Stereo Depth Map Construction

By:

Emmanuel Fuentes
Pawel Mikolajczyk
Naji Barnes-McFarlane

Stereo Depth Map Construction

By:
Emmanuel Fuentes
Pawel Mikolajczyk
Naji Barnes-McFarlane

Online:
< http://cnx.org/content/col11252/1.1/ >

C O N N E X I O N S

Rice University, Houston, Texas

Collection structure revised: December 14, 2010
PDF generated: February 6, 2011
For copyright and attribution information for the modules contained in this collection, see p. 29.

Table of Contents

Chapter 1

Introduction to Computer Vision[1]

1.1 Stereo Imaging

The human eyes both take independent “photos” at different angles and positons. The remarkable thing about the human brain is that it is able to use both images to recreate the 3D world. It does this very quickly and remarkably accuratly.

To replicate this in computers is a crucial part of image processing and robotics. A computer's ability to understand its own position compared to its surroundings is a crucial step forward. This would allow the ability for robots to go where humans could not and “learn” useful information about their surroundings.

In the scope of this project we try to replicate basic Stereo Imaging given two photos taken at different horiztonal positions from eachother. We are going to rebuild the 3D space possible with two images, with depths from the camera plane.

[1]This content is available online at <http://cnx.org/content/m36393/1.1/>.

Chapter 2

Correspondence Problem[1]

2.1 Correspondance Problem

2.1.1 Stereo Image Matching

Given any two images of the same scene one may want to understand where given parts of the picture exist in the other. Most often is it virtually impossible to take identical picutres with the same scene in exactly the same positions for two or more images.

There are two methods of corresponding two images. The classical method is to analyze a location on an image and see where is it most like on the other image. This is called the **Correlation-Based Method**. The more robust and practical method is the **Feature-Based Method** that indentifies unqiue features in one image and finds the same features in the other image.

A **Feature** is best desribed as a unique piece of the image that is not repeated anywhere else in the picture. There are many methods to indentifing features in a given image, however the method that lent itself well to our partcular problem of replicating stereo imaging is the SIFT Alogrithm.

[1]This content is available online at <http://cnx.org/content/m36386/1.1/>.

Chapter 3

SIFT Alogrithm[1]

3.1 SIFT Alogrithm

3.1.1 Scale Invariant Feature Transform Alogrithm

This algorithm was developed by David Lowe in 1999 at the Univeristy of British Columbia. Through it methodology the **SIFT Alogrithm** is invariant to scale, light,noise and other commons changes that can effect an object's representation in an image. The alogrithm best identifies objects with clear edges, points of high contrast, and stable fundamental geometry that will not change from picture to picture. The features the method displays, not only have their respective x,y corrdinates in the photo, but also have an oriention in radians, and a scale factor to describe the size of the feature. This allows for accurate description of the uniquness of the feature. All of this lends itself well to solving the correspondance problem presented with two or more images.

Regardless of scale or orientation or position the SIFT alogrith will be able to tell where a particular feature is based on the geometry of the photo and the features neariest to the feature being matched, as well as its descriptors.

[1]This content is available online at <http://cnx.org/content/m36387/1.1/>.

Chapter 4

SIFT Alogrithm Feature Extraction Example[1]

4.1 Feature Extraction

4.1.1 Application of Feature Extraction

Two images of the same bear on a desk taken at different horitzonal positions relative to the subject were taken and anaylzeed using the SIFT algorithm implemented in Matlab with the aid of a toolbox by VLFeat.

[1]This content is available online at <http://cnx.org/content/m36388/1.1/>.

Picture A

Figure 4.1: Taken on a DSLR Nikon D3000

Picture B

Figure 4.2: A horizontal shift from camera position in Picture A

After the SIFT algorithm has been applied corresponding features are very clear and intuitive.

Zoomed In Feature Extraction Picutre A

Figure 4.3: Features(Green) with scale and orientation

Zoomed In Feature Extraction Picutre B

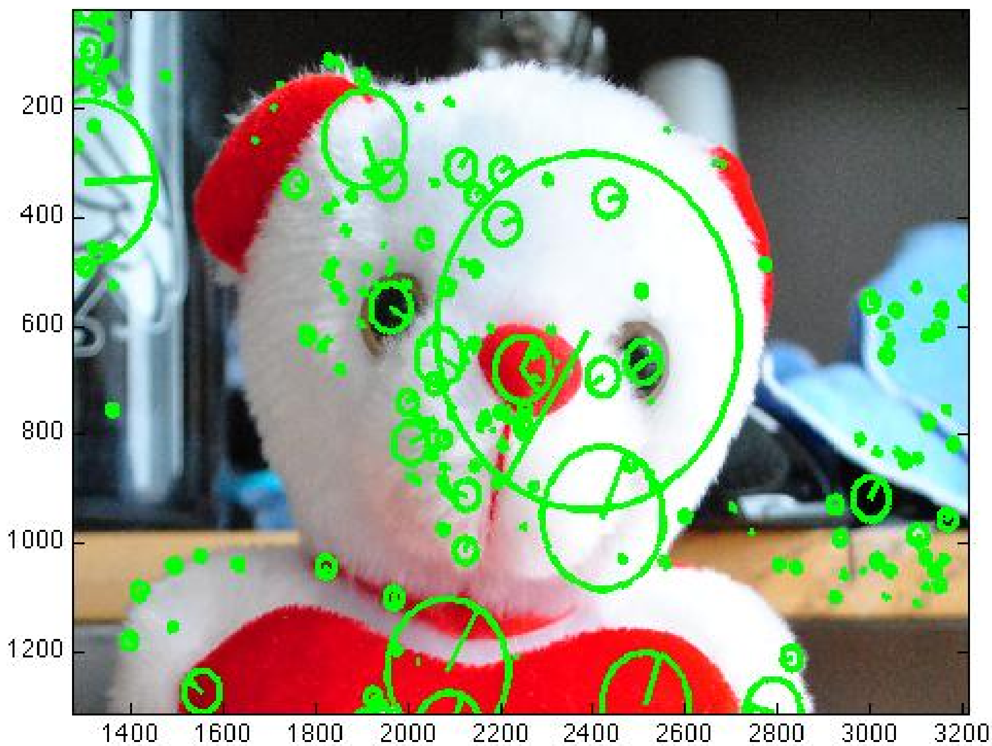

Figure 4.4: Features(Green) with scale and orientation

It is clear that the SIFT program chooses the same features of the same scale and orientation in each image, and there is a clear match between the two images.

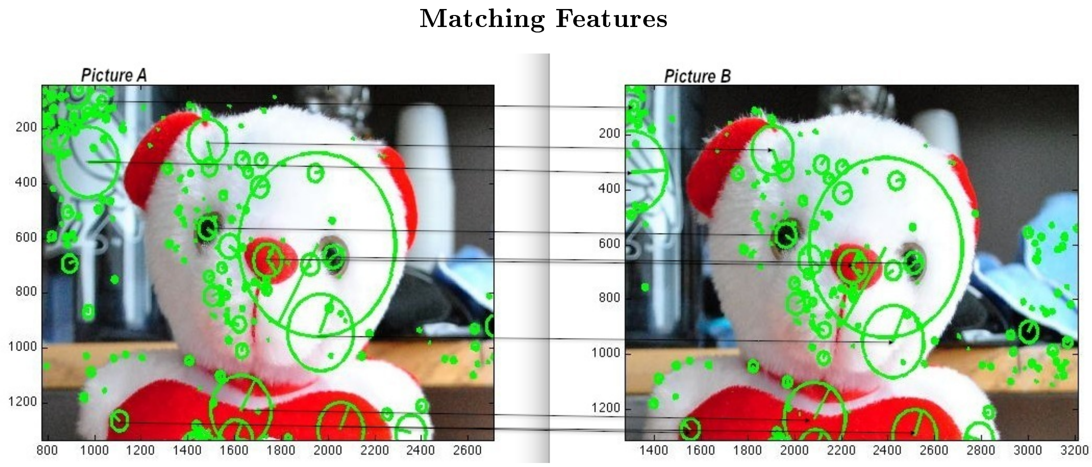

Figure 4.5: Black line drawn from center of corresponding features.

Chapter 5

Triangulation[1]

5.1 Triangulation

Once the SIFT algorithm has been applied, one has positions of corresponding features in each image. To rebuild each point in 3D space, to undo the projection induceded by the camera, we represent our camera using the pin hole model.This assumption allowas us to draw a line from the camera position to the feature, obtaining an equation for a line that can be used to approximate furthur points down that line beyond our feature. Obtaining equations for each photo and its respective camera we can build a system of equations to see where these two lines intersect. The position of camera one and camera two are varied by some horizontal distance or **disparity**, similarly as features in photo one and photo two. The figure shows the general idea behind this reconstuction

[1]This content is available online at <http://cnx.org/content/m36390/1.1/>.

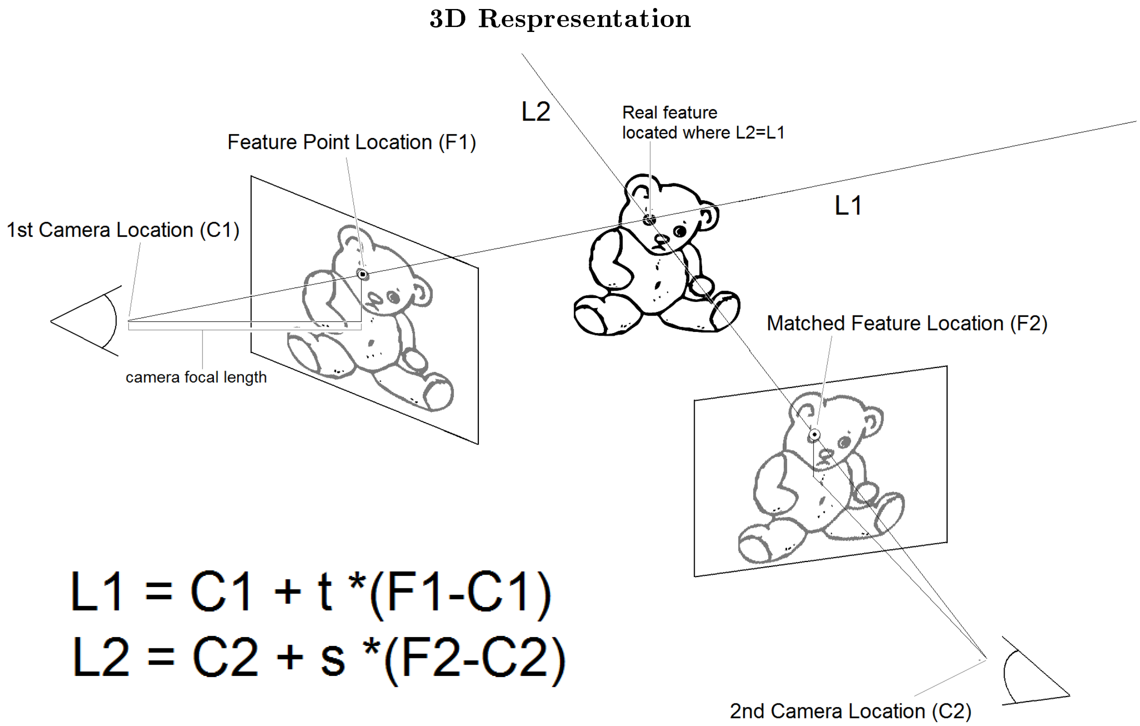

Figure 5.1: Bear at two different angles with corresponding feature.

However this is a "perfect" model that inherently has many flaws. One there is going to be a scale problem because we are not sure if we took the picture with a small or big camera of a small or big "world." This trick can be seen in old movies when you see Godzilla destroying a city, when it is actually a man in a costume jumping on blocks or when a big building is simply a clay miniature shot up close.

The second problem is that this method requires we know the position and the disparity of the cameras before hand to solve this eqation.

Lastly, due to imperfections that will always exist in a photograph, the corresponding features have both vertical and horizontal disparities, not just one or the other, therefore the lines will not truly ever intersect.

Chapter 6

Triangulation of Feature Points[1]

6.1 Minimization of Corresponding Feature's Line Equations

6.1.1 Least Squares Method

Granted in our attempt to solve this problem, one knows the **position** and **disparity** of the cameras, our problem becomes a **minimization problem**. Due to the imperfections inherent in photography, the features of one photo will never truly intersect with that of another. So instead of trying to solve for the intersection of two lines, one can perform **a least squares minimzation** to see where the two lines in space gets closest. Once this value has been obtained, it can be used back into either line equation to find the corresponding point in 3D space. This point with x,y,and z occridnates, z being the distance the point exists away from the camera, now has a porpotial depth to that of the camera, and image plane. If this process is done for each corresponding set of features in both images, we arive at depths of all our feature points from the camera.

6.1.2 Disparities Error

The least squares method is the most realistic. Given the focal length, and positions of the camera the entire world can be recreated, with porportional scale and correct orientation. However, this method in practice is the most unstable, a couple of incorrect matches with the SIFT algorithm can reprt incorrect depths and throw off a whole region of the depth map. In addition even with the least squares method, if the feature points are incorrect the depth will be wrong. This is why a very high threshold for the SIFT algorithm is desirable to get rid of less correlated matches. Our group would rather have fewer features but be correct, than a lot of features with high error.

In response to this instability, we also plotted porportial depth maps simply using disparities of the camera and the corresponding features to get a sense of the relative depths. This method does not out put correct values for Z but is much more stable than potential incorrect least sqaures method pinpointing.

[1]This content is available online at <http://cnx.org/content/m36391/1.1/>.

Chapter 7

Depth Maps[1]

7.1 Depth Map Generation Using SIFT and Minimzation Alogorithms

7.1.1 Feature Depths to Depth Map

Using the griddata and surf functions in Matlab, we are able to interpolate the z data of each feature in 3D space, to see what the rest of the regions in between features is doing. Then by laying an original picture over this data we can see visually how deep regions of our photographs are in space.

[1]This content is available online at <http://cnx.org/content/m36392/1.1/>.

7.1.1.1 Given these Two Images:

Picture A

Figure 7.1: Bear on a desk, with a wall and a deep hallway.

Picture B

Figure 7.2: Same scene,different camera position.

7.1.1.2 Applying our least squares alogroithm:

Depth Map with SIFT and Least Squares Minimzation

Figure 7.3

NOTE: Aside from the extremes in the very front and back left corner, depths make sense. Shallow over bear, deeper to the left with the near by wall, and very deep to the right with the door.

As was mentioned earlier, the least squares method is more unstable and is visably so with the extremes in the front of the photo and in the back left corner. However the rest of the photo makes intuitive sense. All the reported Z values are off by a scale factor due to the explained ambiguity.

Generalized Range of Depths (proportional to real value in inches)

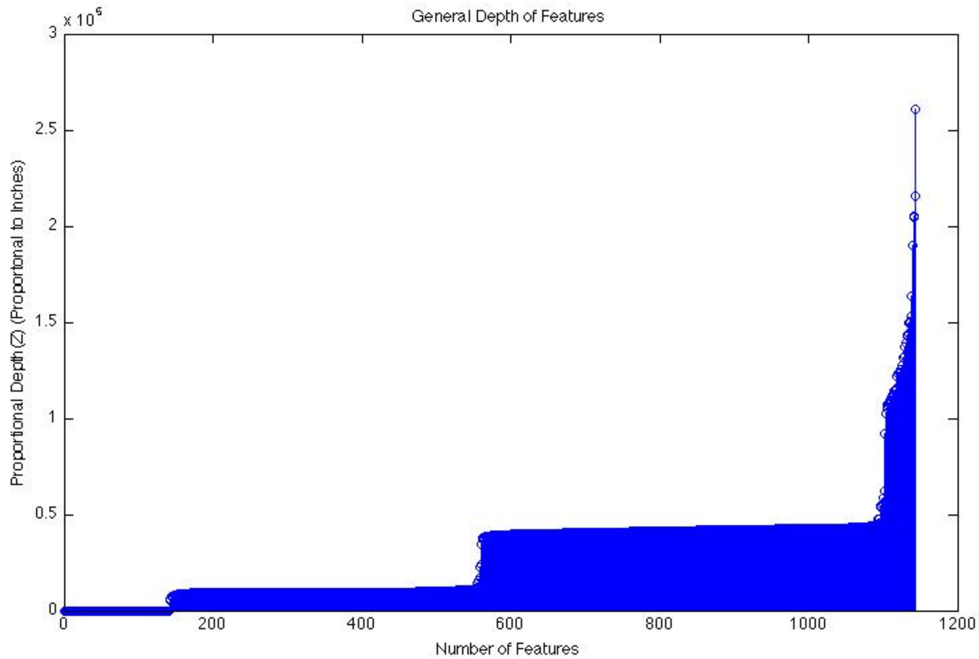

Figure 7.4

NOTE: Region 200 to 600 is the bear's depth, 600 to 1000 is the background(wall), and 1000 to the end is the extreme depth represented by the hallway.

Porportionally this graph makes sense because the wall was in fact about 3 times as far as the bear was to the camera, and the depth of the hallway was obviously extreme compared to the wall because we could not see the end of it in the photo, it seemed to go to infinity.

Therefore these results and the corresponding depth map of the two steoro images fits our expectations.

7.1.2 Depth Map Simply Using Disparites

Depth Map just using Disparites as the Basis

Figure 7.5: Bear raised,hallway w/ extreme depth

This method is much more stable and demesntrates better quick changes in depth than the previous method. This algorithm understand that there is a huge drop from the depth of the bear to the hallway to the right of it.

Data for this section is not important because all values do not take into consideration 3D space, simply the disparities of the features and the camera positions.

7.1.3 Another Example

To further test our alogorithm on a large depth, we shot two stero images of the Duncan wall way and used the Disparity method as shown above to get an idea of the relative depths.

Orignal Photo of Duncan Hall

Figure 7.6: Large Depth w/terminating end,and shallow walls.

Depth Map using Disparites Angle 1

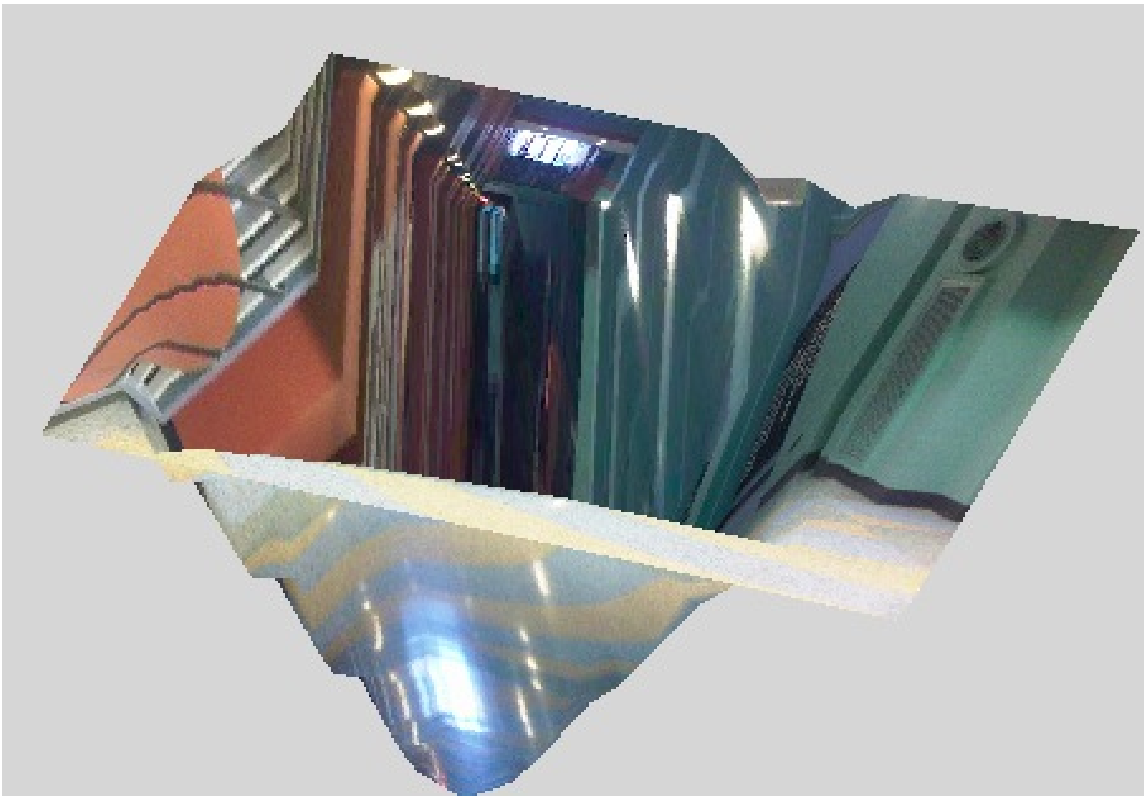

Figure 7.7

NOTE: Notice large Depth that terminates correctly where the end of the hall should be.

Depth Map using Disparites Angle 2

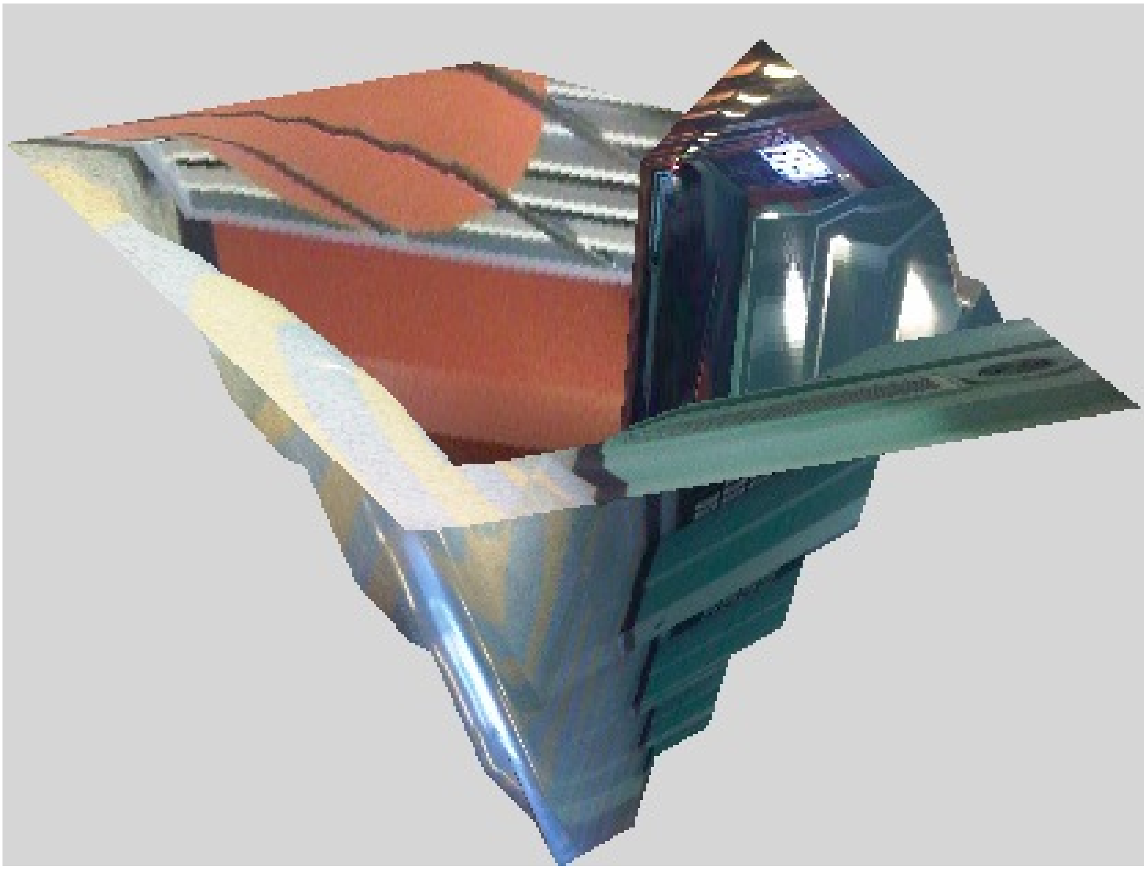

Figure 7.8

NOTE: Notice how Side Wall Depth is accurate with jagged depths that match the surface of the real walls in Duncan Hall.

Once again we see that this algorithm is very powerful in transforming two stero images into a depth map, that is also very accurate in its proportional depth from camera.

Chapter 8

Conclusion and Acknowledgements[1]

Conclusion

In conclusion replicating Stereo Imaging was a success. We achieved proportional depth maps, that accurately map the 3D space of a scene given two images of the same scene at different angles. Granted there was a lot of control in this experiment the given camera positions and disparities, it was prone to many instabilities. These instabilities did not ruin the depth map, but threw off certain regions that could be accounted for. However when two photos were deliberately taken with terminating ends in view, the algorithm work very well. Future scope of this project would be given many images of the same scene at many different angles and without camera positions we could recreate a complete 3D replication of the scene, object. This could then be used to replicate intricate worlds and scenes with incredible detail.

Acknowledgements

A big thanks to our Professor Richard Baraniuk, who led a fantastic class, our TA Matthew Moravec for putting in all his effort and work,the rest of the ELEC 301 Team (Connexions Group,ECE Department,Study Session TAs, and our peers of the ELEC Class of 2012). A special thanks to Aswin Sankaranarayanan for being a fantastic mentor and teaching us material that was beyond the scope of our current knowledge to achieve the goal of this project.

Thanks for a Great Semester. Naji Barnes McFarlane Pawel Mikolajczyk Emmanuel Fuentes

[1]This content is available online at <http://cnx.org/content/m36394/1.1/>.

Index of Keywords and Terms

Keywords are listed by the section with that keyword (page numbers are in parentheses). Keywords do not necessarily appear in the text of the page. They are merely associated with that section. *Ex.* apples, § 1.1 (1) **Terms** are referenced by the page they appear on. *Ex.* apples, 1

Attributions

Collection: *Stereo Depth Map Construction*
Edited by: Emmanuel Fuentes, Pawel Mikolajczyk, Naji Barnes-McFarlane
URL: http://cnx.org/content/col11252/1.1/

Module: "Introduction to Computer Vision"
By: Emmanuel Fuentes
URL: http://cnx.org/content/m36393/1.1/
Page: 1

Module: "Correspondence Problem"
By: Emmanuel Fuentes
URL: http://cnx.org/content/m36386/1.1/
Page: 3

Module: "SIFT Alogrithm"
By: Emmanuel Fuentes
URL: http://cnx.org/content/m36387/1.1/
Page: 5

Module: "SIFT Alogrithm Feature Extraction Example"
By: Emmanuel Fuentes
URL: http://cnx.org/content/m36388/1.1/
Pages: 7-12

Module: "Triangulation"
By: Emmanuel Fuentes
URL: http://cnx.org/content/m36390/1.1/
Pages: 13-14

Module: "Triangulation of Feature Points"
By: Emmanuel Fuentes
URL: http://cnx.org/content/m36391/1.1/
Page: 15

Module: "Depth Maps"
By: Emmanuel Fuentes
URL: http://cnx.org/content/m36392/1.1/
Pages: 17-25

Module: "Conclusion and Acknowledgements"
By: Emmanuel Fuentes
URL: http://cnx.org/content/m36394/1.1/
Page: 27

Stereo Depth Map Construction
This project conducted in Fall 2010 at Rice University is about recreating a depth map of a scene/objects given two or more images of that scene at different angles/positions.