University of Maryland

College Park

PROJECT 4

ENPM673 - PERCEPTION FOR AUTONOMOUS ROBOTS

Stereo Vision

Author:

Aaqib Barodawala (119348710)

Instructor and TAs:

Dr. Samer Charifa

Arunava Basu, Madhu NC,

Ritvik Oruganti

PROBLEM:
Implement the concept/pipeline of Stereo Vision System with the given three datasets.

SOLUTION:

1. Calibration:
   The solution begins with extracting the matching feature points from both the images in each dataset. This was done by utilizing SIFT extractor and BF matcher inbuilt functions from OpenCV.

   The proceeding step involves estimating the fundamental matrix using RANSAC. The fundamental matrix is defined by the equation:

   $$x'^{T}_{i} F x_i = 0,$$

   Where x_i and x'_i are the respective matching points from the two images of a dataset.

   We can rewrite the above equation in the form of $Af = 0$, where A is of the form:

   $$A = \begin{bmatrix} x_1x'_1 & x_1y'_1 & x_1 & y_1x'_1 & y_1y'_1 & y_1 & x'_1 & y'_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_nx'_n & x_ny'_n & x_n & y_nx'_n & y_ny'_n & y_n & x'_n & y'_n & 1 \end{bmatrix}$$

   After forming the A matrix, fundamental matrix F can be calculated by performing SVD of A. When using SVD, we get U, S, V where the last column of V is the solution of f. While using NumPy, it returns V^T after SVD. Hence, the last row of V^T will be the solution. The f obtained will be a 9×1 matrix which is then reshaped into a 3×3 matrix.

   Since F is supposed to be rank deficient i.e., its rank should be 2, the last row of the S matrix is made 0. To get the F matrix, the U, S, V are multiplied again (obtained from SVD), and the last element of that matrix is made equal to 1.

   To solve the above equation, 8 matching pairs are required. Hence, while implementing RANSAC, 8 random matching pairs are taken from the list of all the matching pairs extracted from the two images. For implementing RANSAC, the error is calculated using the equation, $x'^{T}_{i} F x_i$, where x_i and x'_i are the respective matching points from the two images of a dataset. The error is then compared to the threshold value set (for RANSAC) for each point. If the error is less than the threshold value, its added to the count of inliers obtained and this way, a close estimate of the fundamental matrix F can then be obtained.

   Proceeding ahead, the Essential Matrix E can be computed using the equation:

   $$E = K^T F K,$$

   Where K is the intrinsic matrix of the camera sensor.

The Essential matrix E can then be further decomposed to rotation and translation matrices by performing SVD of E. The two translation components are obtained by extracting the last column of U and negative U, where U, S, V is obtained by computing SVD of E. The two rotation components are obtained by using the following equations:

$$R_1 = UWV$$

$$R_2 = UW^TV$$

$$\text{Where } W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Since there are two pairs of rotation components and two pairs of translation components, there are total four solution sets of R and T. However, only one set provides a positive depth. To solve this issue, the pixel coordinates are triangulated for each R and T solution set using "cv.triangulatePoints()" function and then number of positive values are counted in the z-direction for each coordinate. The R and T pair with the maximum number of positive values for each coordinate is then selected.

2. Rectification:
   The epipolar lines can then be visualized by first computing the homography matrices for the two images using "cv.stereoRectifyUncalibrated()" inbuilt OpenCV function. Then by using "cv.computeCorrespondEpilines()" inbuilt OpenCV function, the epipolar lines are drawn on the two copies of the images. Using the homography matrices obtained, the images can be warped and rectified using the "cv.warpPerspective()" function. The epipolar lines obtained will be horizontal with respect to the matching points obtained.

3. Correspondence:
   The disparity is the relative horizontal distance between two matching points from the two images and is calculated along the x-axis. Using a 3×3 window and Sum of Squared Differences (SSD) method, the disparity can be calculated between the two images. For each pixel, the 3×3 window is formed for each image and the pixel with minimum SSD is chosen as correspondence. The disparity is then calculated.

   Using "cv.normalize()" and "cv.applyColorMap()" inbuilt functions, the disparity color and grayscale image map can then be visualized.

4. Compute Depth Image Map:
   With the disparity obtained above for each pixel, the depth can be calculated similarly for each pixel by using the following equation:

$$Depth = \frac{Focal\ Length \times Baseline}{Disparity}$$

Using “cv.normalize()” and “cv.applyColorMap()” inbuilt functions, the depth color and grayscale image map can then be visualized.

PROBLEMS FACED:

1. Major problem faced was understanding the underlying functions used. After research and practising the said functions, the process was understood, and the problem was resolved.
2. Making the RANSAC robust also posed the biggest problem. Using “np.random.seed()” and “np.random.choice()” within RANSAC helped solve the issue.

RESULTS:

1. DATASET 1 – Artroom:

Fundamental Matrix:

[[1.01869756e-07 -2.17134676e-05 2.06923778e-03]
 [2.43606036e-05 -3.37878202e-06 -1.79286338e-01]
 [-3.25455623e-03 1.77131076e-01 1.00000000e+00]]

Essential Matrix:

[[0.30620565 -65.2675218   -16.67227633]
 [73.22442731 -10.15612674 -280.5487633]
 [17.38407162 274.09947205   -0.89874932]]

Rotation Matrix:

[[0.99967534 -0.0065272   0.02462951]
 [0.00699876 0.99979291 -0.01910891]
 [-0.02449968 0.01927508 0.999514]]

Translation Matrix:

[0.97162309 -0.05847524 0.22919252]

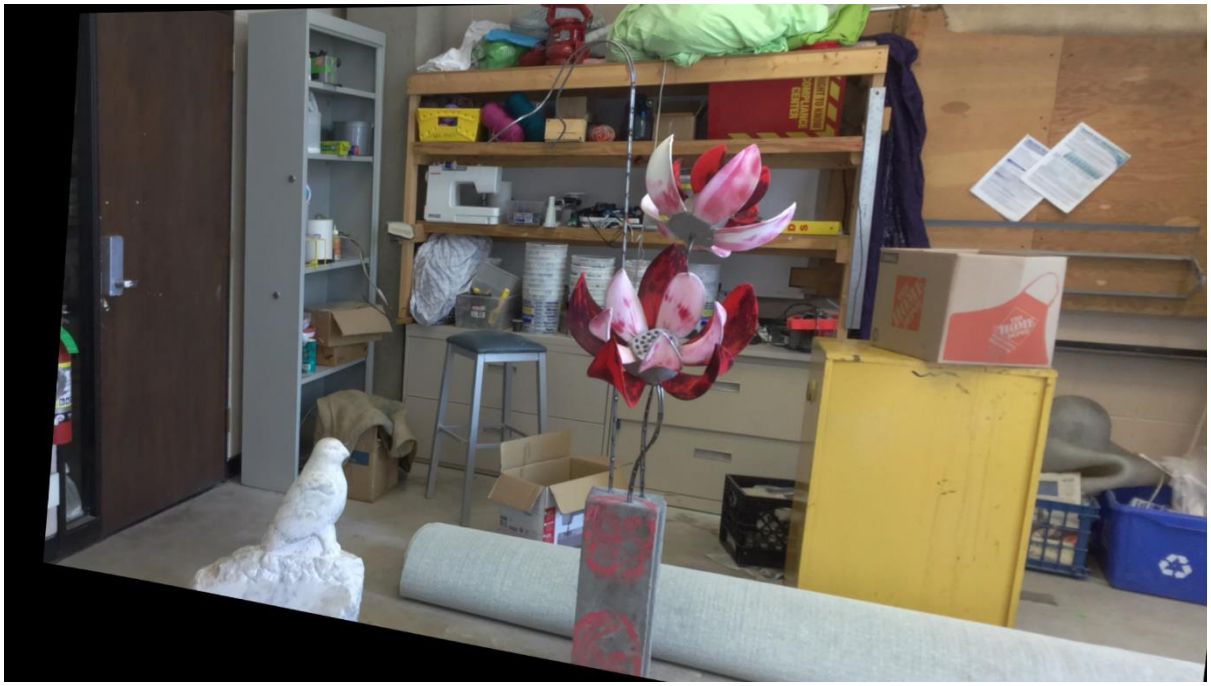
Homography matrix for image 1:

[[-1.48640197e-01 8.95884018e-03 -1.92150422e+01]
 [3.54816132e-03 -1.77665333e-01 -3.09672534e+00]
 [2.75223890e-05 -5.32036761e-06 -2.02360817e-01]]

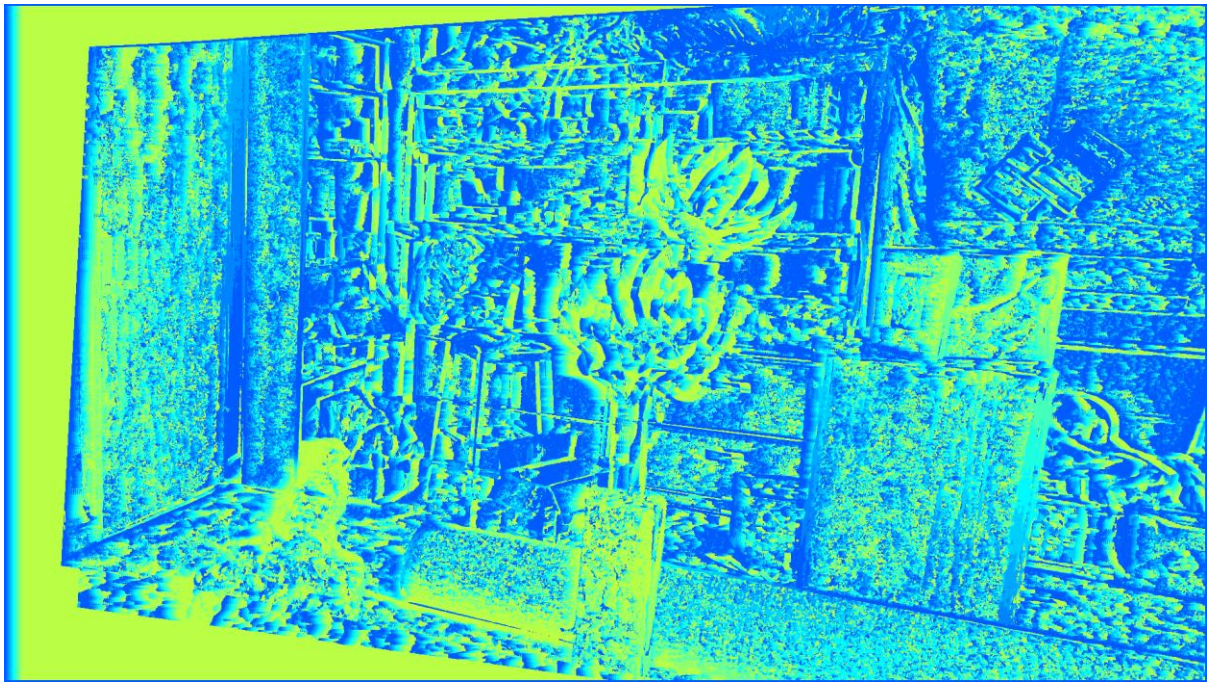
Homography matrix for image 2:

[[8.64961325e-01 -5.30440894e-02 1.58280936e+02]
 [-1.36941064e-02 1.00271843e+00 1.16783881e+01]
 [-1.38712042e-04 8.50656991e-06 1.12857001e+00]]

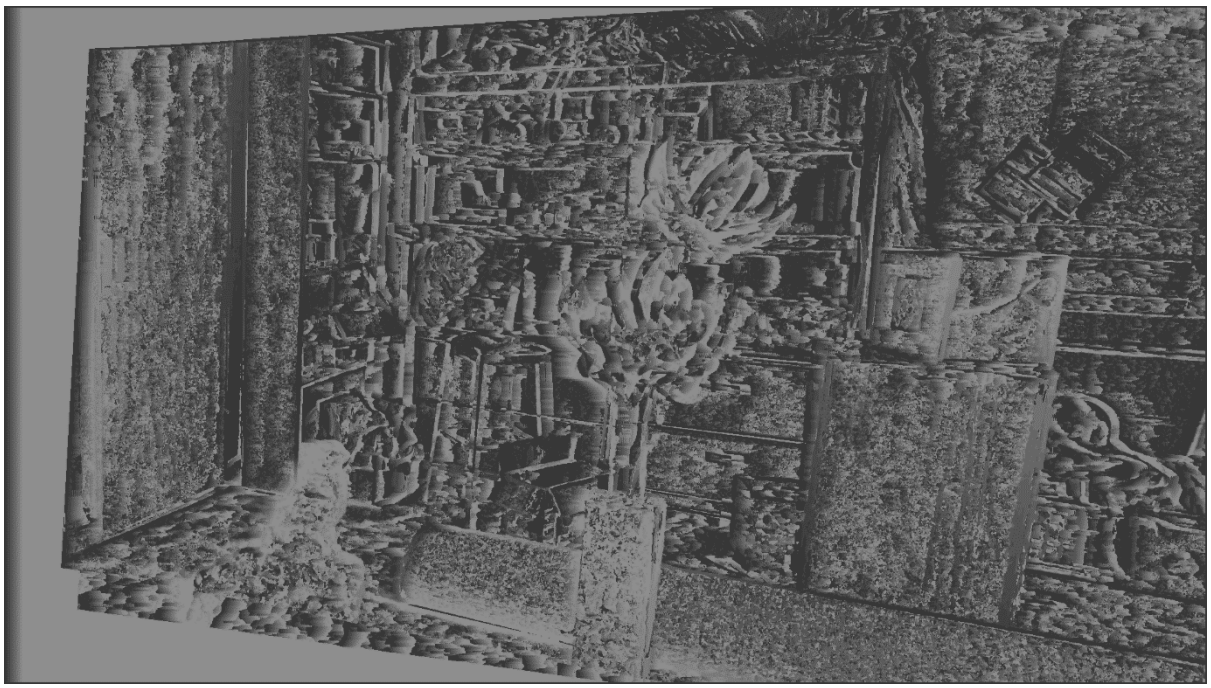
Rectified Images:



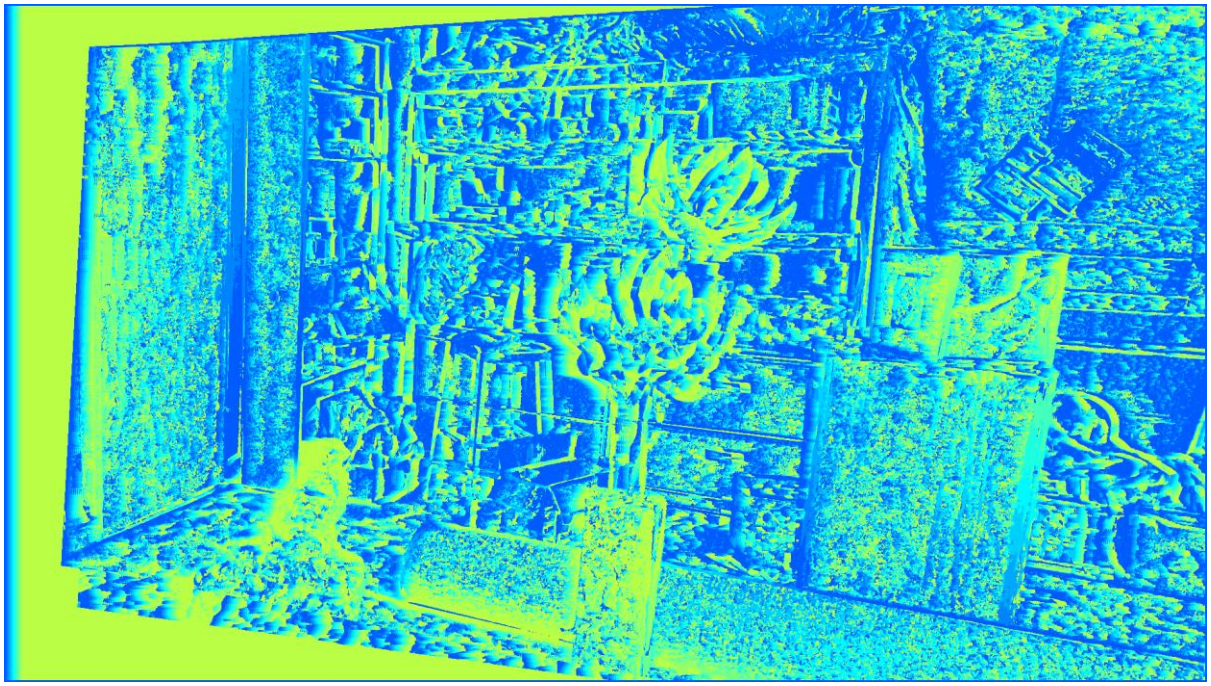
Disparity Color Map:



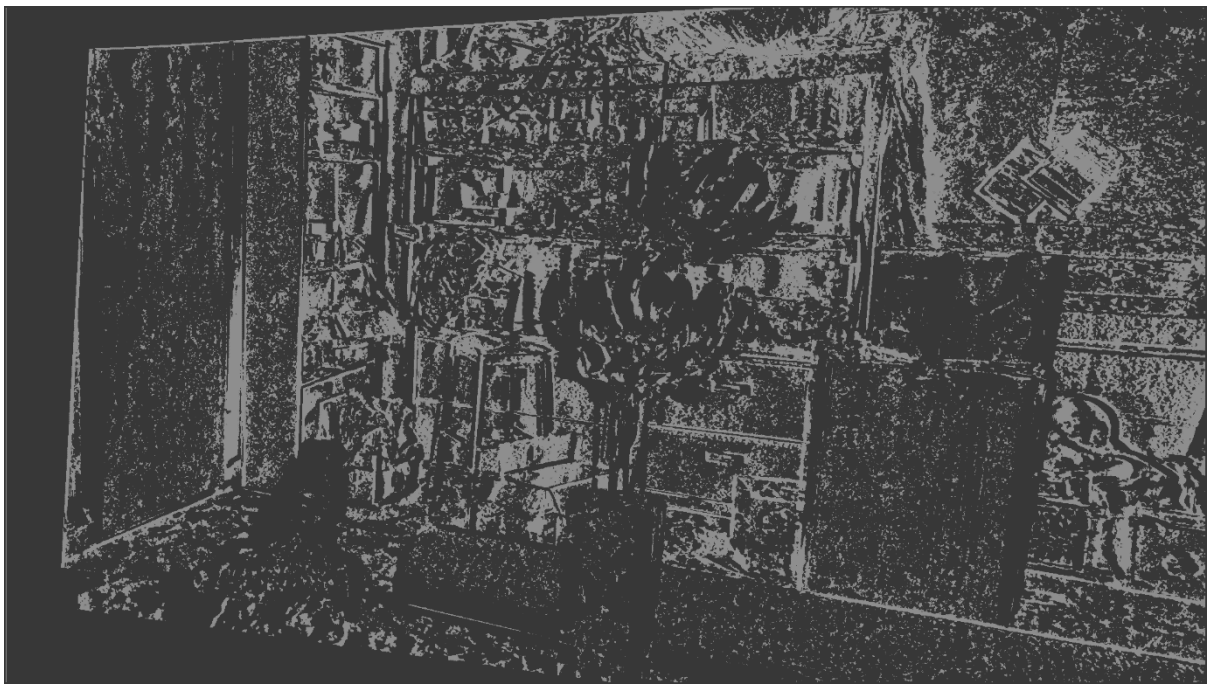
Disparity Grayscale:



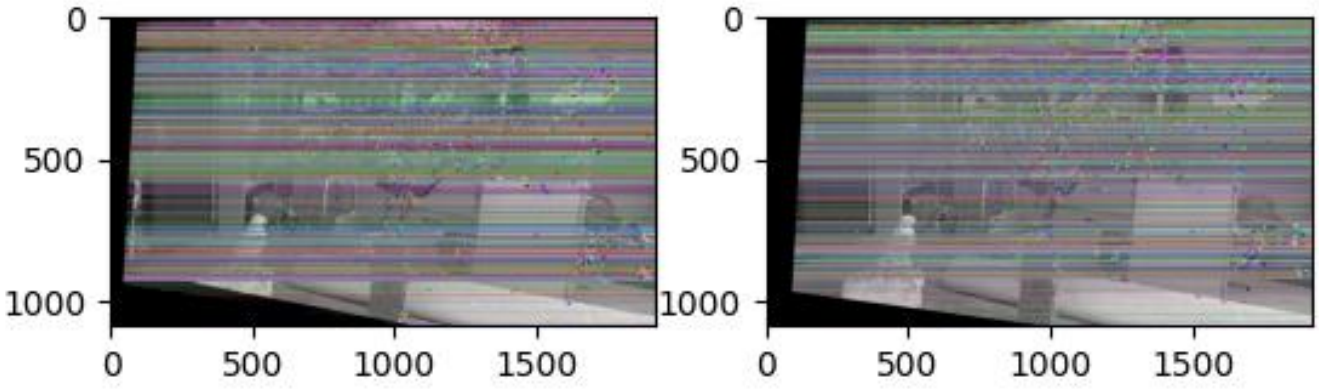
Depth Color Map:



Depth Grayscale:



Epipolar lines:



2. DATASET – Chess:

Fundamental Matrix:

[[1.00020552e-07 -1.69326471e-05 1.04246976e-02]
 [1.61169213e-05 -2.77404399e-07 1.02752851e-02]
 [-9.92599128e-03 -1.18043881e-02 1.00000000e+00]]

Essential Matrix:

[[0.30920081 -52.34512362   2.0177465]
 [49.82341115 -0.8575604   41.29253586]
 [-1.64209445 -45.70947209   0.17836487]]

Rotation Matrix:

[[9.99651568e-01 -6.58557286e-04 -2.63876724e-02]
 [7.77905982e-04 9.99989514e-01 4.51288587e-03]
 [2.63844237e-02 -4.53184056e-03 9.99641598e-01]]

Translation Matrix:

[-0.65774427 0.0288892   0.75268712]

Homography matrix for image 1:

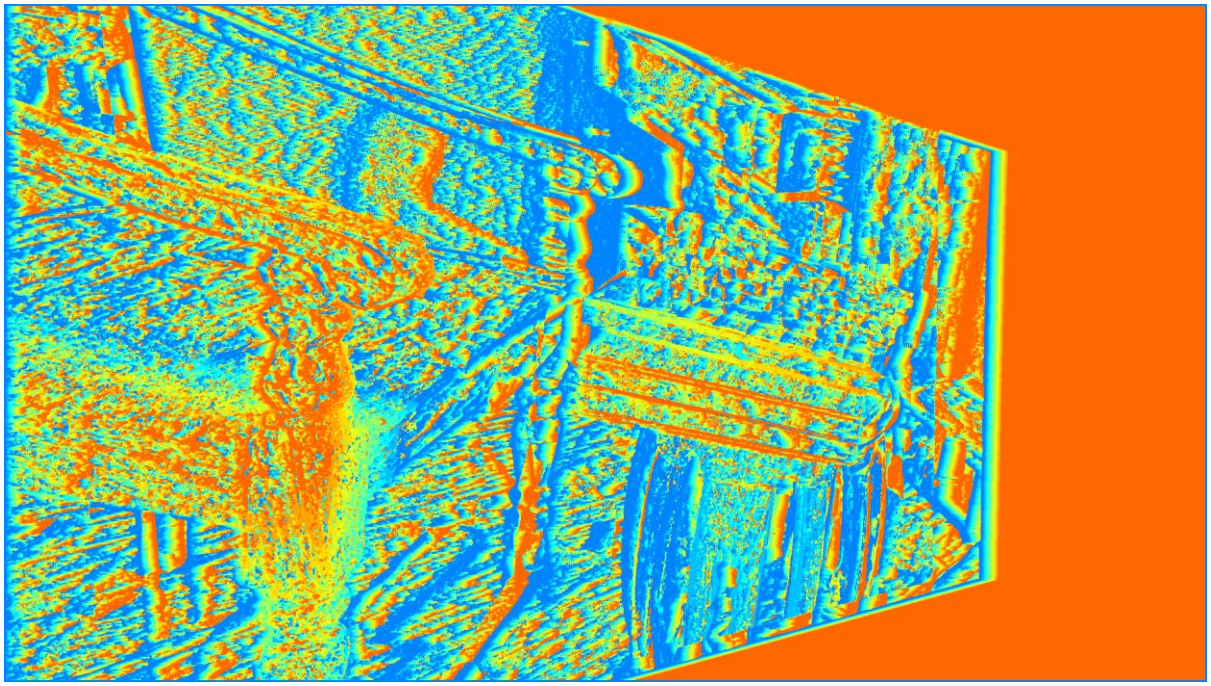
[[-2.38933499e-02 1.14776219e-03 1.07468345e+01]
 [-5.53576355e-03 -1.55662094e-02 6.05478951e+00]
 [-9.07537528e-06 6.15019905e-07 -6.06674573e-03]]

Homography matrix for image 2:

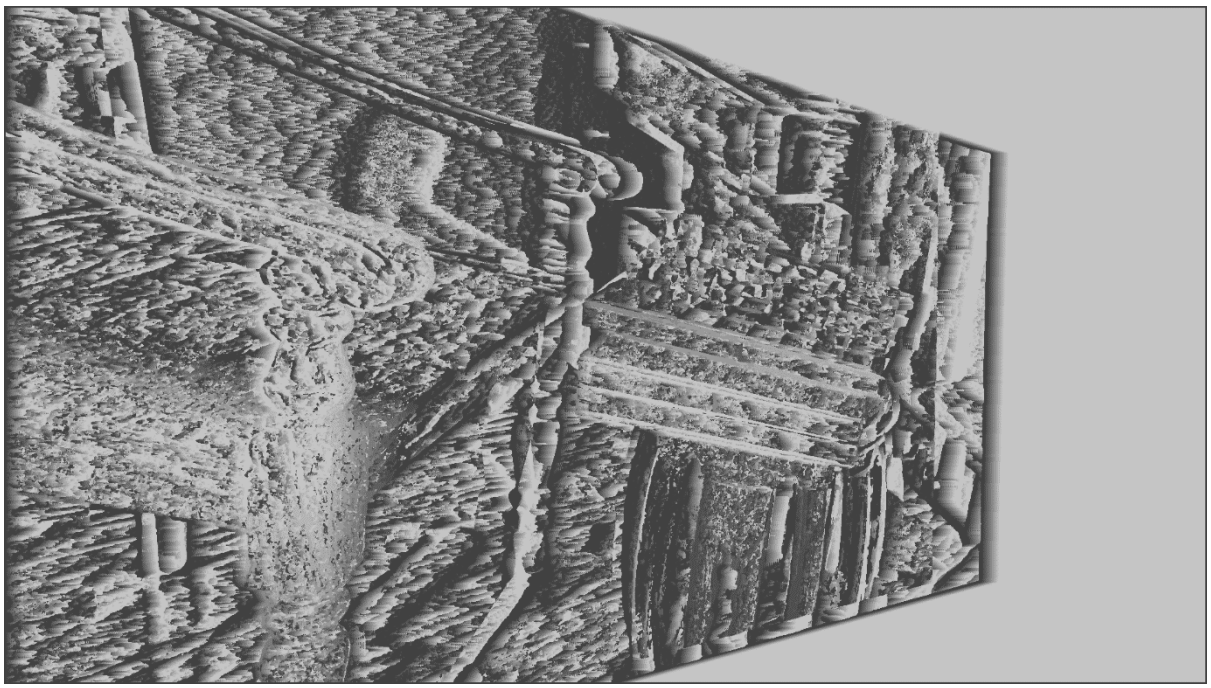
[[1.57329344e+00 -7.57379343e-02 -5.09463215e+02]
 [3.71212261e-01 9.83287986e-01 -3.47339283e+02]
 [5.98385567e-04 -2.88061246e-05 4.41105163e-01]]

Rectified Images:

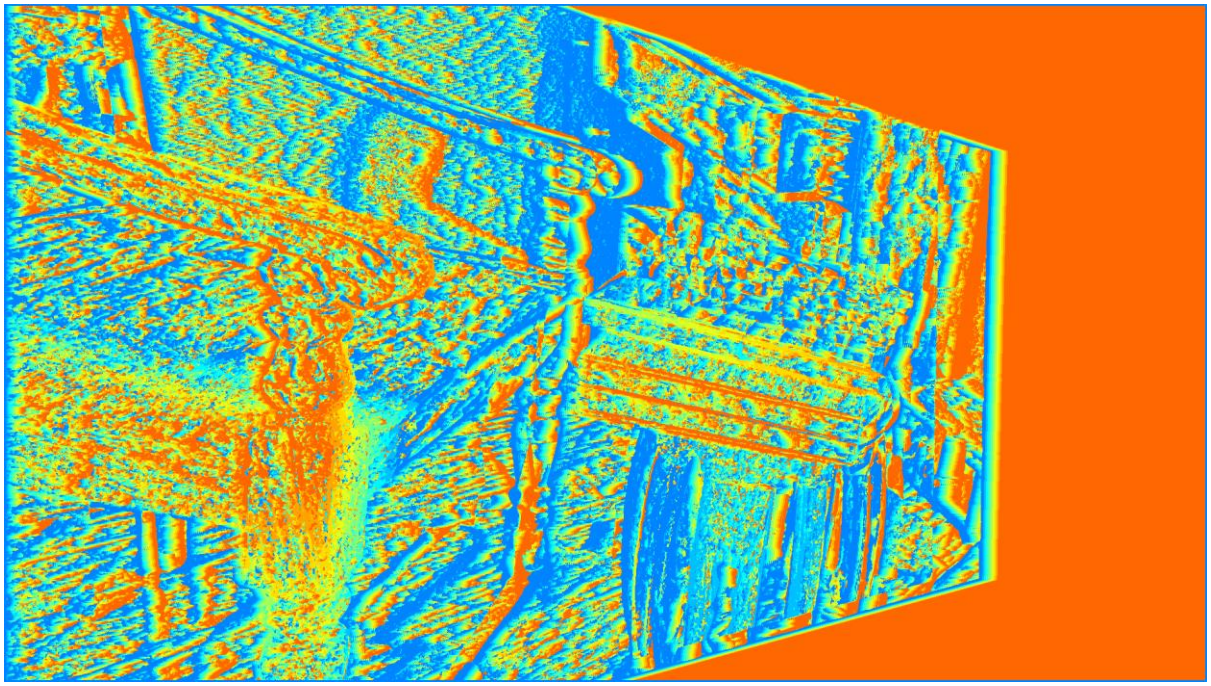
Disparity Color Map:



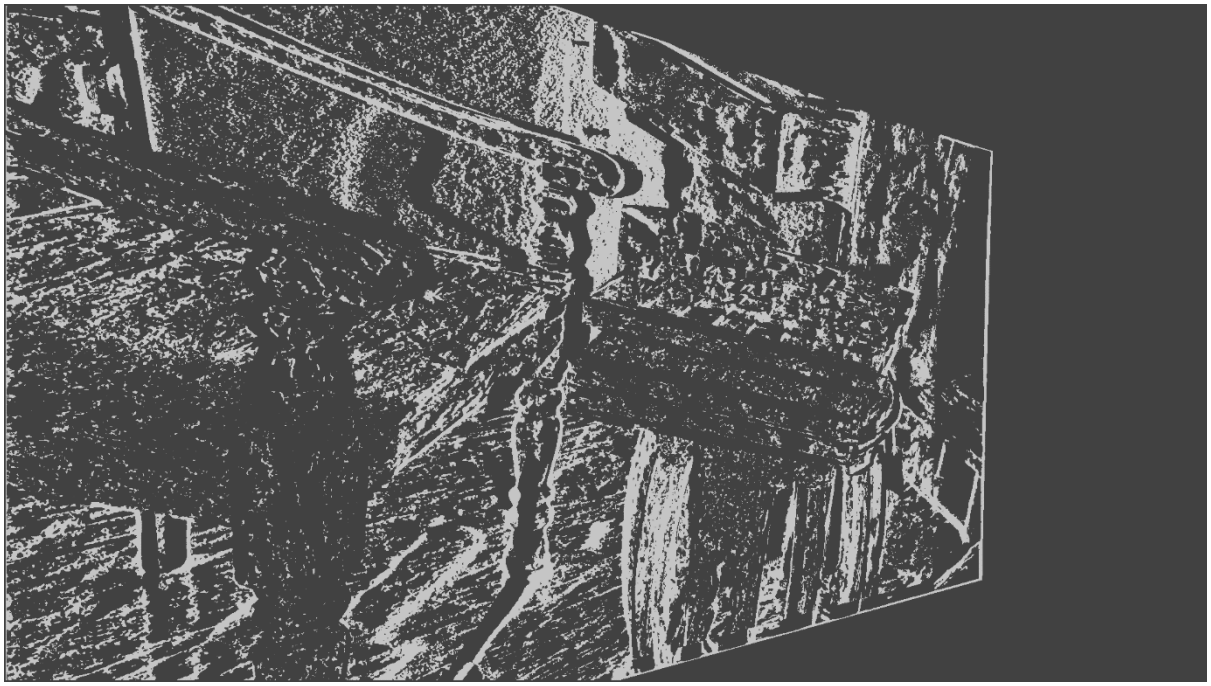
Disparity Grayscale:



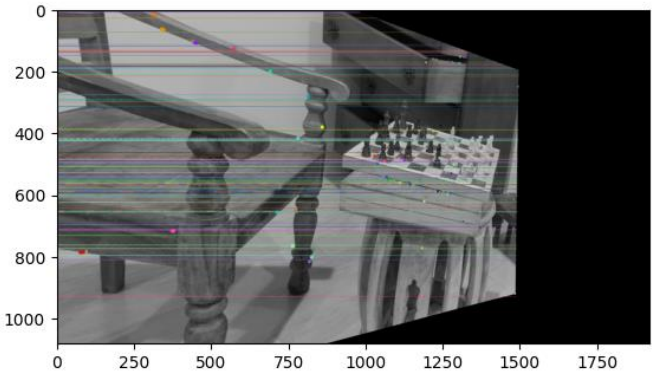
Depth Color Map:

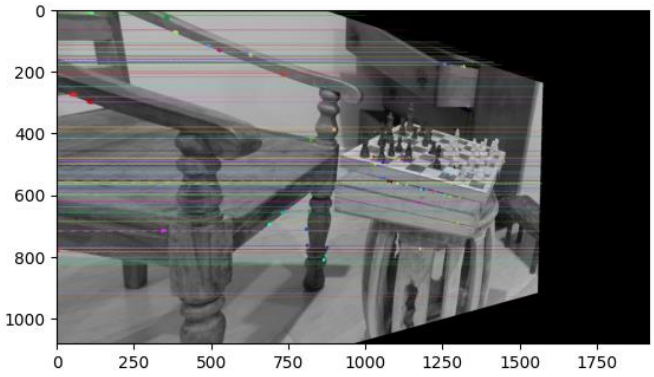


Depth Grayscale:



Epipolar lines:





3. DATASET – Ladder:

Fundamental Matrix:

[[-3.39220768e-08 -7.55687966e-06 5.11179483e-03]
 [8.59344039e-06 3.11505635e-07 6.87045327e-02]
 [-6.10702855e-03 -7.01573315e-02 1.00000000e+00]]

Essential Matrix:

[[-1.02014231e-01 -2.27258866e+01 -3.71003832e+00]
 [2.58431470e+01 9.36794295e-01 1.24631993e+02]
 [3.66706227e+00 -1.25516831e+02 -1.10427295e-01]]

Rotation Matrix:

[[9.99670688e-01 2.93663556e-04 -2.56598757e-02]
 [-3.85696157e-04 9.99993511e-01 -3.58175907e-03]
 [2.56586573e-02 3.59047647e-03 9.99664315e-01]]

Translation Matrix:

[0.98362192 0.0291228 -0.17787573]

Homography matrix for image 1:

[[-7.50954993e-02 -1.31810099e-03 7.32023617e+00]
 [-5.79812853e-03 -7.05018925e-02 3.72311391e+00]
 [-8.13525212e-06 -5.03609932e-07 -6.48926215e-02]]

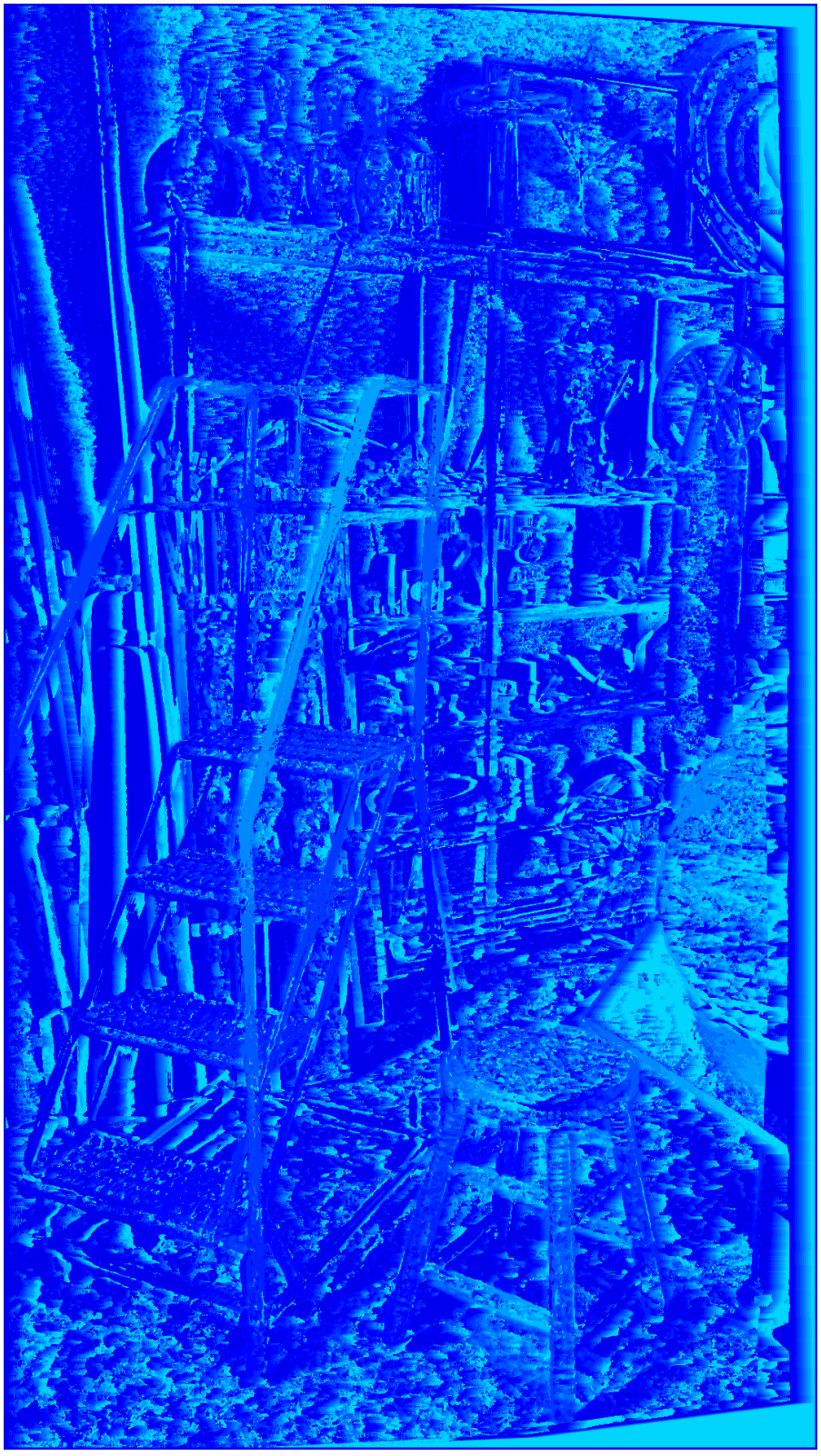
Homography matrix for image 2:

[[1.05465145e+00 3.07774948e-02 -5.90581803e+01]
 [6.87444546e-02 1.00243187e+00 -3.94565964e+01]
 [1.01994436e-04 2.97646507e-06 9.42065598e-01]]

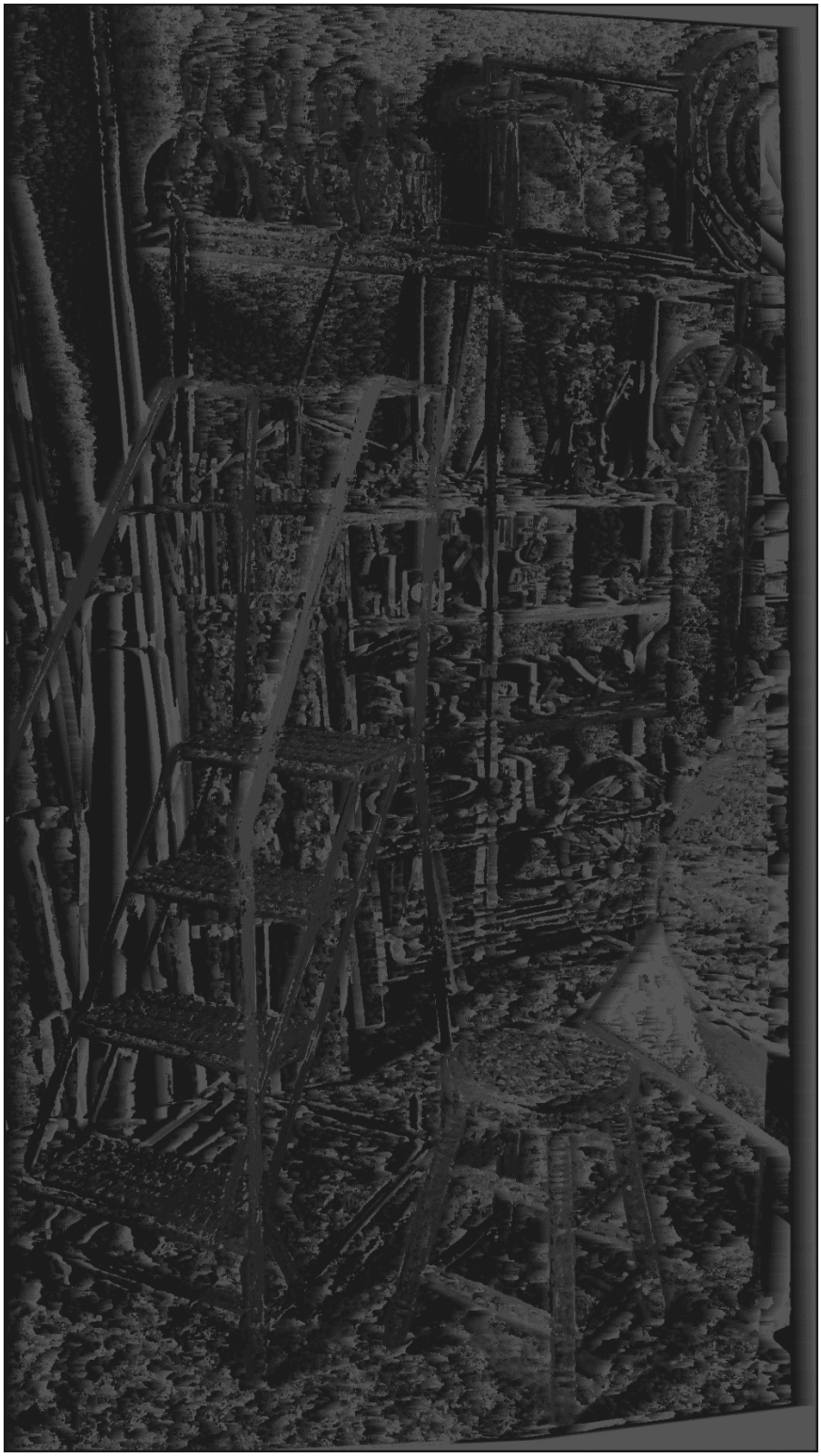
Rectified Images:


ATCC

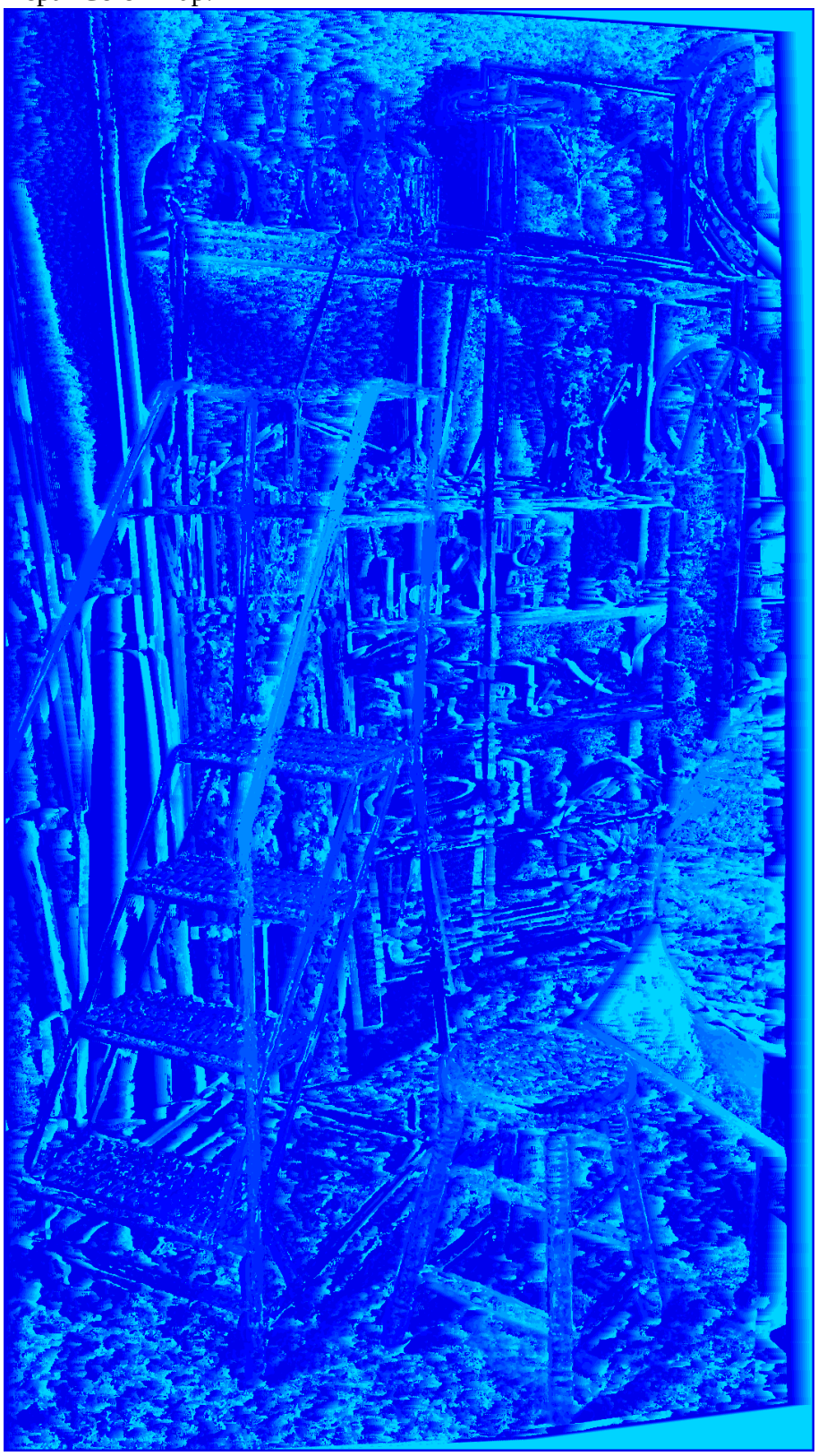
Disparity Color Map:



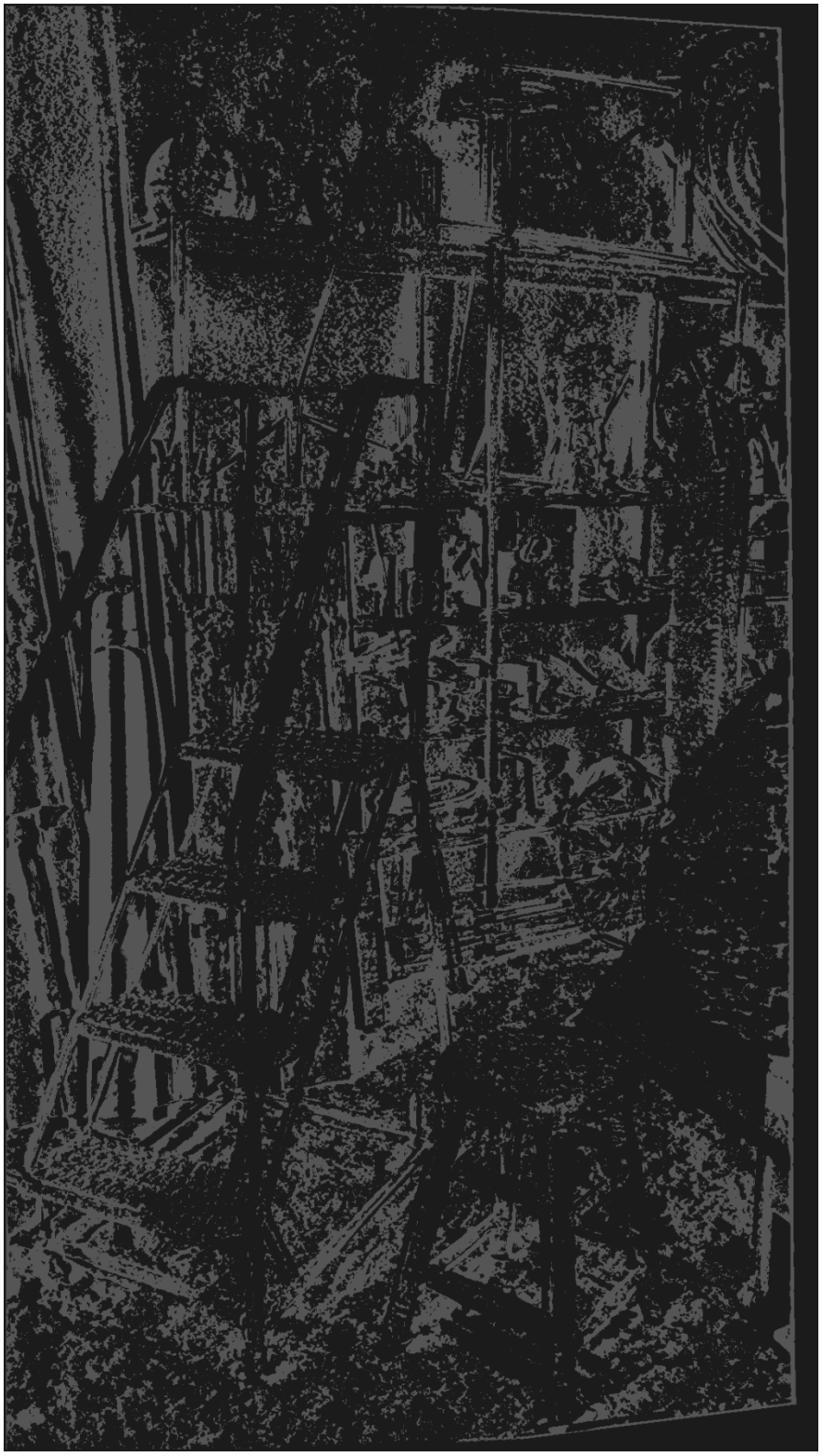
Disparity Grayscale:



Depth Color Map:



Depth Grayscale:



Epipolar lines: