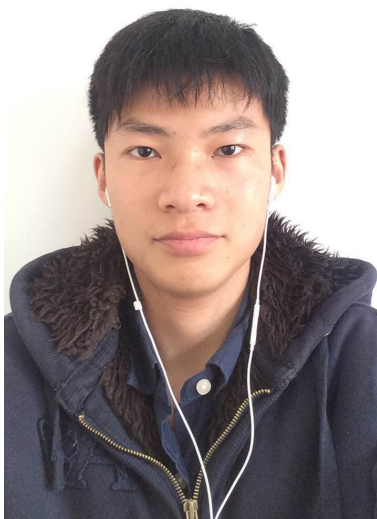


Assignment 5: Depth from Stereo

Suozhi Qi
3/11/17

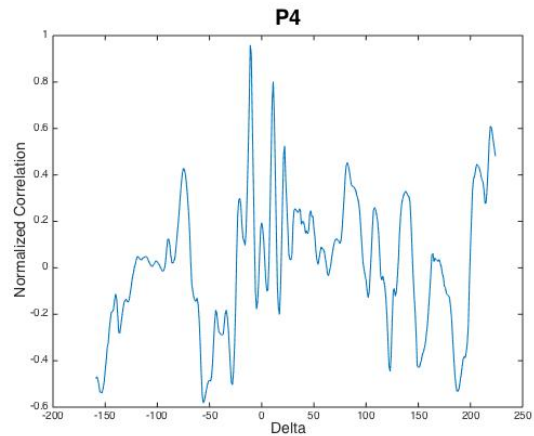
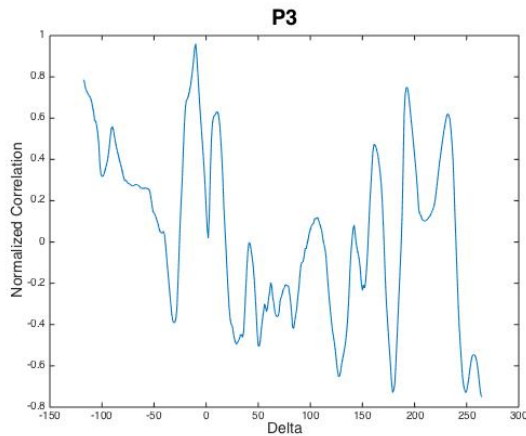
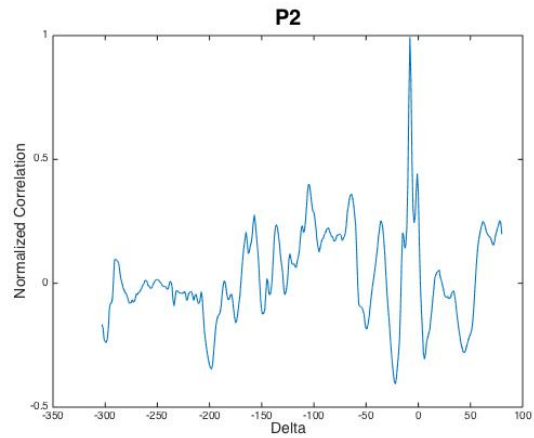
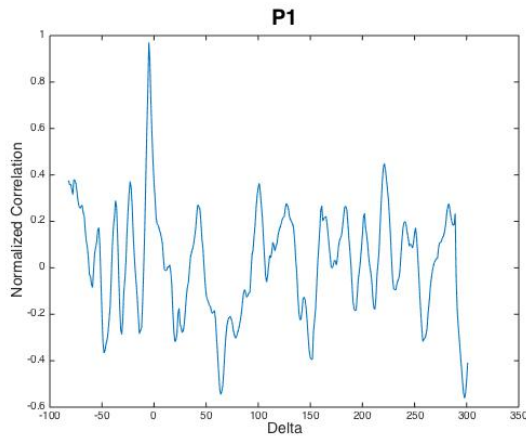


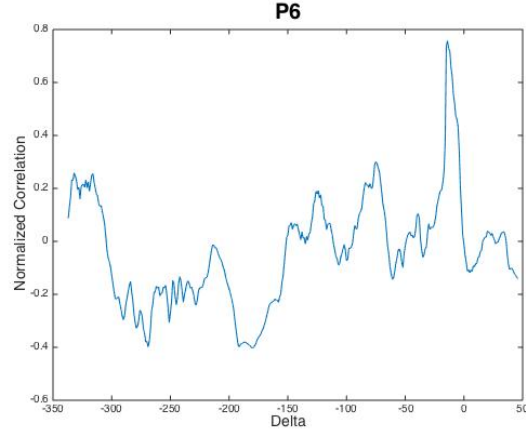
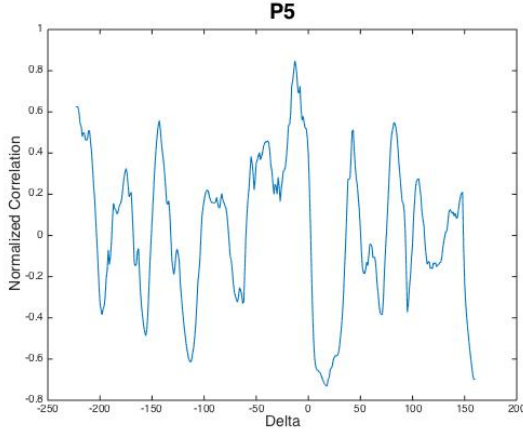
1 Depth from Binocular Stereo

1.1 Solving correspondence

1 I chose patch size to be $23 * 23$ after many tests. The similarity metric I chose is normalized correlation. The reason why I chose $23 * 23$ is because this size is big enough to take the surrounding pixels of a target pixel into consideration. At the same time, it is not too big to waste too much time in calculation. As to the similarity metric, I first tried to use SSD. But then I found it is hard to normalize across all the pixels in an image in this way. So I switched to normalized correlation.

2 I have used my matlab code to profile all of the given 6 pixel points, and got their corresponding similarity profile. From left to right, and up to bottom, are p1, p2, p3, p4, p5, p6 respectively.

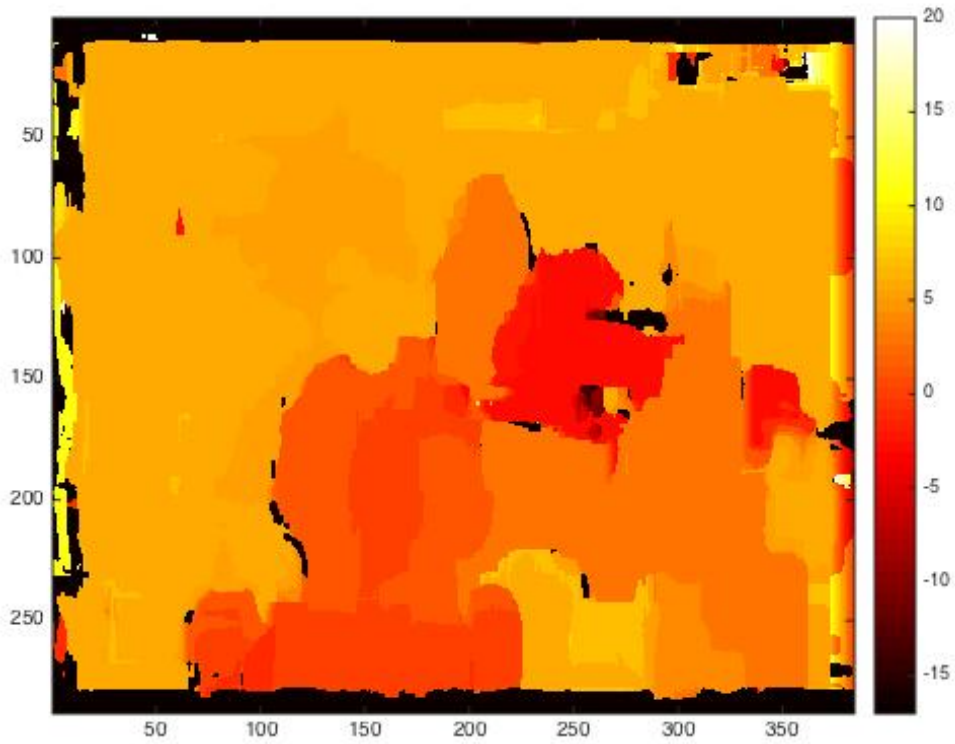




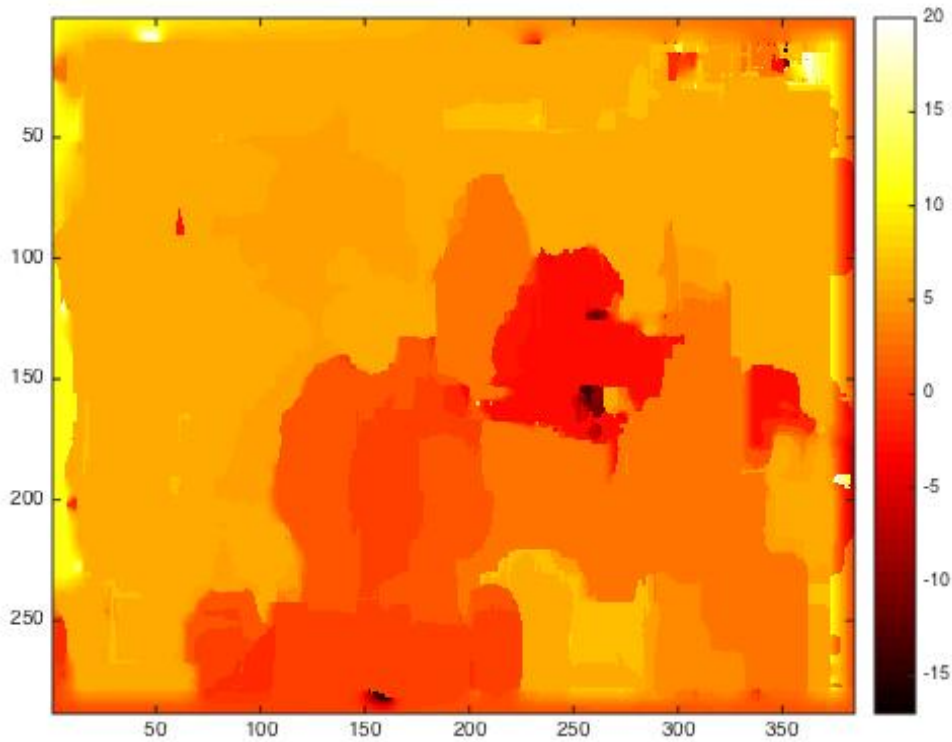
From my 6 similarity profile images, we can clearly see that there is a reliable extremum in each of them. These extremum all happen around 0, or more precisely, a little to the left of 0, indicating that these two images are taken by two cameras located very closely to each other. These extremums are all close to 1, indicating very good correlation. The reason that these profile pics have a reliable extremum is because the left image and right image are closely correlated with each other. Also we can see that the local extremum of each pic is located to the left of 0 (negative numbers), leading us to conclude that the disparity is to the left of the image.

3 I first use the `getMatch` function in my matlab code to get all of the similarity profiles of each individual pixels. Next, I find all of the maximum normalized correlation in each of the pixel points. I used two criteria to determine whether I will accept that correlation as the most similar patch: first, the maximum value has to be larger than my threshold. In this question, I set the threshold as 0.55; second, the correlation point with the maximum correlation value has to be within a certain pixel distance from the point to be matched. In this case, I set this threshold distance to be 20 pixels. If a pixels point matches both of these criteria, then I update its value with the disparity. Otherwise I will just set it as NaN.

4 Refer to my matlab code. It takes about 10 minutes to run the whole thing. The disparity map is show below.



5 To fill in the undefined values, I took advantage of an imported function `inpaint_nans` by John D'Errico. Among all of the methods in the imported function, I chose method 2, which uses `del` 2 but solving a direct linear system of equations for nan elements. I chose this method because this one is the fastest since it uses the sparsest possible system of equations. The disparity map is shown below.



6 Smoothness: It is a measure of how similar colors that are close together are. It tends to fail at discontinuities of depth [1].

Uniqueness: Almost always, a given pixel or feature from one image can match no more than one pixel or feature from the other image.

This constraint fail if transparent objects are present in the scene [1].

Ordering: The ordering of features is preserved across images.

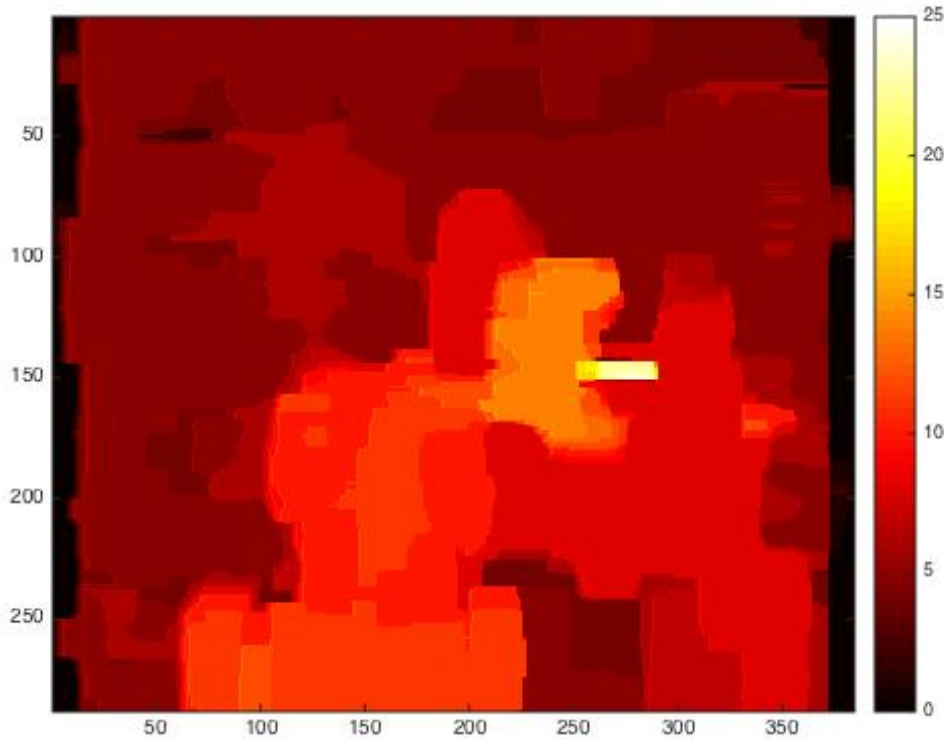
This constraint fails at regions known as the forbidden zone [1].

1.2 Scanline stereo

1 Key points: This paper presents a purely computational approach to get an optimal matching surface in a 3D space. Basically, this algorithm consists of two searches: one called as "inter-scanline search" to find correspondence of connected edges, and the other as "intra-scanline search" to search for correspondence of every edge. They propose a dynamic programming method for both of these two searches (inter and intra) simultaneously.

Until section 3.2 of this paper, it mainly talks about the dynamic programming method of "intra-scanline search". This search can be treated to find matching paths on a 2D plane whose axes are the right and left scanlines. But intra-scanline searches does not consider mutual dependency between scanlines, and this is the place where inter-scanline searches come to our eyes. Smoothness, uniqueness and ordering all can improve the quality of correspondences. Smoothness helps to choose the most possible edges, and uniqueness and ordering helps to get rid of false-positive pairs and narrow down the range of selection.

2 Please refer to my matlab code for execution details. Basically, I use DP to find the best path to maximize the correlation of all points, thus to get the optimal disparity for each pixel point. Shown below is the image.



2 Grad Credits: Active 3D Cameras

Two Kinect range sensing technologies, structure-light (SL) and Time-of-Flight (ToF) are compared in this paper using a framework of seven different experimental setups. In 2010, Microsoft launched a SL based range sensing camera and later released the newest ToF based Kinect camera.

The SL approach is an active stereo-vision technique. From a projector, certain known patterns are sequentially projected to the object then observed from a camera in another direction. The distortion of object can be calculated based on parameters of the camera and the baseline between camera and projector. By analysis, the depth information of pixels can be extracted. Based on the standard structured light principle, the kinect^{SL} camera is made up of a color RGB camera, a monochrome NIR camera and a projector. The computation of depth information is implemented by simple triangulation techniques. The other ToF technology is based on the timing of light traveling to an object and then comes back. The kinect^{ToF} takes advantage of Continuous Wave (CW) Intensity Modulation technique. By applying periodic light, a phase shift would occur, which can be computed into sensor-object distance.

There are multiple sources of errors in each of these two systems. First, both ToF and SL cameras may suffer from ambient background light, which may lead to over-saturation in case of long exposure times. Second, interference problems may occur when multiple Kinect cameras are used simultaneously. Third, temperature drift is also a potential problem, and TOF approach produces more heat than SL approach, leading to the installation of an active cooling device in the TOF cameras. Besides, both TOF and SL cameras suffer from systematic errors in their depth measurement. As to the SL cameras, the errors are mainly caused by inadequate calibration and restricted pixel resolution. For TOF cameras, on the other hand, errors are from the approximation during distance calculation.

Other sources of errors also contribute to the whole sources of errors in the active 3D system. Depth inhomogeneity, for example, which is caused by lack of depth information as a result of occlusion may act as a source of error. Intensity-related distance error as a result of reduced SNR also is another source of error. Besides all of these mentioned factors, semitransparent and scattering media, which causes imperfect reflection of incident light, and dynamic scenery, which may alter the true depth, are all contributing to the errors of both TOF and SL cameras.

ToF and SL cameras both has their advantages and disadvantages. SL cameras delivers more robust depth values than ToF cameras throughout the ambient background light range. As to multi-device interference, the active frequency pattern of the ToF camera has a stronger interference than for the SL cameras. As to device warm-up, ToF cameras generally has less error than SL cameras and SDA and RMSE are nearly constant over time. For rail-depth tracking, SL cameras deliver much stronger out-of-plane errors than ToF ones. Besides, in face of semitransparent liquid, SL cameras have great performance for thicker semitransparent liquids but fails for thinner cases, while ToF cameras often fail to indicate the pixels' invalidity. As to reflective board, SL cameras show less problems with indirect lighting setup, but they also have limitations for low angles and they deliver a higher invalid depth for low incident angles. ToF cameras, on the other hand, have more problems with the superposition of the indirect illumination. They are capable of detecting some of the corrupted pixels but they tend to fail at low angles. Last, as to turning Siemens star, in the situation of increasing speed, ToF cameras are more reliable than SL cameras.

Compared with passive 3D system, active imaging system has some advantages. First, it could give accurate depth information. Second, it gives fast 3D information. Third, it could deliver real time information. Last but not least, active 3D system is of low cost and is portable.

Of course, active imaging system also have some disadvantages. For example, active imaging systems use standard optics to focus the reflected light onto the chip area. Therefore, typical optical effects need to be corrected. Besides, active imaging system may introduce multi-device interference errors into the system. Last but not least, media that does not perfectly reflect the incident light potentially brings errors into active 3D system.

3 References

[1] http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT11/node5.html