Project 3

Neel Bhalla, Eric Grimaldi

December 4, 2021

1 Abstract

In this project, we explore techniques for stereo vision for a camera rig with unknown parameters. In such a setting, we work with two images from the same time frame that share nearly all of their features. However they are slightly warped relative to each other due to the difference in camera position. We begin with an application of our Harris Corner Detector and Normalized Cross Correlation algorithms from Project 2 to this new setting. We then extend these methods by using the correspondences we find to estimate the *Fundamental Matrix*, and applying a new version of our RANSAC algorithm from Project 2 to eliminate outliers. We close the project with an exploration of *dense disparity maps* of our experiments, as a milestone before true 3D reconstruction.

2 Approach

Our approach to estimating the Fundamental Matrix for cameras with unknown parameters begins by finding point correspondences between the images from the two cameras. This method is largely the same as our method from Project 2. First, the two images are pre-processed to filter any small noise (using a Gaussian filter), and then corners are extracted using our Harris Corner Detector (HCD). Due to the large number of outputs from HCD, we use non-max suppression to remove any features which do not exist at a local maximum. Using the sparse set of corners from non-max suppression, we find point correspondences across images by finding the maximum Normalized Cross Correlation (NCC) values between the neighborhoods of corners. We provide a more in depth explanation of these methods in Project 2, and as such will refrain from repetition.

Given a set of at least 8 point-to-point correspondences, we estimate the Fundamental Matrix using the least squares solution. Due to correspondence outliers skewing the solution significantly, this is done after running RANSAC and keeping inliers for some progressively strict threshold.

Finally, we compute a Dense Disparity Map (DDM) between the two images. In order to do this, we use NCC again to find correspondences for every point in the image. We use the Fundamental Matrix to greatly restrict our search space; when searching for an arbitrary correspondence, instead of searching the entire image we search only along the epipolar line given by the Fundamental Matrix. We visualize our disparity information as 3 images: a grayscale image for the horizontal component, a grayscale image for the vertical component, and a color image for both components where the direction of the disparity vector is encoded as hue and the magnitude of the disparity vector is encoded as saturation.

2.1 Fundamental Matrix

In our approach we estimate the Fundamental Matrix within RANSAC from 8 point to point correspondence in order to remove outliers. Once outliers have been removed, we recalculate the Fundamental Matrix using all inliers. For both of these calculations, we use the same method.

By definition we know

$$p_{r_i}^T F p_{l_i} = 0$$

For convenience, let us change subscripts

$$p'^T F p = 0$$

This can be rearranged as

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0$$

Which becomes

$$xx' f_{11} + xy' f_{21} + x f_{31} + yx' f_{12} + yy' f_{22} + y f_{32} + x' f_{13} + y' f_{23} + f_{33} = 0$$

For a number of point correspondences n, we can construct

$$\begin{bmatrix} x_1x'_1 & x_1y'_1 & x_1 & y_1x'_1 & y_1y'_1 & y_1 & x'_1 & y'_1 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_nx'_n & x_ny'_n & x_n & y_nx'_n & y_ny'_n & y_n & x'_n & y'_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{12} \\ f_{22} \\ f_{32} \\ f_{13} \\ f_{23} \\ f_{33} \end{bmatrix} = 0$$

For convenience, we will refer to this as

$$Af = 0$$

Since f has 9 unknowns, but the norm of f is arbitrary, we can look for a solution with $||f|| = 1$. This constraint means we need at least 8 correspondences to solve this problem, however we can use an arbitrarily large number of points to improve the accuracy of our calculation.

This means that we are essentially solving

$$\min_f ||Af||^2 = f^T A^T Af \text{ such that } ||f||^2 = f^T f = 1$$

To do this, we will optimize the Lagragian cost

$$min_f[\mathcal{L}(f) = f^T A^T Af - \lambda(f^T f - 1)]$$

Taking the derivative with respect to λ, we get our original norm constraint

$$f^T f - 1 = 0$$

$$f^T f = 1$$

Taking the derivative with respect to x, we find an eigenvector problem

$$A^T Af - \lambda f = 0$$

$$A^T Af = \lambda f$$

As such, f is the eigenvector of $A^T A$ with eigenvalue λ. This simplifies our Lagrangian cost optimization to

$$min_f[\mathcal{L}(f) = \lambda$$

So we must choose f to be the eigenvector of $A^T A$ with the smallest eigenvalue. We accomplish this by computing the Singular Value Decomposition of A, and choosing f to be the last colum of V, which corresponds to the eigenvector of $A^T A$ with the smallest eigenvalue, finally giving us the Fundamental Matrix F.

2.2 Dense Disparity Map

Calculation of the DDM is actually relatively simple. Essentially, for each pixel, we calculate the epipolar line for that pixel in the other image using the Fundamental Matrix.

$$l_{r_i} = Fp_{l_i}$$

We then search for a correspondence between the original pixel and some pixel on epipolar line in the other image. We use our NCC again to evaluate these possible correspondences, though this time we use a patch the true image, instead of the sparse feature set used in previous sections of our approach.

With these correspondences, disparity itself is a simple calculation; simply find the difference in x and y coordinates for the two corresponding points.

$$d_i = p_{l_i} - p_{r_i}$$

As a final task, we construct 3 images as a visualization of the DDM: a grayscale image for the horizontal component of d_i $\forall i$, a grayscale image for the vertical component of d_i $\forall i$, and a color image for both components of d_i $\forall i$ where the direction of the disparity vector is encoded as hue and the magnitude of the disparity vector is encoded as saturation.

As a note: computation of the DDM can be time consuming. For a stereo pair of square images of size nxn, the process of finding correspondences for the entire image requires an NCC calculation for roughly n^3 pixels. As such, we take two steps to further save time. Firstly, we simply downsize our images by a factor of 5 in both directions. Secondly, in our experiments, we know which image is "left" and which image is "right"; further we can see by inspection that both stereo pairs were recorded by stereo rigs with small baselines. With this knowledge, we can restrict the search along the epipolar line to a reasonable subset of the line, instead of the entire line.

3 Experiments

In this project, we conducted our simple experiment on two data sets. We estimate the Fundamental Matrix and calculated the DDM for two pairs of stereo images. The "Cone" stereo pair in Figure 1 is generally feature rich, with lots of texture. The "Prison" stereo pair in Figure 2 is also reasonably feature rich, but also includes patches of sky which are largely homogeneous.

(a) left im (b) right im

Figure 1: "Cone" stereo pair of images

(a) left im

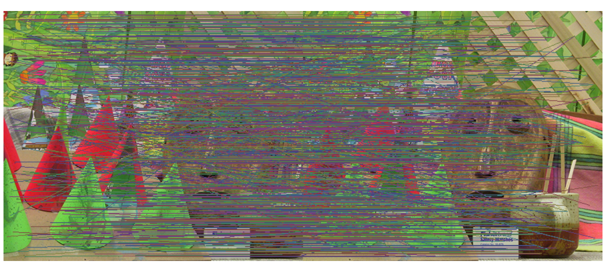
(b) right im

Figure 2: Prison stereo pair of images

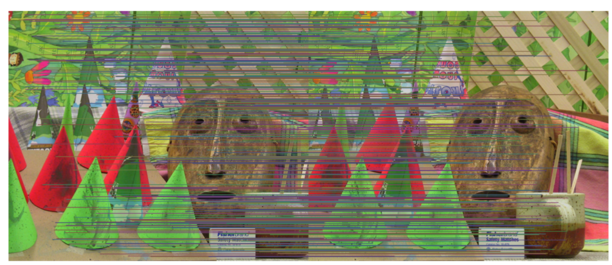
4 Results

4.1 Cones Pair

The cones dataset features two cones and a mask, with a very intricate background. The images are feature rich, allowing NCC to be effective at finding actual matches between the two images. As a result, the HCD and NNC algorithms found a lot of features (with 11 pixel windows), many of which are not aligned with the true fundamental matrix. The removal of these outliers can be seen clearly in Figure 3.



(a) left im



(b) right im

Figure 3: Cone pair pre and post RANSAC correspondences.

After applying RANSAC to eliminate the outliers and estimate the Fundamental Matrix, it was revealed that the two images differ by a translation in camera position. One can confirm this in Figure 3 which shows that all the correspondences are connected across images with horizontal lines, or Figure 7 which shows that there is 0 vertical disparity across the entire image.

It follows that the fundamental matrix will yield horizontal epipolar lines, meaning that the two cameras different in orientation only by a small translation. It will thus typically suffice to scan lines that nearly mimic rows of the images in order to find correspondences. The detections found by our program confirm this, as one can see in Figures 4-6.

Figure 4: Example p_l and p_r found by our program. The corresponding points are highlighted with blue rectangles. The correspondence is found by sampling NCC from templates centered on the epipolar line (highlighted in yellow) found with the fundamental matrix.

Figure 5: Another example

The results of our DDM calculations can be seein in Figures 7-9. Due to Fundamental Matrix yielding only nearly horizontal lines, it effectively is impossible for there to be any vertical disparity. Figure 7 confirms this. Given this knowledge of the geometry of the stereo rig, the horizontal disparity is then a function of the depth of the image. This is evident in Figure 8.

Figure 6: Final example

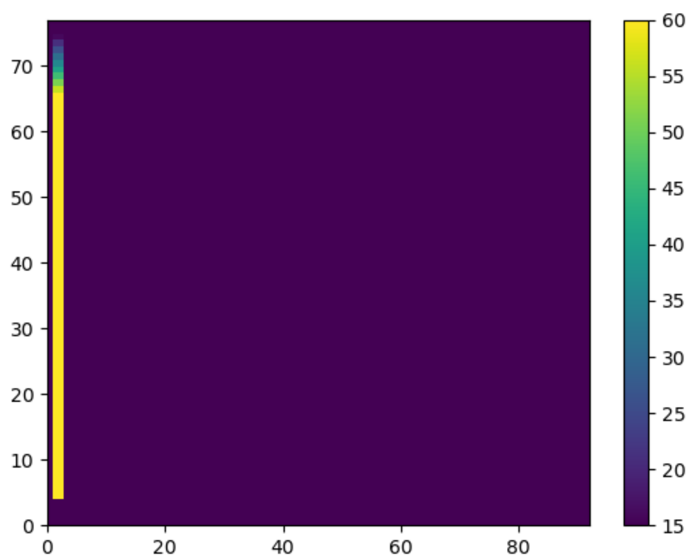


Figure 7: Vertical disparities in image

It is possible to see the different items in the horizontal disparity image, as they were placed at different depths in the scene. The wooden lattice background pattern can also been seen faintly in the top right of the image. There are outliers in the disparity as the normal cross correlation mixed up some prominent recurring features in the images, primarily the tops of cones and the sticks in the mug, where parallax motion may have obliterated some features. In general, our algorithm effectively found matches when given an epipolar constraint.

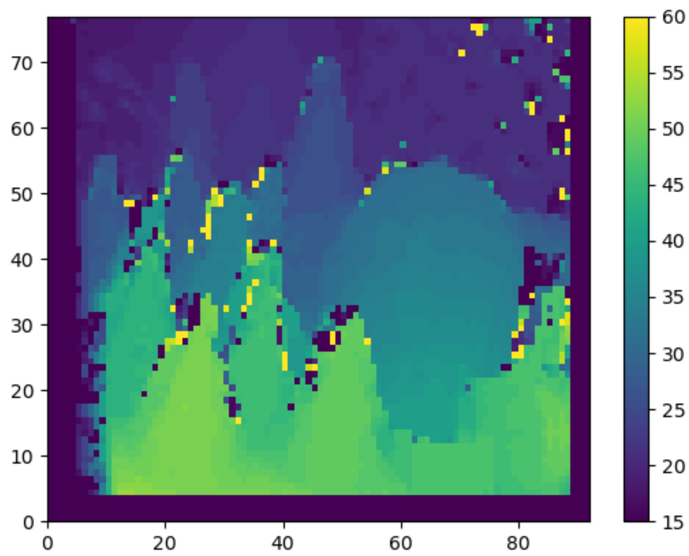


Figure 8: Horizontal disparities in image

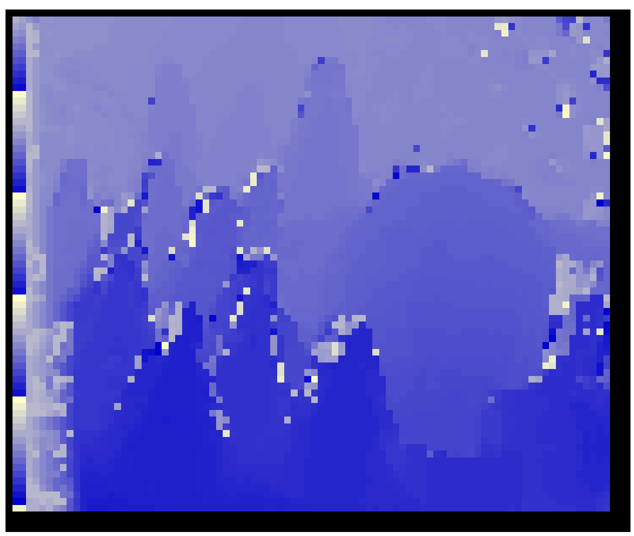
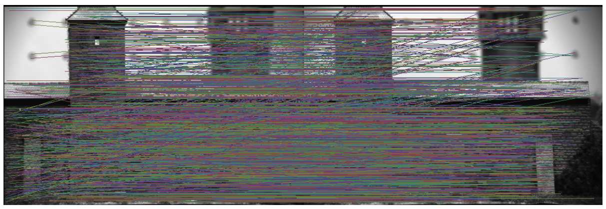


Figure 9: Vectorized HSV version of the disparity map. Observe that there are no vertical disparities, so every pixel is a shade of 0°, i.e. blue. Darker blue refers to closer elements whereas further away is lighter.

4.2 Prison Pair

The prison dataset features a sharply textured building in the foreground, a tower out of focus in the background, and a highly homogeneous sky. There are a lot of features in the texture of the foreground building, but the sky is a mostly homogeneous region; the sky will provide some issue with feature and point correspondence due to its homogeneity.

Due to the plethora of features on the image, the feature extraction and correlation portion of our program was successful. Running RANSAC to estimate the Fundamental Matrix revealed that the cameras were again separated by a horizontal translation, as demonstrated again in Figures 10 and 14.



(a) left im

(b) right im

Figure 10: Prison Pair pre and post RANSAC correspondences.

Looking at the correspondences, it seems that there will again be horizontal epipole lines. When we leverage the estimated Fundamental Matrix to find point correspondences, we again confirm this in Figures 11-13.

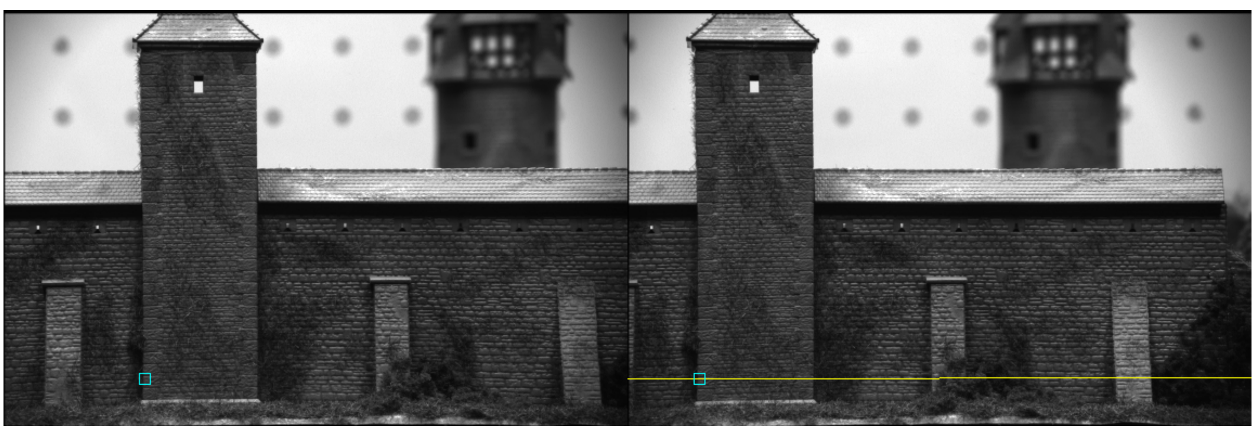


Figure 11: Example p_l and p_r found by our program. The corresponding points are highlighted with blue rectangles. The correspondence is found by sampling NCC from templates centered on the epipolar line (highlighted in yellow) found with the fundamental matrix.

While the towers and dots are easy to find, the sky region is far too homogeneous to correctly estimate correspondences for many of these pixels. The epipolar line for points in the sky cut through the whole sky, meaning that our approach for calculating DDM has a high chance of finding false correspondences in the sky due to high NCC results. Figure 13 depicts this problem (which persists even for a large window size of 11x11).

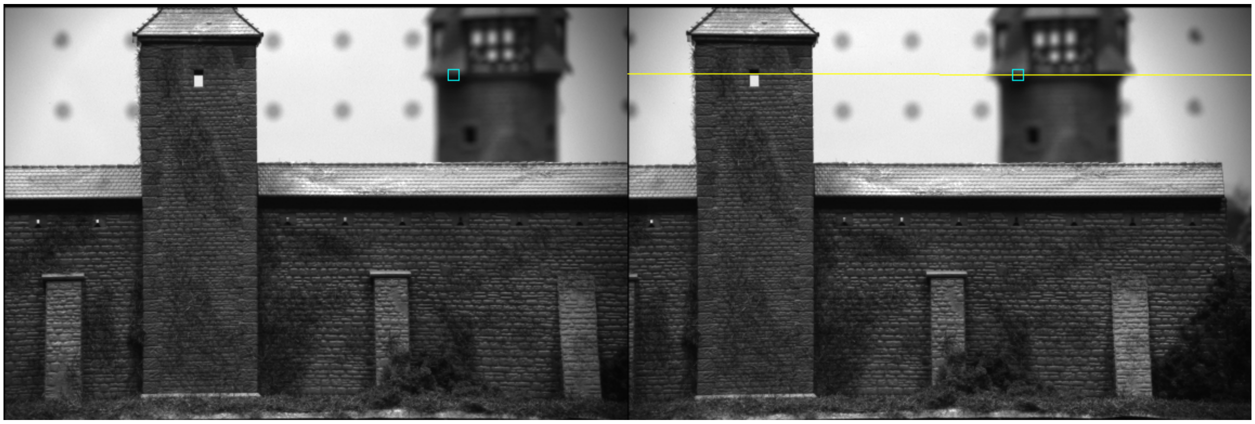


Figure 12: Another example

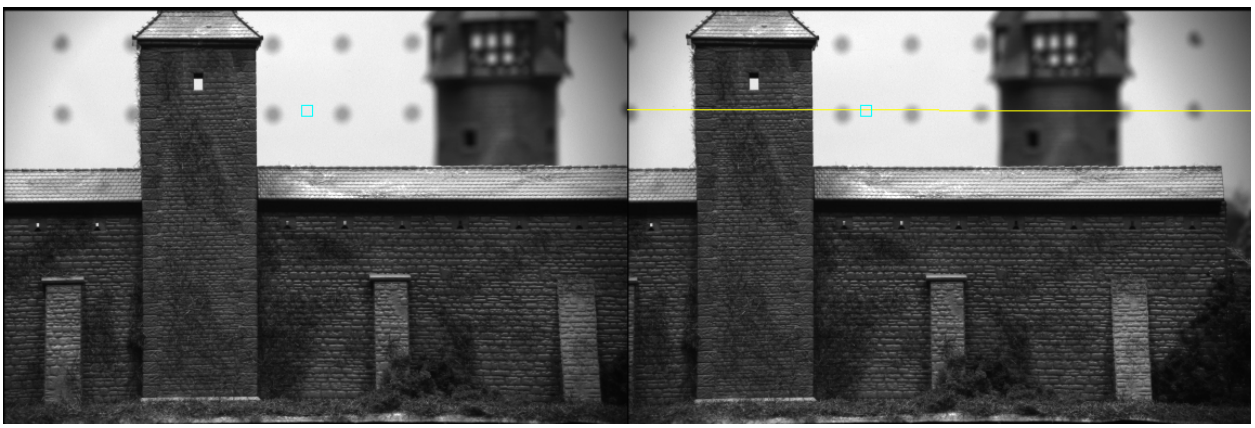


Figure 13: It is clear that the patch of sky on the left is not the same patch of sky on the right, but it does satisfy the epipolar constraint, and scores highly in NCC due to its practically identical contents. This is a case where the disparity calculations will not estimate the true disparity of the point.

The results of our second DDM are shown in Figures 14-16. Similar to the cones dataset, the epipolar lines are horizontal meaning that there is no vertical disparity at all. Again, this also means that the horizontal disparity is essentially a function of the depth of the image, as evident in Figure 15. However, it is apparent that the sky in between the two towers is extremely noisy due to a high number of nonsense correspondences, even with the various constraints used to improve our correspondence search.

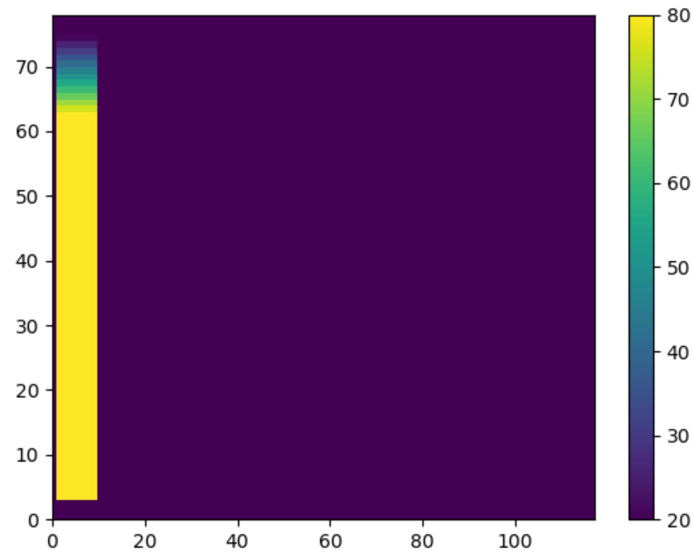


Figure 14: Vertical disparities in image

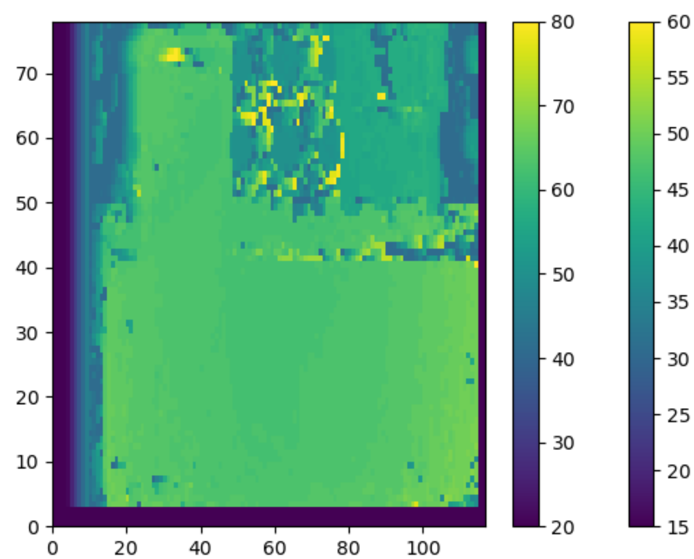


Figure 15: Horizontal disparities in image.

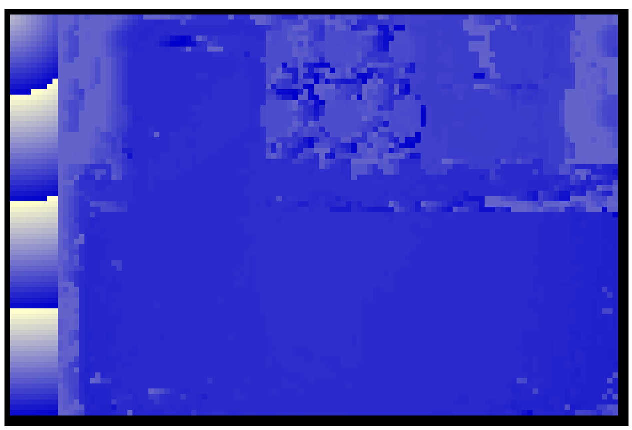


Figure 16: Vectorized HSV version of the disparity map. Observe that there are no vertical disparities, so every pixel is a shade of $0°$, i.e. blue. Darker blue refers to closer elements whereas further away is lighter.

5 Appendix: Code

Please see our git repository for copies of our code.

$$https://github.com/EAGrimaldi/EECE5639\_Project3$$