Stereo Vision
COMP90072 PROJECT

It is clearly obvious that most animals have two eyes, referred to as Binocular vision. Why is this so ? Well one good reason is that having two is advantageous as it provides a much wider field of view. For example, humans have a maximum horizontal field of view of approximately 190 degrees with two eyes ! Approximately 120 degrees of this is the binocular field of view (seen by both eyes) flanked by two uni-ocular fields (seen by only one eye) of approximately 40 degrees. Why is this so ? In other words, why overlap so much of each field of view? Well more importantly than the wide field of view, having two eyes means that we are able to perceive a single three-dimensional image of our surroundings. This is referred to as stereopsis, in which binocular disparity (or parallax) provided by the two eyes' different positions on the head gives precise depth perception.

In stereo vision, two cameras instead of eyes are displaced horizontally from one another, and used to obtain two differing views on a scene, in a manner similar to human binocular vision. By comparing these two images, the relative depth information can be obtained in the form of a disparity map, which encodes the difference in horizontal coordinates of corresponding image points. The values in this disparity map are inversely proportional to the scene depth at the corresponding pixel location.

Computer stereo vision involves the extraction of 3D information from digitally acquired stereo images of the same scene. By comparing information about a scene from two vantage points, 3D information can be extracted by examination of the relative positions of objects in the two panels.

But how do you know what to compare? As humans, this is easy, we look for similar objects and compare. With a computer, this presents a challenge, as to we need to work out how to define and locate objects. There are many techniques, and all involve image processing. For example, “smoothness”, a measure of how similar colors that are close together are can be used to compare images. The underlying premise here is that objects are more likely to be colored with a small number of colors. So if we detect two pixels with the same color they most likely belong to the same object. The method described above for evaluating smoothness is based on information theory, and an assumption that the influence of the color of a voxel influencing the color of nearby voxels according to the normal distribution on the distance between points. The model is based on approximate assumptions about the world. Another method based on prior assumptions of smoothness is auto-correlation.

Smoothness is a property of the world, not of an image. For example, an image constructed of random dots would have no smoothness, and inferences about neighboring points would be useless. Luckily, we don’t live in such a world. The human vision system seems to have evolved to use smoothness to interpret the world around us.

In this project, we will investigate how Binocular/Stereo vision works. To do this, we will firstly explore how template matching works. Part of this will involve looking for a known object in a busy environment. This will involve procedures such as Gaussian fitting and correlations. These are very powerful techniques with a wide range of application. We will then explore how to correct images for parallax and other distortions, cross calibrate images and use this information to create a mapping algorithm to extract depth data from two planar images. Armed with these techniques, and the software to implement them, we will tackle a full blown stereo 3D problem with an “unknown” object.