Camera Projection, Triangulation, & Epipolar Geometry

Date of current version November 25th 2024.

JASMINE KHALIL
Pennsylvania State University, State College PA (e-mail: jkk5987@psu.edu)

This work was carried out as part of a project for the CMPEN/EE 454 course at Penn State University instructed by Dr. Robert Collins

ABSTRACT Understanding the principles of 3D reconstruction from 2D images is pivotal for applications in computer vision, robotics, and augmented reality. These fields include techniques that map 3D points to 2D image planes, compute correspondences between views, and recover spatial information through triangulation. Motion capture (Mocap) systems are commonly used for precise motion tracking. They create an ideal framework for exploring these concepts with high accuracy and practical relevance. In this project, we address the problem of reconstructing and analyzing 3D structures from synchronized and calibrated camera views of a motion capture setup. We show how intrinsic and extrinsic camera parameters project 3D mocap points into 2D-pixel locations and recover 3D information through triangulation. Additionally, we compute the Fundamental matrix using both camera calibration parameters and the eight-point algorithm, evaluate the accuracy of the Fundamental matrices, and adapt them to cropped views. Our results demonstrate the importance of robust calibration and the sensitivity of the Fundamental matrix to the chosen estimation method. Comparisons reveal that while theoretical approaches align with the expected results, practical challenges such as noise and cropping require adjustments to maintain accuracy. This work highlights the relationship between theory and practice in camera geometry and emphasizes the significance of quantitative evaluation for reliable 3D reconstruction. By bridging theory with real-world implementation, our project contributes to the broader understanding of camera-based 3D reconstruction. These findings have implications for advancing technologies in computer vision and improving applications in various scientific and engineering disciplines.

INDEX TERMS 3D Reconstruction, Camera Projection, Computer Vision, Epipolar Geometry, Eight-Point Algorithm, Fundamental Matrix, Motion Capture Systems, Pinhole Camera Model, Triangulation.

I. INTRODUCTION

This project provides a practical exploration of critical concepts in camera projection, triangulation, and epipolar geometry, fundamental topics in computer vision and 3D reconstruction. We used data provided from a motion capture (Mocap) lab, where a person performs a movement while wearing a suit with infrared-reflecting markers. These markers are tracked by an array of precisely calibrated and synchronized infrared cameras to generate accurate 3D point measurements. We were given the following input data from the experiment:

1) Two images captured by two of the visible-light cameras representing views taken at precisely the same time, which had already been processed to remove non-linear radial lens distortion, making them 'corrected' images.
2) Two Matlab files, one for each of the camera views containing the intrinsic parameters of the camera: focal length, principal point, aspect ratio, and skew, which combine to form the intrinsic matrix Kmat and the extrinsic parameters: camera position and rotation matrix Rmat which combine to form the projection matrix Pmat.
3) A Matlab file containing 3D point locations in world coordinates of 39 markers on the performer's body.

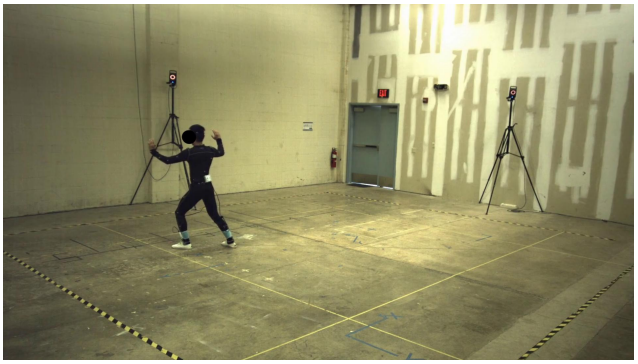


FIGURE 1: Provided image taken by the first camera view.

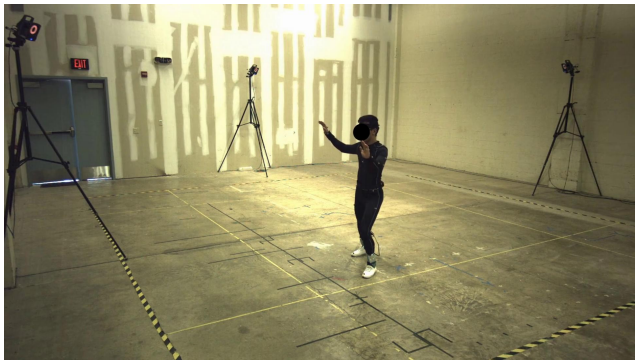


FIGURE 2: Provided image taken by the second camera view.

Using the provided intrinsic and extrinsic camera calibration parameters, which describe the cameras' relationship to the 3D world, we aim to understand the geometric relationship between the two camera views and how 3D structures are mapped into 2D images.

By working with real-world data and leveraging theoretical frameworks, this project bridges the gap between abstract mathematical concepts and their practical applications in fields such as robotics, augmented reality, and human motion analysis.

II. PINHOLE CAMERA MODEL

From the Pinhole Camera Model, we can extract the location of a point in camera coordinates using the following matrix equation[1].

$$\underbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}_{\text{Camera Point}} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -c_u \\ 0 & 1 & 0 & -c_v \\ 0 & 0 & 1 & -c_w \\ 0 & 0 & 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix}}_{\text{World Point}}.$$

In compact form:

$$P_C = R(P_W - C). \tag{1}$$

The above matrix equation can be simplified to:

$$\underbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}_{\text{Camera Point}} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix}}_{\text{World Point}},$$

and as a variable equation, this becomes:

$$P_C = RP_W + T. \tag{2}$$

Setting (1) equal to (2), we extract the following expression for the location of the camera, C:

$$C = -R^T T. \tag{3}$$

Using the provided rotation matrices and camera position vectors, we extracted the following camera locations.

$$\textbf{Camera 1 Location} = \begin{bmatrix} -3740.7 & -2143.2 & 5968.5. \end{bmatrix}$$

$$\textbf{Camera 2 Location} = \begin{bmatrix} 4210.3 & -1713.4 & 5673.2. \end{bmatrix}$$

We verified that the cameras' locations and their rotation matrices combine in the expected way to yield the appropriate entries in the projection matrices. Then, we compared the results between the given and constructed Pmats for both camera views, which matched as expected.

III. PROJECTING 3D MOCAP POINTS INTO 2D PIXEL LOCATIONS

To project the 39 given 3D Mocap points into 2D-pixel coordinates, we used the complete pinhole model camera model[1]:

$$\underbrace{\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}}_{\substack{\text{Pixel} \\ \text{Location}}} \sim \underbrace{\begin{bmatrix} \pm 1/s_x & 0 & o_x \\ 0 & \pm 1/s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}}_{\substack{\text{Film Plane} \\ \text{to Pixels}}} \underbrace{\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\substack{\text{Perspective} \\ \text{Projection}}}$$

$$\ldots \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -c_u \\ 0 & 1 & 0 & -c_v \\ 0 & 0 & 1 & -c_w \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\substack{\text{World} \\ \text{to Camera}}} \underbrace{\begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix}}_{\substack{\text{World} \\ \text{Point}}}.$$

By looping through each 3D point and substituting the X, Y, and Z coordinates with U, V, and W, respectively, as well as the provided camera parameters, we obtained 39 2D pixel coordinates. To ensure the accuracy of our results, we plotted them over the two provided image views:

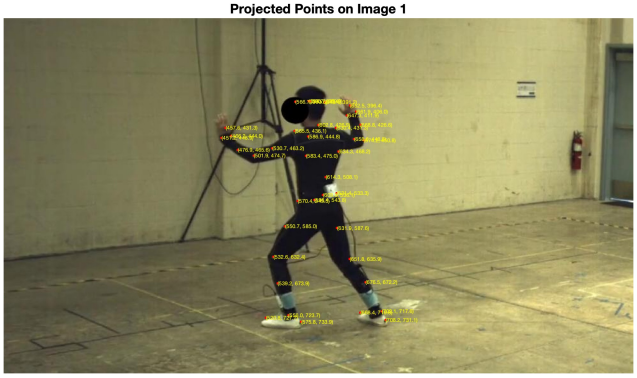


FIGURE 3: 2D locations of the Mocap points on the person in the first view.

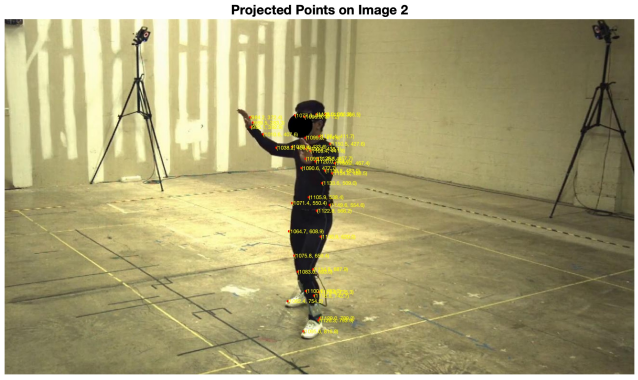


FIGURE 4: 2D locations of the Mocap points on the person in the second view.

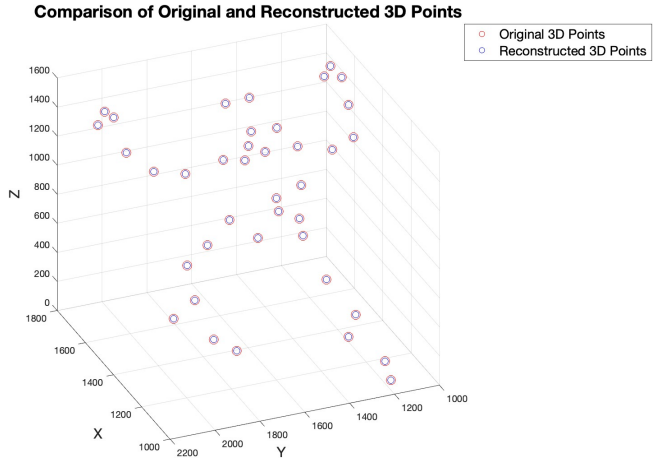


FIGURE 5: Visiual comparison between the actual 3D Mocap points and the reconstructed 3D points.

IV. TRIANGULATION

Triangulation determines the 3D position of a point in space using its 2D coordinate in multiple camera views[2]. Using the provided intrinsic and extrinsic camera parameters, we traced rays from the camera centers through to different 2D image points in both views and estimated 3D point coordinates as the intersection of these rays. We first did this using the calculated 2D points from the previous projection step and repeated it with user-selected 2D points in both views to make measurements about the scene[3].

A. RECOVERING 3D MOCAP POINTS FROM TWO VIEWS

Using our Mocap 2D points, we looped through each coordinate and calculated rays passing through each point in both views, found the direction of the cross-product of the two rays, calculated the closest points on each ray, and obtained the final, reconstructed 3D point as the midpoint between the closest points.

After applying this process to all 39 mocap points, we calculated a mean squared error between the actual and reconstructed 3D points and found that our set of reconstructed 3D points was very close to the original set:

$$MAE = 2.455e - 24.$$

We plotted a comparison between the actual 3D points and the reconstructed 3D points, showing that both sets of 3D points are very similar.

B. MAKING MEASUREMENTS ABOUT THE SCENE

By prompting the user to select points in both image views, we gathered another set of 2D coordinates. The following points were selected to make some measurements about the scene:

1) 3 corresponding floor points in both views.
2) 3 corresponding points on the painted wall in both views.
3) Point at the top of the doorway in both views.
4) Point at the bottom of the doorway in both views.
5) Point at the top of the person's head in both views.
6) Point at the person's shoes in both views.
7) Center of the camera near the striped wall in both views.

For each pair of corresponding 2D coordinates, we used triangulation to recover the 3D coordinate and plotted a 3D reconstruction of the scene:

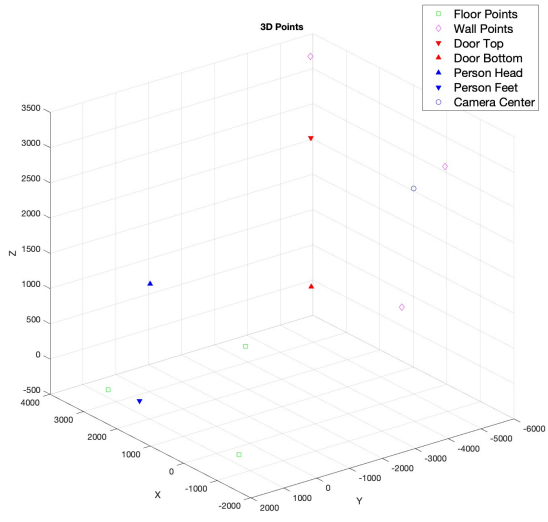


FIGURE 6: Reconstruction of 3D points of significant parts of the scene.

Using he 3D coordinates from the triangulation of each pair of user-selected points, we made the following measurements

about the scene:

- Height of the doorway ≈ 2146.36
- Height of person ≈ 1632.45
- 3D location of the center of the camera near the painted wall $\approx (-46.96, -4938.97, 2439.29)$

V. FUNDAMENTAL MATRIX

The fundamental matrix is a 3x3 matrix with rank 2 that plays a central role in epipolar geometry by encapsulating the geometric relationship between two views of the same scene.

From the Longuet-Higgins equation, we know that $P_r^T E P_l = 0$ Where E is the essential matrix that uses calibrated camera views while the fundamental matrix does not. Both matrices relate to each other through the intrinsic matrices from both views as[4]:

$$F = K_2^{-1} F K_1. \tag{4}$$

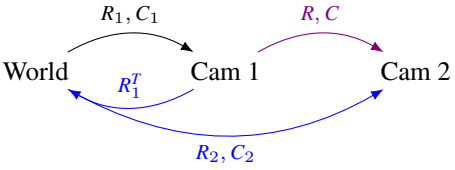
The epipolar constraint equation:

$$P_r^T F P_l = 0, \tag{5}$$

is also closely related to the Longuet-Higgins equations.

A. FUNDAMENTAL MATRIX FROM KNOWN CAMERA CALIBRATION PARAMETERS

Using the provided camera calibration parameters for both cameras,



The relative rotation between the two camera views is derived to be:

$$R = R_2 R_1^T.$$

The location of camera 2 with respect to the coordinate system of camera 1 is:

$$C = R_1(C_2 - C_1),$$

where C_1 and C_2 are the positions of camera 1 and camera 2, respectively, with respect to the world coordinate system. We formed the skew-symmetric matrix, T, using our results and obtained the following matrix:

$$T_{\text{skew}} = \begin{bmatrix} 0 & -C(3) & C(2) \\ C(3) & 0 & -C(1) \\ -C(2) & C(1) & 0 \end{bmatrix}.$$

We calculated the essential matrix using $E = R * T_{skew}$, from which we calculated the fundamental matrix using (4). For consistency, we chose the same set of points used in calculating F using the eight-point algorithm[4].

From the mapped points in both views, we calculated the epipolar line equations in both views using

$$l_2 = F * P_1, \tag{6}$$

and

$$l_1 = P_2^T * F, \tag{7}$$

and plotted all the lines to capture where the other camera is in each view (the epipoles).

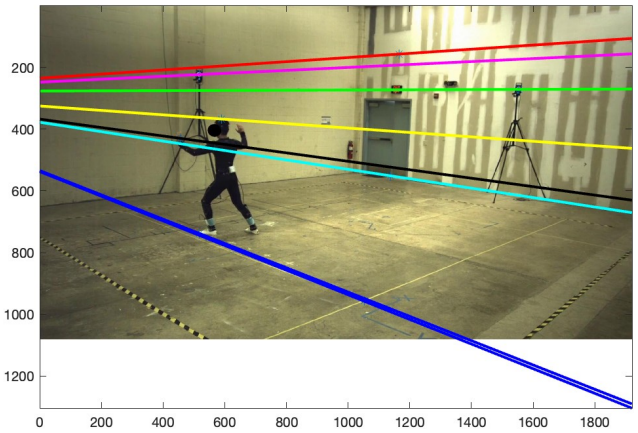


FIGURE 7: Epipolar lines plotted on the first view toward the second camera (the epipole) using F from known camera calibrations.

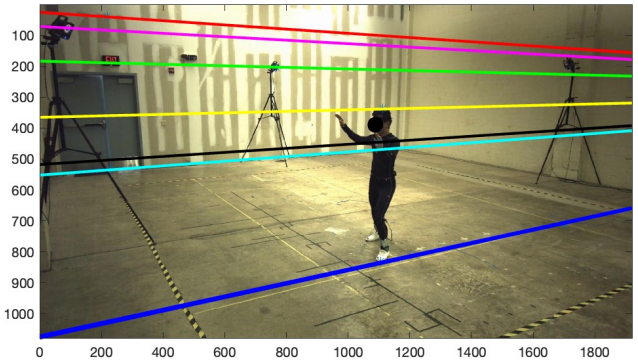


FIGURE 8: Epipolar lines plotted on the second view toward the first camera (the epipole) using F from known camera calibrations.

B. FUNDAMENTAL MATRIX FROM EIGHT-POINT ALGORITHM WITH HARTLEY PRECONDITIONING

The eight-point algorithm is used to gather the entries of the fundamental matrix from point correspondences in both image views. The general steps are:

1) Select m point correspondences that do not all lie on one plane or the same line, $m \geq 8$.
2) Construct the mX9 matrix, A.
3) Find the eigenvalues/eigenvectors of $A^T A$.
4) The entries of F are the components of the eigenvector corresponding to the smallest eigenvalue[5].

As a result of the coordinates of corresponding points possibly having a wide range of values and numerical instabilities, normalizing the points first such that they have a mean of 0 and standard deviation of 1, then denormalizing the fundamental matrix is a better approach. This process is Hartley Preconditioning by Richard Hartley[6].

Using the eight-point algorithm and implementing Hartley preconditioning, we calculated the fundamental matrix from user-selected points (roughly the same points used to calculate F from the known camera calibration parameters). We calculated the epipolar lines in both views using (6) and (7) and plotted them as below.

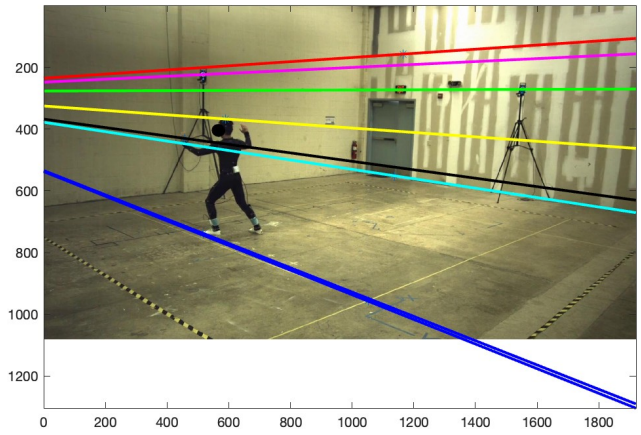


FIGURE 9: Epipolar lines plotted on the first view toward the second camera (the epipole) using F from the 8-point algorithm with preconditioning.

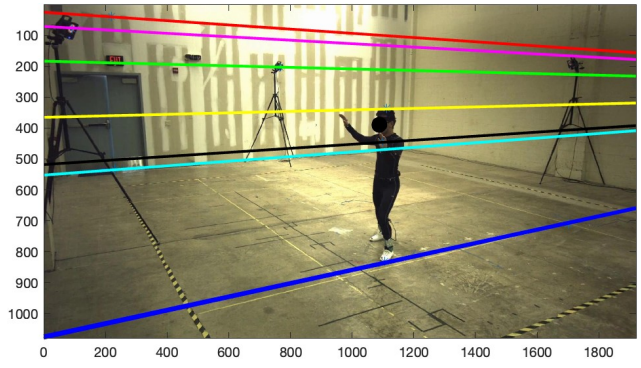


FIGURE 10: Epipolar lines plotted on the second view toward the first camera (the epipole) using F from the 8-point algorithm with preconditioning.

C. FUNDAMENTAL MATRIX FROM EIGHT-POINT ALGORITHM WITHOUT HARTLEY PRECONDITIONING

To test the significance of Hartley preconditioning in this case, we computed the fundamental matrix and plotted the epipolar lines using the same user-selected points but without the preconditioning. To do this, we built A without normalization and removed the step that does the 'unnormalization' of F[5].

The epipolar geometry still looks reasonable, but there are slightly larger gaps between the epipolar lines and the points selected compared with the epipolar geometry from the eight-point algorithm with Hartley preconditioning. Nevertheless, the plots below indicate that including Hartley preconditioning in this case does not significantly affect the results of the fundamental matrix.

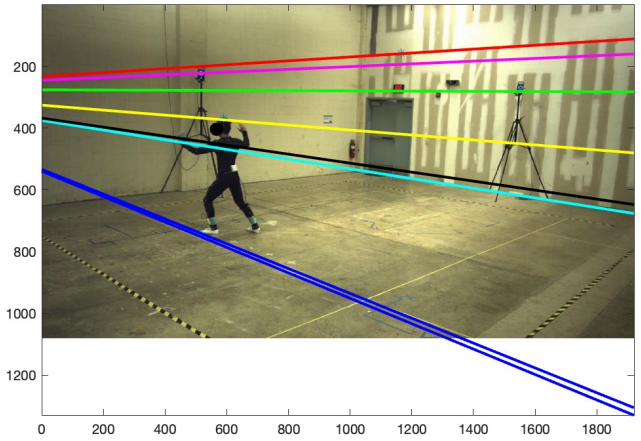


FIGURE 11: Epipolar lines plotted on the first view toward the second camera (the epipole) using F from the 8-point algorithm without preconditioning.

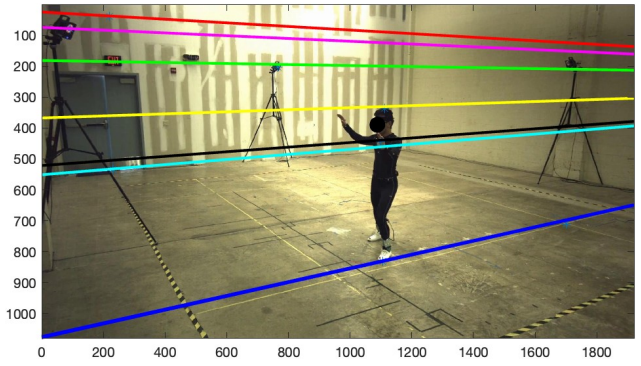


FIGURE 12: Epipolar lines plotted on the second view toward the first camera (the epipole) using F from the 8-point algorithm without preconditioning.

VI. QUANTITATIVE EVALUATION OF ESTIMATED F MATRICES

Using the set of 39 accurate 2D mocap points from section III, where one pair of those coordinates is (x1,y1) in image 1 and (x2,y2) in image 2, we computed the epipolar line in image 2 from (x1,y1) and computed the squared geometric distance of the point (x2, y2) from that line using the following equation:

$$SED = (ax + by + c)^2/(a^2 + b^2).$$

Repeating this process by mapping (x2,y2) in image 2 into an epipolar line in image 1 and measuring the square distance of (x1,y1) to that line, we accumulated all of these distances over the 39 3D point matches to get the SED error.

For each of the three fundamental matrices formed, we calculated the SED error. Our results are summarized below.

- SED for F from known camera calibration parameters ≈ 0.00
- SED for F from eight-point algorithm with Hartley preconditioning ≈ 9.45
- SED for F from eight-point algorithm without Hartley preconditioning ≈ 12.92

As expected, our SED error for F from the camera calibration parameters is very small compared with the error from either of the F matrices computed using the eight-point algorithm. Removing the preconditioning also seemed to affect the SED error reasonably.

VII. MODIFYING THE FUNDAMENTAL MATRIX FOR CROPPED VIEWS

We played around with the cropping parameters: starting pixel in the x direction (x_{start}), starting pixel in the y direction (y_{start}), width, and height for each of the images such that the images focus tightly around the person. We then used user-selected points to form a set of 2D points and adjusted the intrinsic matrices for both cameras using the following adjustments:

$$K_{\text{cropped}} = \begin{bmatrix} f_x & 0 & u_0 - x_{start} \\ 0 & f_y & v_0 - y_{start} \\ 0 & 0 & 1 \end{bmatrix}.$$

Cropping the views does not change the relative rotation matrices, R, or the skew-symmetric matrix, T_{skew} so E has not changed, but from (4), the fundamental matrix will change as a result of the updated intrinsic parameters for the cropped views. Using the updated fundamental matrix, we plotted the epipolar geometry and achieved the following plots as expected.

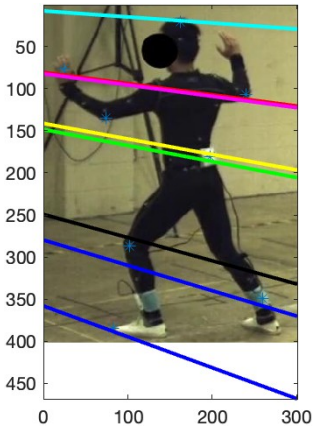


FIGURE 13: Epipolar lines plotted on the cropped first view toward the second camera (the epipole) using modified F.

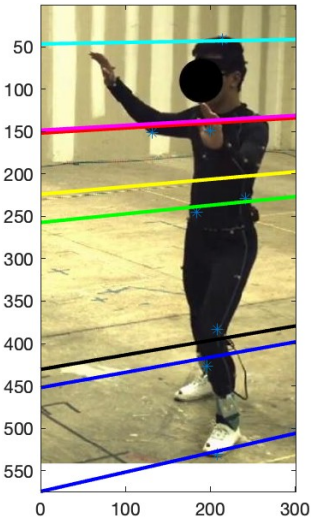


FIGURE 14: Epipolar lines plotted on the cropped second view toward the first camera (the epipole) using modified F.

CONCLUSION

This project successfully demonstrated key concepts in camera projection, triangulation, and epipolar geometry by integrating theoretical principles and practical implementation. By working with synchronized and calibrated camera views and known intrinsic and extrinsic parameters, we accurately projected 3D motion capture points into 2D-pixel coordinates. Furthermore, we used triangulation to reconstruct 3D scene geometry and computed the fundamental matrix, using both camera calibration parameters and the eight-point algorithm, providing insights into the relationship between two scene views.

Quantitative evaluation of the estimated fundamental matrices revealed the relative accuracy of these methods, highlighting the strengths and limitations of each approach. Adjusting the fundamental matrix to account for cropped camera views further demonstrated the adaptability of the methodology to real-world scenarios. These results align with theoretical expectations and reinforce the importance of precise calibration and synchronization in achieving accurate 3D reconstruction in computer vision applications.

Overall, this work contributes to a deeper understanding of multi-view geometry and its applications in MOCAP systems and 3D scene reconstruction. We emphasize the relationship between camera parameters, projection, and epipolar constraints, establishing a basis for further research into advanced computer vision techniques.

REFERENCES

[1] R. Collins, "Lecture 14: Pinhole Camera Model," 454: Fundamentals of Computer Vision, Penn State University, State College, PA.

[2] R. Collins, "Lecture 15: Triangulation," 454: Fundamentals of Computer Vision, Penn State University, State College, PA.

[3] R. Collins, "Lecture 16: Epipolar Geometry," 454: Fundamentals of Computer Vision, Penn State University, State College, PA.

[4] R. Collins, "Lecture 17: E/F Matrices," 454: Fundamentals of Computer Vision, Penn State University, State College, PA.

[5] R. Collins, "Lecture 18: The 8-Point Algorithm," 454: Fundamentals of Computer Vision, Penn State University, State College, PA.

[6] R. I. Hartley, "In defense of the eight-point algorithm," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 6, pp. 580-593, June 1997, doi: 10.1109/34.601246. keywords: Iterative algorithms;Layout;Cameras;Equations;Image reconstruction;Algorithm design and analysis;Stereo vision;Computer vision;Least squares methods,