$$

2.2 Joint calibration with the eight point algorithm

An alternative method to compute the fundamental matrix is through a system of linear equations formed from the relationship between left and right 2D point correspondences and the fundamental matrix. A 2D point in the left view U_l and corresponding 2D point in the right view U_r are related to the fundamental matrix F by the following:

$$U_l^t F U_r = 0 \quad (19)$$

Each row of F can be expressed as F_1, F_2 and F_3 so that (19) can be expressed as:

$$U_l^t F_1 c_r + U_l^t F_2 r_r + U_l^t F_3 = 0 \quad (20)$$

Suppose V is a 9x1 vector containing each value of F. Thus (20) can be rearranged as a product of V.

$$\left(U_l^t c_r \quad U_l^t r_r \quad U_l^t \right) V = 0 \quad (21)$$

Each pair of 2D points form a linear equation expressed by (21). Therefore eight 2D points form eight equations in a system of linear equations AV = 0, where V is a 9x1 vector representing each value of F and A is given by:

$$A = \begin{pmatrix} U_{l_1}^t c_{r_1} & U_{l_1}^t r_{r_1} & U_{l_1}^t \\ \vdots & \vdots & \vdots \\ U_{l_N}^t c_{r_N} & U_{l_N}^t r_{r_N} & U_{l_N}^t \end{pmatrix} \quad (22)$$

Taking the singular value decomposition of A will yield $UD(S^t)$ where D is a diagonal matrix containing the singular values of A in descending order, and U and S are orthonormal matrices. The last column of S is equivalent to the null vector of A, and as a result is equal to V. Therefore each value of F can be recovered from V.

Results

Using singular value decomposition on AV = 0, where the A matrix is formed from eight randomly selected corresponding 2D points in equation (22), the following fundamental matrix was obtained:

$$F = \begin{bmatrix} 2.115 \cdot 10^{-6} & 3.26 \cdot 10^{-6} & -3.53 \cdot 10^{-3} \\ -4.97 \cdot 10^{-6} & 1.94 \cdot 10^{-7} & 1.39 \cdot 10^{-3} \\ 5.97 \cdot 10^{-4} & 5.24 \cdot 10^{-4} & 9.99 \cdot 10^{-1} \end{bmatrix} \quad (23)$$

Discussion

The fundamental matrix obtained from separate calibration is different than the one obtained from the eight point algorithm. The eight point algorithm is more susceptible to 2D point error, so the fundamental matrix and subsequent parameters obtained from it will have more numerical instability due to noise in the essential matrix. Therefore, the fundamental matrix and camera parameters derived from the separate calibration procedure will be used for the following image rectification and 3D reconstruction tasks.

3. Image Rectification

Theory

Two stereo images can be rectified by rotating them such that the image planes are parallel to the base line and their conjugate epipolar lines are collinear. Once the intrinsic and extrinsic parameters of both cameras are known, the rotation matrices needed to rectify them can be computed.

For the left camera, a rotation matrix R_l can be applied so that the camera frame is parallel to the base line. The unit vector of the relative translation with respect to the left camera is defined by:

$$\frac{T}{\| T \|} = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \quad (24)$$

Suppose $(1\ 0\ 0)^t$ is the x axis after rotation. Because $R_l T/(\|T\|) = (1\ 0\ 0)^t$, the first row of R_l is defined as:

$$r_{l_1} = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}^t \quad (25)$$

Because the rows of R_l must be orthogonal, the second row can be defined as:

$$r_{l_2} = \frac{1}{\sqrt{t_x^2 + t_y^2}} \begin{pmatrix} t_y \\ -t_x \\ 0 \end{pmatrix}^t \quad (26)$$

Then the third row will be:

$$r_{l_3} = r_{l_1} \times r_{l_2} \quad (27)$$

Suppose $U_l = (c_l, r_l)$ is a 2D point in the left camera image. Its corresponding point in the rectified left camera image $U_l' = (c_l', r_l')$ can be mapped from the original point using R_l and the intrinsic camera matrices with:

$$\lambda \begin{pmatrix} c_l' \\ r_l' \\ 1 \end{pmatrix} = W_l R_l W_l^{-1} \begin{pmatrix} c_l \\ r_l \\ 1 \end{pmatrix} \quad (28)$$

Because this method of forward mapping is known to create gaps in the rectified image, the rectified image can be created through "backward mapping" with equation (29) below. Each coordinate of the rectified image (c_l', r_l') can be multiplied by the rectification transform to produce the corresponding coordinates (c_l, r_l) in the original image. The RGB value at (c_l, r_l) will equal the one at (c_l', r_l'). Therefore one can iterate through each pixel of the rectified image and find its value from the original image to avoid gaps.

$$\lambda \begin{pmatrix} c_l \\ r_l \\ 1 \end{pmatrix} = W_l R_l^t W_l^{-1} \begin{pmatrix} c_l' \\ r_l' \\ 1 \end{pmatrix} \quad (29)$$

The rotation matrix for the right camera is given as:

$$R_r = R_l R \quad (30)$$

Therefore points from the original right image can be mapped to the rectified image using backward mapping with the following:

$$\lambda \begin{pmatrix} c_r \\ r_r \\ 1 \end{pmatrix} = W_r R_r^t W_r^{-1} \begin{pmatrix} c_r' \\ r_r' \\ 1 \end{pmatrix} \quad (31)$$

Results

Applying equation (29), the rectified left image was found in Figure 1:

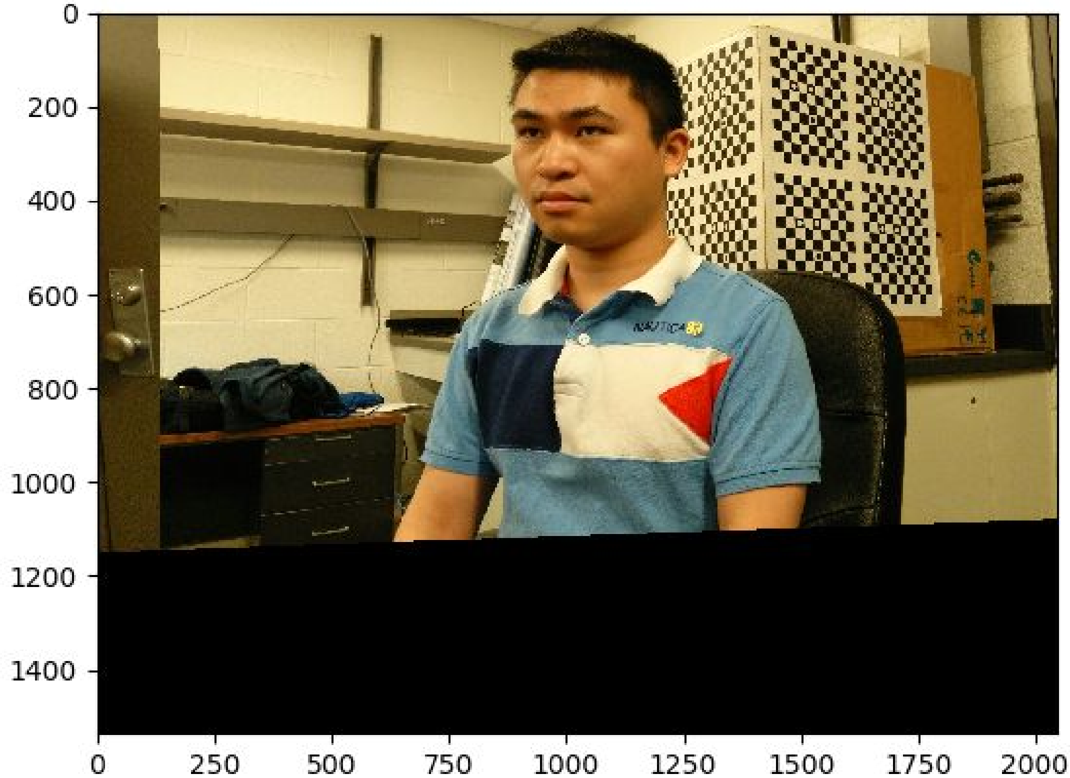

Figure 1: Rectified left image

Applying equation (31), the right rectified image was found in Figure 2:

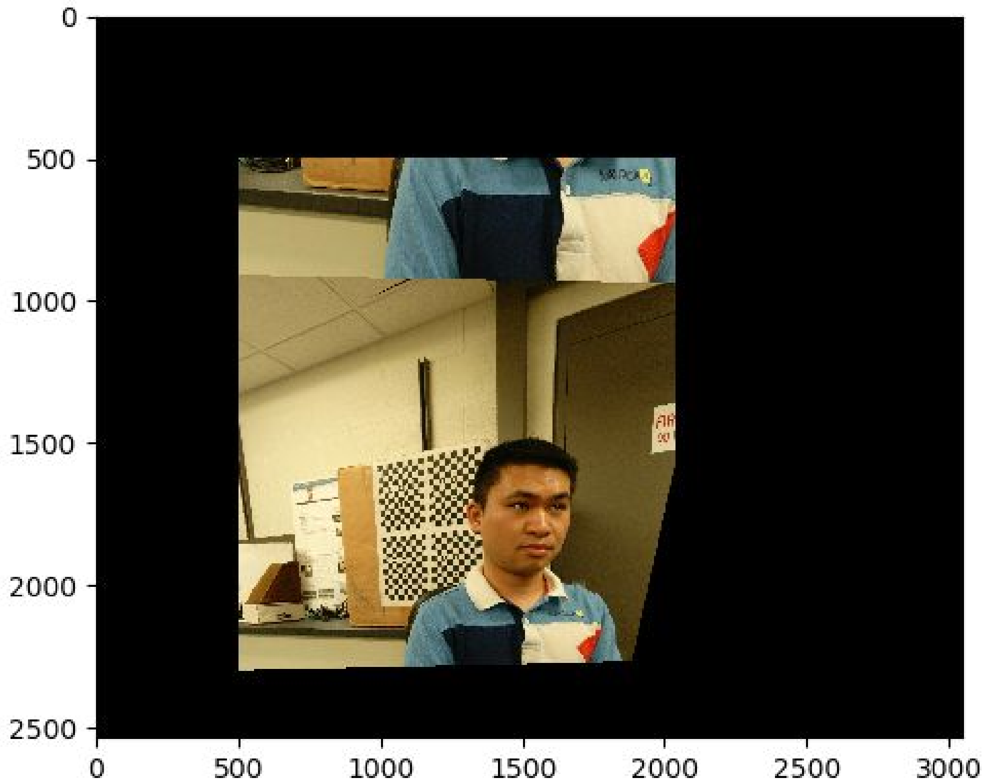

Figure 2: Rectified right image

The epipoles were calculated from F. The right epipole was calculated from the null eigenvector of F, and the left epipole was calculated from the null eigenvector of F^t. The null eigenvector of each matrix was computed using singular value decomposition. The epipoles were found to be:

$$e_r = [2.19 \cdot 10^3, 1.9 \cdot 10^4] \quad (32)$$
$$e_l = [7.55 \cdot 10^2, -6.94 \cdot 10^3] \quad (33)$$

The epipolar lines were found for each facial feature point by multiplying their 2D coordinates by the fundamental matrix. This yielded a vector of coefficients (a,b,c) for the line equation $ac_r+br_r+c=0$. Each epipolar line was drawn from the facial feature point to its x-intercept. Figures (3) and (4) show the epipolar lines drawn from the facial feature points in the left and right original images.

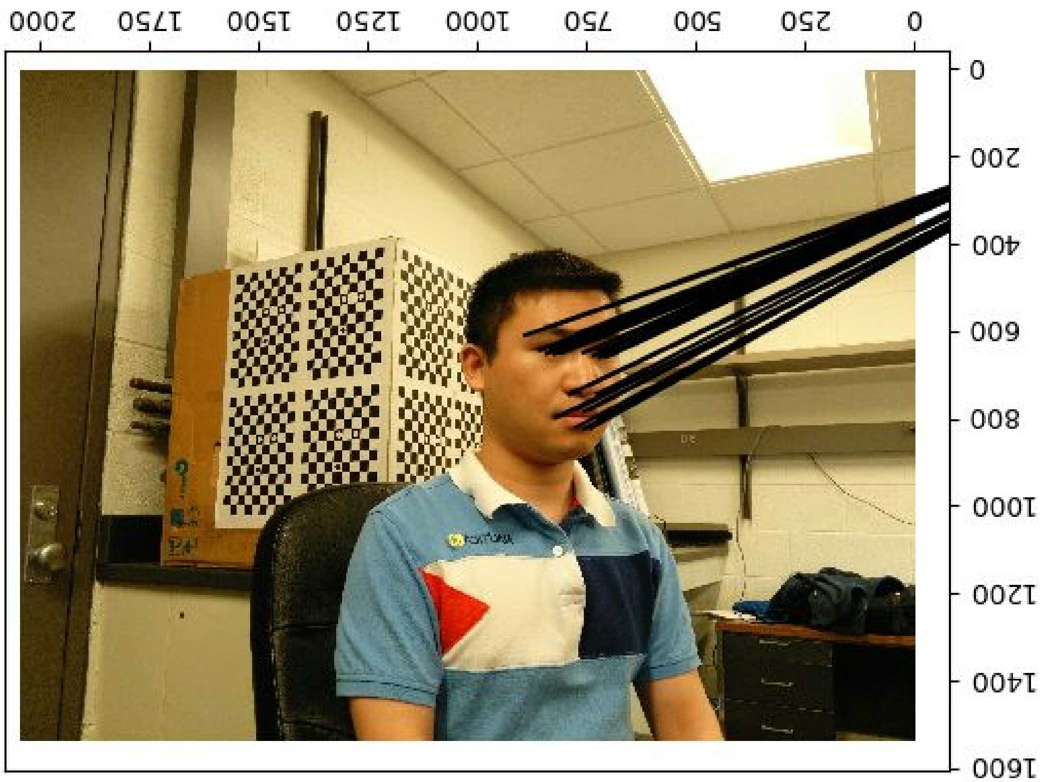

Figure 3: Left image epipolar lines

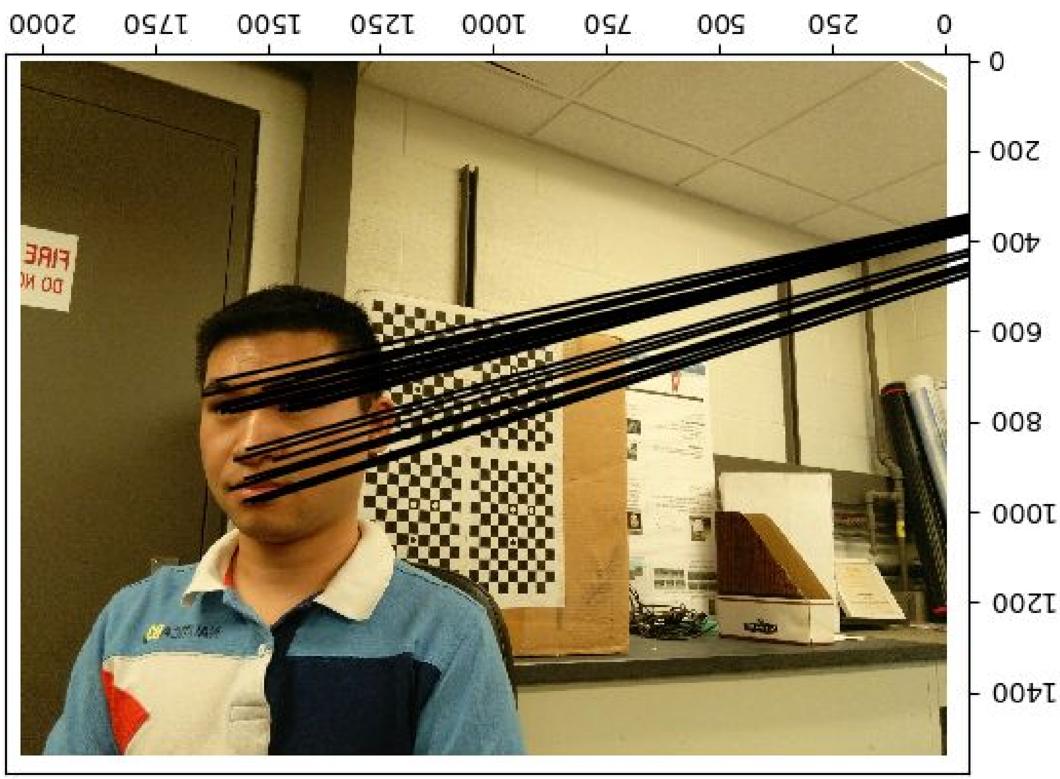

Figure 4: Right image epipolar lines

The epipolar lines are shown on the rectified images in figures (5) and (6):

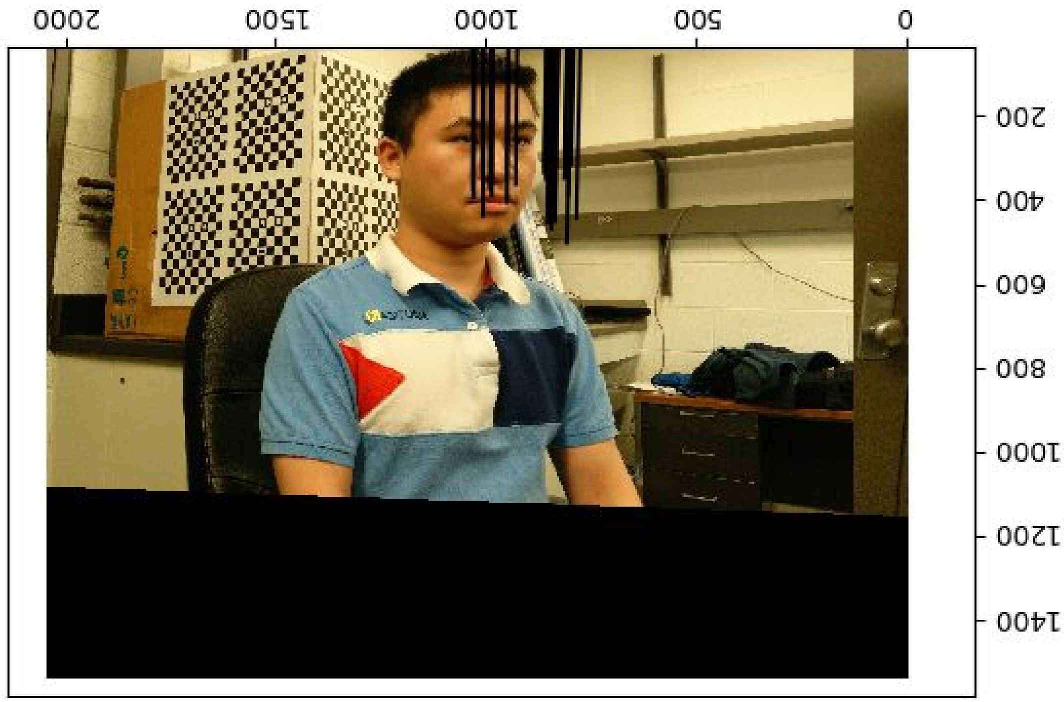

Figure 5: Rectified left image epipolar lines

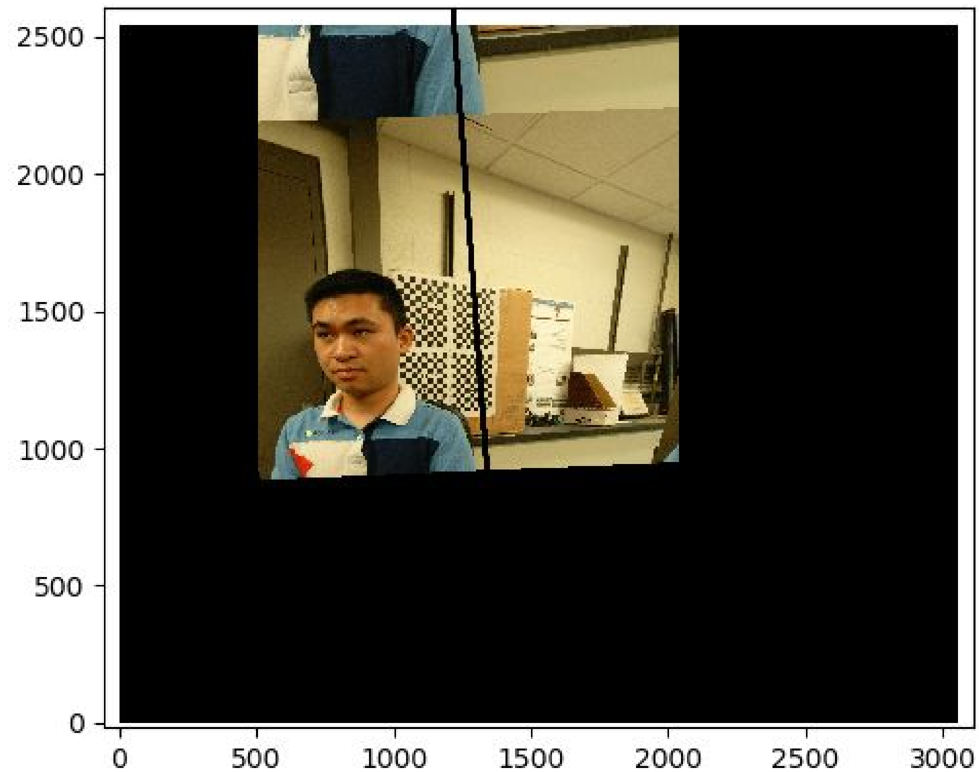

Figure 6: Rectified right image epipolar lines

Discussion

Due to a persistent issue with image rotation after rectification, the epipolar lines on the rectified images are on the wrong axes but exhibit the expected parallel behavior. RGB values were determined from nearest-neighbor estimations of coordinates that fell between pixels.

4. 3D reconstruction

Theory

Full 3D reconstruction can be performed if intrinsic camera parameters W_l, W_r and relative orientation *R, T* are known using the full perspective projection relationship between 3D points in the object frame and their corresponding 2D projected points in the image frame. A system of equations can be formed from the following:

$$\lambda_l \begin{pmatrix} c_l \\ r_l \\ 1 \end{pmatrix} = W_l \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (34)$$

$$\lambda_r \begin{pmatrix} c_r \\ r_r \\ 1 \end{pmatrix} = W_r \begin{bmatrix} R \ T \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (35)$$

Because each 2D point in the left camera frame *(c_l,r_l)* and its corresponding 2D point in the right camera frame *(c_r,r_r)* share the same 3D point in the object frame, (34) and (35) can be used to form a system of linear equations to solve for 3D point given a pair of 2D points using singular value decomposition by arranging the equations in the form AV = 0. Each 2D point is related to the rows of its projection matrix P and corresponding 3D point X by the following system of three equations, where P^N is the Nth row of the projection matrix:

$$\begin{aligned} c\left(P^3X\right) - \left(P^1X\right) &= 0 \\ r\left(P^3X\right) - \left(P^2X\right) &= 0 \\ c\left(P^2X\right) - r\left(P^1X\right) &= 0 \end{aligned} \quad (36)$$

The first two equations can be rearranged as:

$$\begin{bmatrix} cP^3 - P^1 \\ rP^3 - P^2 \end{bmatrix} X = 0 \quad (37)$$

From (37) an A matrix can be formed from two sets of corresponding 2D points from the left and right views:

$$A = \begin{bmatrix} c_l P_l^3 - P_l^1 \\ r_l P_l^3 - P_l^2 \\ c_r P_r^3 - P_r^1 \\ c_r P^3 - P_r^2 \end{bmatrix} \quad (38)$$

Performing singular value decomposition on (38) will yield the null eigenvector of A, which contains the values of X.

Results

The 3D points of the 2D facial feature points were computed for each pair of 2D points in the left and right view using equation (38) and plotted in figure (7). Their rough 3D approximation is included in figure (8).

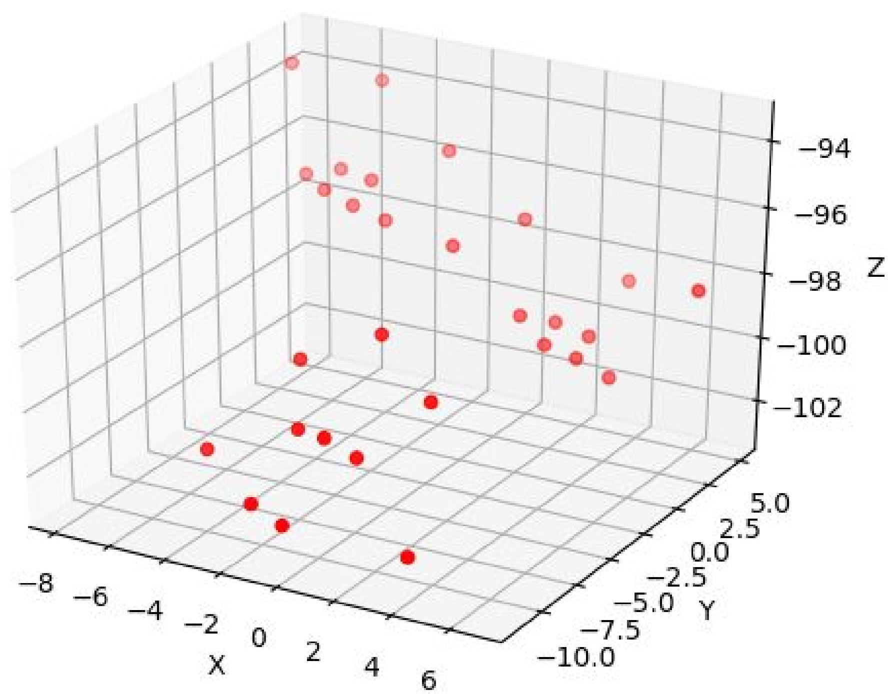

Figure 7: Plotted 3D points

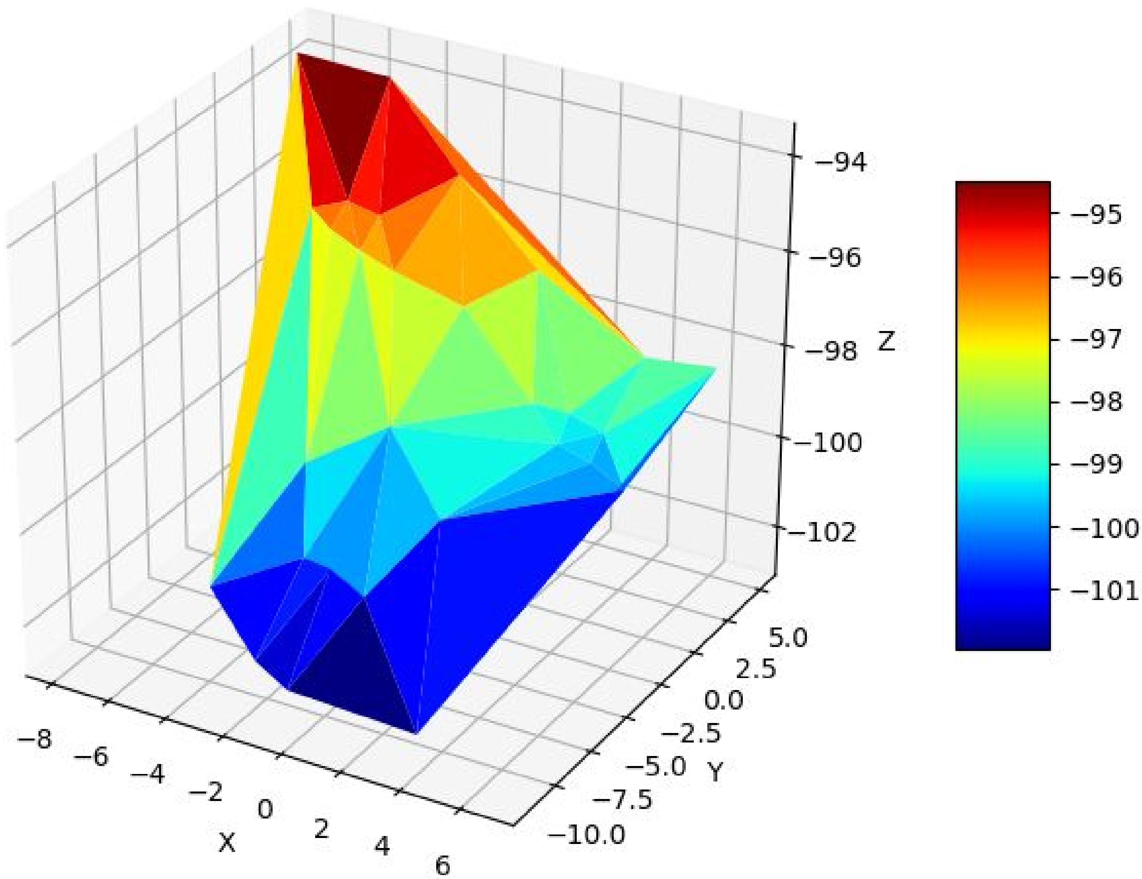

Figure 8: 3D map from figure 7

The width of each eye was calculated to be 21.5, and the width of the mouth was calculated to be 14.01.

5. Summary and Conclusion

Image rectification and 3D reconstruction were successfully performed, though there were persistent issues with the resulting images being rotated.