1 Introduction

How to make a robot view the real world like out human beings? The best method is to mimic human's eye vision system, which is called stereo vision system. It is common sense that only through two eyes, we could get the accurate 3D information of any objects; thus, the aim for this mimic process is to rebuild disparity map from two stereo photo pairs.

same area from a slightly different angle. The two eye views have plenty in common, but each eye picks up visual information the other doesn’t.

Two Eyes = Three Dimensions (3D)! Each eye captures its own view and the two separate images are sent on to the brain for processing. When the two images arrive simultaneously in the back of the brain, they are united into one picture. The mind combines the two images by matching up the similarities and adding in the small differences. The small differences between the two images add up to a big difference in the final picture! The combined image is more than the sum of its parts. It is a three-dimensional stereo picture.

1.1 What is Stereo Vision

The word "stereo" comes from the Greek word "stereos" which means firm or solid. With stereo vision you see an object as solid in three spatial dimensions—width, height and depth—or x, y, and z. It is the added perception of the depth dimension that makes stereo vision so rich and special. Generally speaking, stereo vision is to view the world with two eyes.

Unlike horses, humans have two eyes located side-by-side in the front of their heads. Thanks to the close side-by-side positioning, each eye takes a view of the

1.2 What is Disparity Map

Human eyes are horizontally separated by about 50-75 mm (interpupillary distance) depending on each individual. Thus, each eye has a slightly different view of the world. This can be easily seen when alternately closing one eye while looking at a vertical edge. The binocular disparity can be observed from apparent horizontal shift of the vertical edge between both views. In figure 2, it will show you the theory of how a disparity is generated in a geometry view.

2 The Main Methods to Generate the Disparity Map

2.1 The local method

SSD:SSD stands for sum of squared difference. In this method, we use a template window to do image matching. The window size could be N(N must be an odd number), we calculate the sum of absolute difference of those pixels covered by the window and get the minimum SSD, and the disparity value which is corresponding to the minimum SSD is what we need.

NCC:NCC stands for normalized cross correlation. In this method, we use a template window to do image matching. We use f(x,y) and t(x,y) to denote the pixel covered by the window in left and right stereo pairs. The formulation 1 states how to calculate normalized cross correlation. The disparity value which is corresponding to the best NCC(which is closest approach to 1) is what we need.

Dynamic programming: This is one method usually used in stereo matching, which is relatively fast and obtains good results. It assumes that the best disparity not only rely on one pixel itself, it is also dependendent on pixels next to it in the same scanline. As a matter of fact, to get a best disparity d in dynamic programming, we use recursive method.

3 The Product of Stereo Vision

The Bumblebee2 is the next generation Bumblebee stereo vision camera. It provides a balance between 3D data quality, processing speed, size and price. Developed as a drop-in replacement for the original Bumblebee camera, the Bumblebee2 also features double the frame rate and a GPIO connector for external trigger and strobe functionality.