Electronic and Computer Engineering School (ECE)

Diploma Thesis:

"Real Time Stereo Imaging for Guiding Surgery"

Diploma Committee:

Supervisor: Professor Balas Konstantinos
Members: Professor Pnevmatikatos Dionysios
Associate Professor Mania Aikaterini

Technical University of Crete

Electronic and Computer Engineering School

Field of Electronics

Optoelectronics and Imaging Diagnostics

Campus of Technical University of Crete - Kounoupidiana

73100 Chania - Greece

Στον αγαπημένο μου θείο Ηλία.

Acknowledgments:

First of all, I would like to thank my supervisor Professor Balas Konstantinos for giving me the opportunity to work as one of his undergraduate students and develop my diploma thesis with title: "Real Time Stereo Imaging for Guiding Surgery". I would also like to express my respect and admiration towards him. His help and guidance were very important and gave me the strength to go on with my diploma.
Next, I would like to express my sincere gratitude to all the postgraduate students of the Electronic Laboratory of The Technical University of Crete for the time they devoted on me and the help they provided me.
Last but not least, I would like to thank my family and my friends for being next to me from the start of my university studies.

Contents:

Abstract

Precision in surgery procedures is of outmost importance. Technical progress in imaging sensors allowed the development of 3-Dimensional (3D) stereoscopic imaging systems. The use of 3D stereoscopic imaging aids in making the surgery faster and more precise. It provides a realistic view of the tissue being operated on and depth perception that is lacking from conventional video stream. We present an innovative real time stereoscopic imaging system for guiding surgery. The system consists of two orthogonally oriented imaging sensors. A cube beam splitter is used to project the same frame onto both sensors, with a slight displacement. Each frame from each sensor is processed, using Color Anaglyph 3D. The resulting video stream is displayed on a monitor and viewed through 3D red cyan glasses to provide realistic three-dimensional video.

Chapter 1: Stereo Imaging: Geometry and Description

1.1 Stereo Imaging: Introduction

Stereo imaging is the extraction of 3D information from digital images, such as obtained by a CCD camera. 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. This is similar to the biological process Stereopsis.

Human visual system is the most sophisticated and powerful vision solution to observe the environment and extract information. A similar system with the biological vision, stereo imaging, is designed to extract 3D information from digital images and use these for examining the position of objects in two images, to build an advanced object recognition system that recognizes objects in different arrangements (for example when objects are placed one in front of the other), tracking different objects, etc.

Because a stereo imaging system is similar to the human biological system, some of the features are identical. For example, a human has two eyes to see slightly different views of the same environment. A stereo imaging system has two cameras located at a known distance and take pictures of the scene at the same time. Using the geometry of the cameras, we can apply algorithms and create the geometry of the environment.

Stereo imaging works in a similar way to 3D sensing in human vision. It begins with identifying image pixels that correspond to the same point in a physical scene observed by multiple cameras. The 3D position of a point can then be established by triangulation using a ray from each camera. The more corresponding pixels identified the more 3D points that can be determined with a single set of images. Correlation stereo methods attempt to obtain correspondences for every pixel in the stereo image, resulting in tens of thousands of 3D values generated with every stereo image.

Among the advantages of a stereo imaging system can be included its reliability and effectiveness in extracting various information (like color, or dimension), it can be used for different vision like tracking or detecting objects, and it's a passive sensor which cannot be influenced by the environment.

1.2 Stereo Imaging Outline

In traditional stereo imaging, two cameras, displaced horizontally from one another are 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 disparities, which are inversely proportional to the differences in distance to the objects.

To compare the images, the two views must be superimposed in a stereoscopic device, the image from the right camera being shown to the observer's right eye and from the left one to the left eye. In real camera systems however, several pre-processing steps are required. [1]

1. The image must first be removed of distortions, such as barrel distortion to ensure that the observed image is purely projectional.
2. The image must be projected back to a common plane to allow comparison of the image pairs, known as image rectification.
3. An information measure which compares the two images is minimized. This gives the best estimate of the position of features in the two images, and creates a disparity map.
4. Optionally, the disparity as observed by the common projection is converted back to the height map by inversion. Utilizing the correct proportionality constant, the height map can be calibrated to provide exact distances.

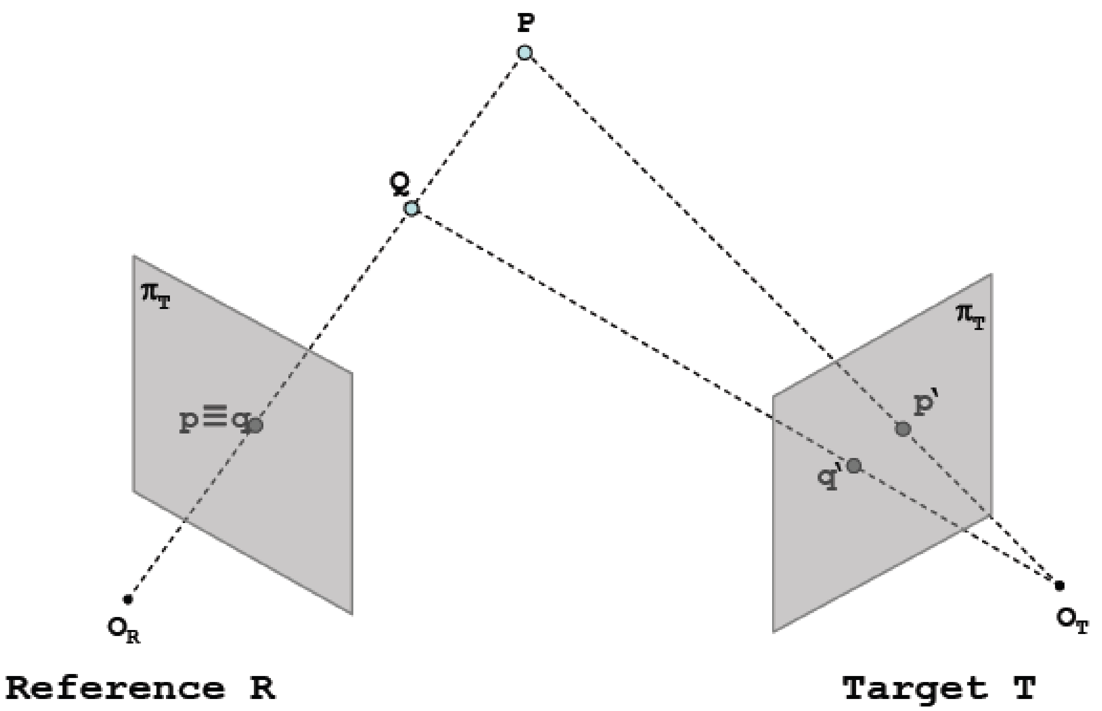

Figure 1. Stereo Camera

1.3 The Geometry of Stereo Imaging

1.3.1 Stereopsis

Stereopsis is a term that is most often used to refer to the perception of depth and 3-dimensional structure obtained on the basis of visual information deriving from two eyes by individuals with normally developed binocular vision. Because the eyes of humans, and many animals, are located at different lateral positions on the head, binocular vision results in two slightly different images projected to the retinas of the eyes. The differences are mainly in the relative horizontal position of objects in the two images. These positional differences are referred to as horizontal disparities or, more generally, binocular disparities. Disparities are processed in the visual cortex of the brain to yield depth perception. While binocular disparities are naturally present when viewing a real 3-dimensional scene with two eyes, they can also be simulated by artificially presenting two different images separately to each eye using a method called stereoscopy. The perception of depth in such cases is also referred to as "stereoscopic depth". [2]

The perception of depth and 3-dimensional structure is, however, possible with information visible from one eye alone, such as differences in object size and motion parallax (differences in the image of an object over time with observer movement), [3] though the impression of depth in these cases is often not as vivid as that obtained from binocular disparities. [4] Therefore, the term stereopsis (or stereoscopic depth) can also refer specifically to the unique impression of depth associated with binocular vision; what is colloquially referred to as seeing "in 3D".

It has been suggested that the impression of “real” separation in depth is linked to the precision with which depth is derived, and that a conscious awareness of this precision – perceived as an impression of interactability and realness – may help guide the planning of motor action. [5]

It has been noted that with the growing introduction of 3D display technology in entertainment and in medical and scientific imaging, high quality binocular vision including stereopsis may become a key capability for success in modern society.

1.3.2 Geometrical basis for stereopsis

Stereopsis appears to be processed in the visual cortex of mammals in binocular cells having receptive fields in different horizontal positions in the two eyes. Such a cell is active only when its preferred stimulus is in the correct position in the left eye and in the correct position in the right eye, making it a disparity detector.

When a person stares at an object, the two eyes converge so that the object appears at the center of the retina in both eyes. Other objects around the main object appear shifted in relation to the main object. In the following example, whereas the main object (dolphin) remains in the center of the two images in the two eyes, the cube is shifted to the right in the left eye's image and is shifted to the left when in the right eye's image.

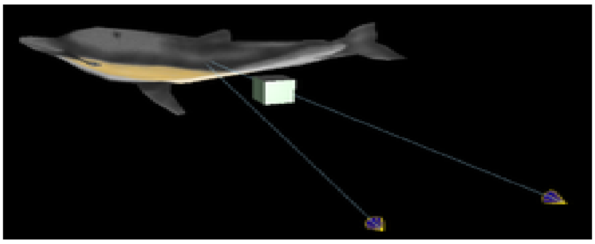

Figure 2. The two eyes converge on the object of attention.

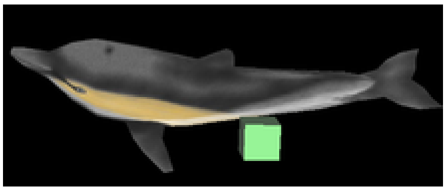

Figure 3. The cube is shifted to the right in left eye's image

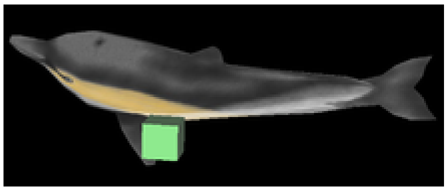

Figure 4. The cube is shifted to the left in right eye's image

Because each eye is in a different horizontal position, each has a slightly different perspective on a scene yielding different retinal images. Normally two images are not observed, but rather a single view of the scene, a phenomenon known as singleness of vision. Nevertheless, stereopsis is possible with double vision. This form of stereopsis was called qualitative stereopsis by Kenneth Ogle.

If the images are very different (such as by going cross-eyed, or by presenting different images in a stereoscope) then one image at a time may be seen, a phenomenon known as binocular rivalry.

1.3.3 Epipolar Geometry

Epipolar geometry is the geometry of stereo vision. When two cameras view a 3D scene from two distinct positions, there are a number of geometric relations between the 3D points and their projections onto the 2D images that lead to constraints between the image points. These relations are derived based on the assumption that the cameras can be approximated by the pinhole camera model.

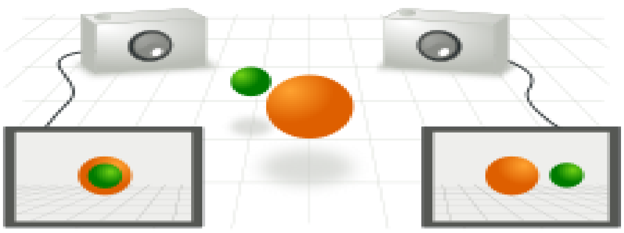

Figure 5. Typical use case for epipolar geometry
Two cameras take a picture of the same scene from different points of view. The epipolar geometry then describes the relation between the two resulting views.

The figure below depicts two pinhole cameras looking at point **X**. In real cameras, the image plane is actually behind the center of projection, and produces an image that is rotated 180 degrees. Here, however, the projection problem is simplified by placing a virtual image plane in front of the center of projection of each camera to produce an unrotated image. $\mathbf{O}_L$ and $\mathbf{O}_R$ represent the centers of projection of the two cameras. **X** represents the point of interest in both cameras. Points $\mathbf{x}_L$ and $\mathbf{x}_R$ are the projections of point **X** onto the image planes.

Each camera captures a 2D image of the 3D world. This conversion from 3D to 2D is referred to as a perspective projection and is described by the pinhole camera model. It is common to model this projection operation by rays that emanate from the camera, passing through its center of projection. Note that each emanating ray corresponds to a single point in the image.

1.3.4 Epipolar point, epipolar line and epipolar plane

Since the centers of projection of the cameras are distinct, each center of projection projects onto a distinct point into the other camera's image plane. As seen in figure 6, these two image points are denoted by $\mathbf{e}_L$ and $\mathbf{e}_R$ and are called epipoles or epipolar points. Both epipoles $\mathbf{e}_L$ and $\mathbf{e}_R$ in their respective image planes and both centers of projection $\mathbf{O}_L$ and $\mathbf{O}_R$ lie on a single 3D line.

The line $\mathbf{O}_L$–$\mathbf{X}$ is seen by the left camera as a point because it is directly in line with that camera's center of projection. However, the right camera sees this line as a line in its image plane. That line ($\mathbf{e}_R$–$\mathbf{x}_R$) in the right camera is called an epipolar line. Symmetrically, the line $\mathbf{O}_R$–$\mathbf{X}$ seen by the right camera as a point is seen as epipolar line $\mathbf{e}_L$–$\mathbf{x}_L$ by the left camera.

An epipolar line is a function of the 3D point **X**, i.e. there is a set of epipolar lines in both images if we allow **X** to vary over all 3D points. Since the 3D line $\mathbf{O}_L$–$\mathbf{X}$ passes through the center of projection $\mathbf{O}_L$, the corresponding epipolar line in the right image must pass through the epipole $\mathbf{e}_R$ (and correspondingly for epipolar lines in the left image). This means that all epipolar lines in one image must intersect the epipolar point of that image. In fact, any line which intersects with the epipolar point is an epipolar line since it can be derived from some 3D point **X**.

As an alternative visualization, consider the points **X**, $\mathbf{O}_L$ & $\mathbf{O}_R$ that form a plane called the epipolar plane. The epipolar plane intersects each camera's image plane where it forms lines—the epipolar lines. All epipolar planes and epipolar lines intersect the epipole regardless of where **X** is located.

1.3.5 Epipolar constraint and triangulation

If the relative translation and rotation of the two cameras is known, the corresponding epipolar geometry leads to two important observations

• If the projection point $\mathbf{x}_L$ is known, then the epipolar line $\mathbf{e}_R$–$\mathbf{x}_R$ is known and the point **X** projects into the right image, on a point **x** R which must lie on this particular epipolar line. This means that for each point observed in one image the same point must be observed in the other image on a known epipolar line. This provides an epipolar constraint which corresponding image points must satisfy and it means that it is possible to test if two points really correspond to the same 3D point. Epipolar constraints can also be described by the essential matrix or the fundamental matrix between the two cameras.

• If the points x_L and x_R are known, their projection lines are also known. If the two image points correspond to the same 3D point **X** the projection lines must intersect precisely at **X**. This means that **X** can be calculated from the coordinates of the two image points, a process called triangulation.

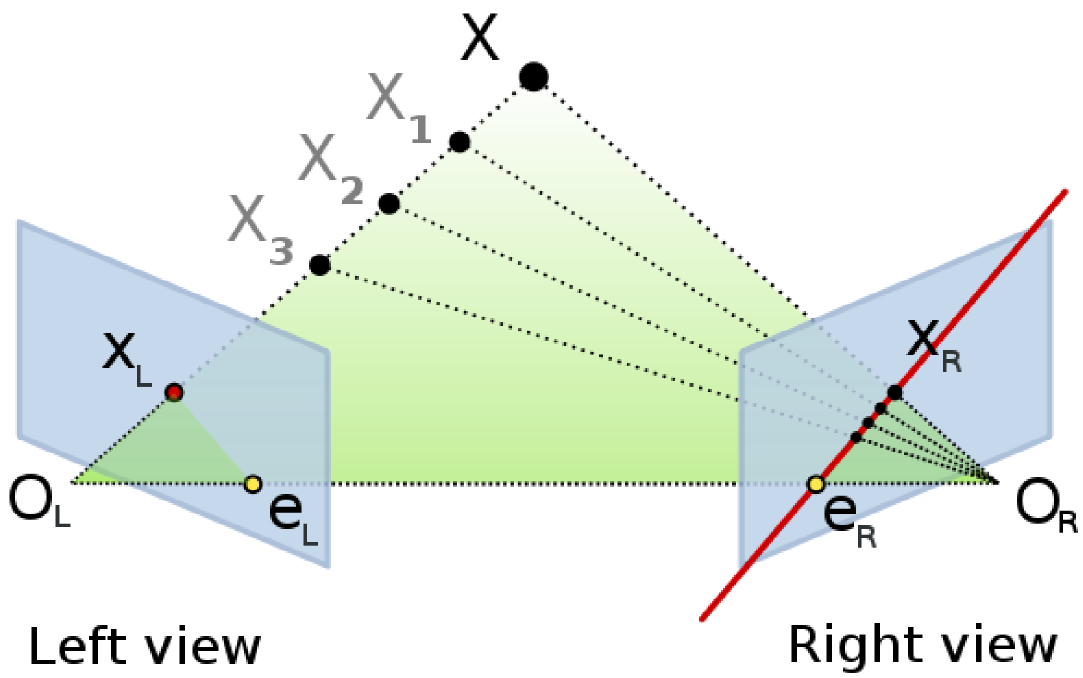

Figure 6. Epipolar Geometry

The epipolar geometry is simplified if the two camera image planes coincide. In this case, the epipolar lines also coincide (E_L–P_L = E_R–P_R). Furthermore, the epipolar lines are parallel to the line O_L–O_R between the centers of projection, and can in practice be aligned with the horizontal axes of the two images. This means that for each point in one image, its corresponding point in the other image can be found by looking only along a horizontal line. If the cameras cannot be positioned in this way, the image coordinates from the cameras may be transformed to emulate having a common image plane. This process is called image rectification.

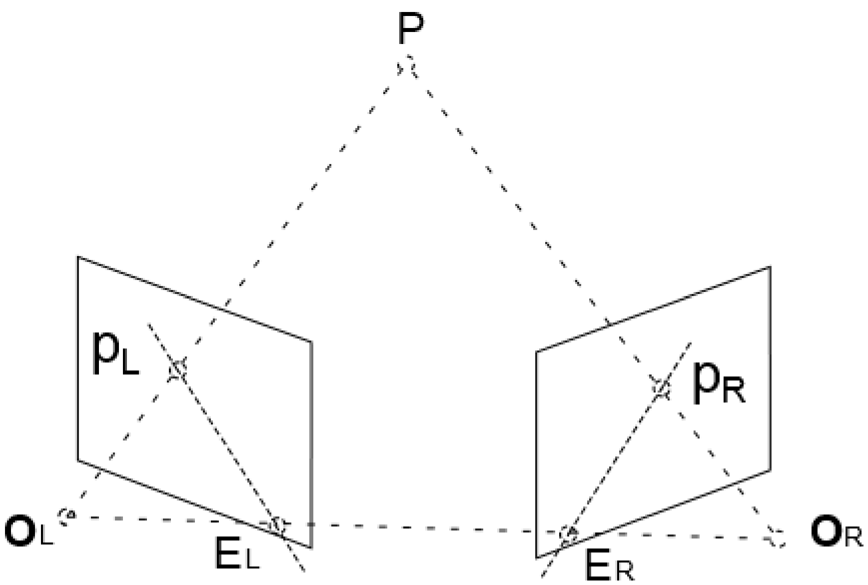

Figure 7.Example of epipolar geometry. Two cameras, with their respective centers of projection points $\mathbf{O}_L$ and $\mathbf{O}_R$, observe a point **P**. The projection of **P** onto each of the image planes is denoted $\mathbf{p}_L$ and $\mathbf{p}_R$. Points $\mathbf{E}_L$ and $\mathbf{E}_R$ are the epipoles.

1.4 Binocular Disparity

Binocular disparity refers to the difference in image location of an object seen by the left and right eyes, resulting from the eyes' horizontal separation (parallax). The brain uses binocular disparity to extract depth information from the two-dimensional retinal images in stereopsis. In computer vision, binocular disparity refers to the difference in coordinates of similar features within two stereo images.

A similar disparity can be used in rangefinding by a coincidence rangefinder to determine distance and/or altitude to a target. In astronomy, the disparity between different locations on the Earth can be used to determine various celestial parallax, and Earth's orbit can be used for stellar parallax.

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 around. 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.

At any given moment, the line of sight of the two eyes meet at a point in space. This point in space projects to the same location (i.e. the center) on the retinae of the two eyes. Because of the different viewpoints observed by the left and right eye however, many other points in space do not fall on corresponding retinal locations. Visual binocular disparity is defined as the difference between the point of projection in the two eyes and is usually expressed in degrees as the visual angle. [6]

The term "binocular disparity" refers to geometric measurements made external to the eye. The disparity of the images on the actual retina depends on factors internal to the eye, especially the location of the nodal points, even if the cross section of the retina is a perfect circle. Disparity on retina conforms to binocular disparity when measured as degrees, while much different if measured as distance due to the complicated structure inside eye.

In the figure 8 the full black circle is the point of fixation. The blue object lies nearer to the observer. Therefore it has a "near" disparity d_n. Objects lying more far away (green) correspondingly have a "far" disparity d_f. Binocular disparity is the angle between two lines of projection in one eye (Mathematically, d_n - d_f, with sign, measured counterclockwise). One of which is the real projection from the object to the actual point of projection. The other one is the imaginary projection running through the nodal point of the lens of the one eye to the point corresponding to the actual point of projection in the other eye. For simplicity reasons here both objects lie on the line of fixation for one eye such that the imaginary projection ends directly on the fovea of the other eye, but in general the fovea acts at most as a reference. Note that far disparities are smaller than near disparities for objects having the same distance from the fixation point.

In computer vision, binocular disparity is calculated from stereo images taken from a set of stereo cameras. The variable distance between these cameras, called the baseline, can affect the disparity of a specific point on their respective image plane. As the baseline increases, the disparity increases due to the greater angle needed to align the sight on the point. However, in computer vision, binocular disparity is referenced as coordinate differences of the point between the right and left images instead of a visual angle. The units are usually measured in pixels.

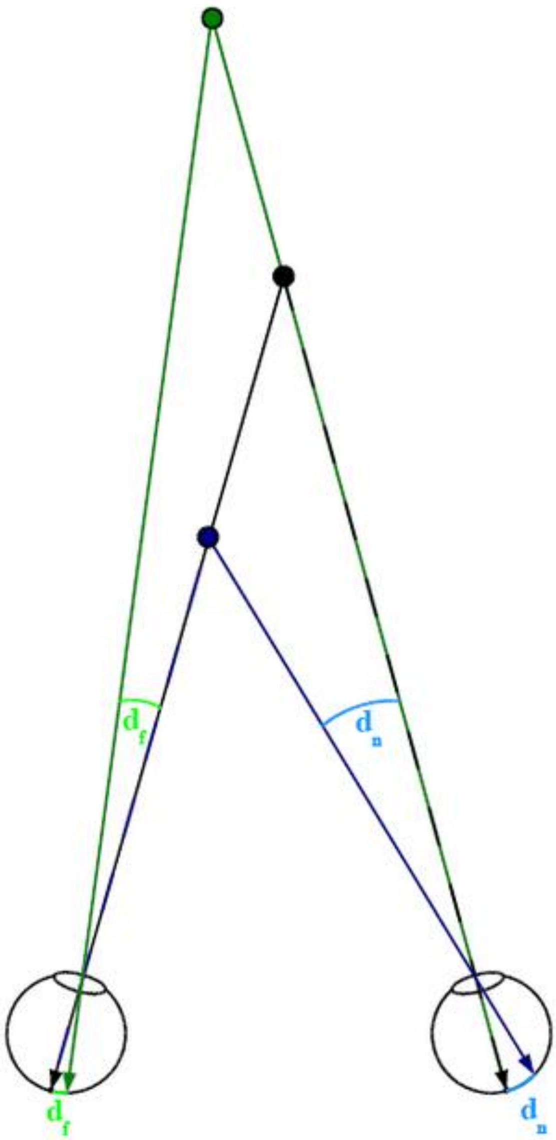

Figure 8. Definition of binocular disparity (far and near).

In the figure 9 the disparity of an object with different depth than the fixation point can alternatively be produced by presenting an image of the object to one eye and a laterally shifted version of the same image to the other eye. The full black circle is the point of fixation. Objects in varying depths are placed along the line of fixation of the left eye. The same disparity produced from a shift in depth of an object (filled coloured circles) can also be produced by laterally shifting the object in constant depth in the picture one eye sees (black circles with coloured margin). Note that for near disparities the lateral shift has to be larger to correspond to the same depth compared with far disparities. This is what neuroscientists usually do with random dot stimuli to study disparity selectivity of neurons since the lateral distance required to test disparities is less than the distances required using depth tests. This principle has also been applied in autostereogram illusions.

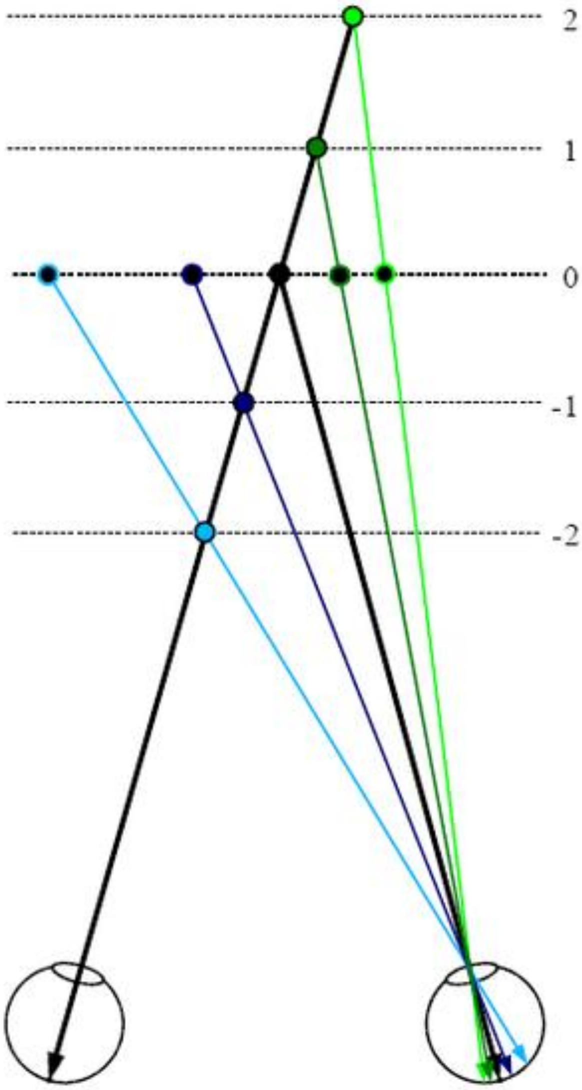

Figure 9. Simulation of disparity from depth in the plane.

1.4.1 Computing disparity using digital stereo images

The disparity of features between two stereo images is usually computed as a shift to the left of an image feature when viewed in the right image. [7] For example, a single point that appears at the x coordinate *t* (measured in pixels) in the left image may be present at the x coordinate *t - 3* in the right image. In this case, the disparity at that location in the right image would be 3 pixels.

Stereo images may not always be correctly aligned to allow for quick disparity calculation. For example, the set of cameras may be slightly rotated off level. Through a process known as image rectification, both images are rotated to allow for disparities in only the horizontal direction (i.e. there is no disparity in the y image coordinates). This is a property that can also be achieved by precise alignment of the stereo cameras before image capture.

After rectification, the correspondence problem can be solved using an algorithm that scans both the left and right images for matching image features. A common approach to this problem is to form a smaller image patch around every pixel in the left image. These image patches are compared to all possible disparities in the right image by comparing their corresponding image patches. For example, for a disparity of 1, the patch in the left image would be compared to a similar-sized patch in the right, shifted to the left by one pixel. The comparison between these two patches can be made by attaining a computational measure from one of the following equations that compares each of the pixels in the patches. For all of the following equations, L and R refer to the right and left columns while r and c refer to the current row and column of either images being examined. "d" refers to the disparity of the right image.

- Normalized correlation:

$$\frac{\sum\sum L(r,c)\cdot R(r,c-d)}{\sqrt{(\sum\sum L(r,c)^2)\cdot(\sum\sum R(r,c-d)^2)}}$$

- Sum of squared differences:

$$\sum\sum(L(r,c)-R(r,c-d))^2$$

- Sum of absolute differences:

$$\sum\sum|L(r,c)-R(r,c-d)|$$

The disparity with the lowest computed value using one of the above methods is considered the disparity for the image feature. This lowest score indicates that the algorithm has found the best match of corresponding features in both images.

The method described above is a brute-force search algorithm. With large patch and/or image sizes, this technique can be very time consuming as pixels are constantly being re-examined to find the lowest correlation score. However, this technique also involves unnecessary repetition as many pixels overlap. A more efficient algorithm involves remembering all values from the previous pixel. An even more efficient algorithm involves remembering column sums from the previous row (in addition to remembering all values from the previous pixel). Techniques that save previous information can greatly increase the algorithmic efficiency of this image analyzing process.

1.4.2 Uses of disparity from images

Knowledge of disparity can be used in further extraction of information from stereo images. One case that disparity is most useful is for depth/distance calculation. Disparity and distance from the cameras are inversely related. As the distance from the cameras increases, the disparity decreases. This allows for depth perception in stereo images. Using geometry and algebra, the points that appear in the 2D stereo images can be mapped as coordinates in 3D space.

This concept is particularly useful for navigation. For example, the Mars Exploration Rover uses a similar method for scanning the terrain for obstacles. The rover captures a pair of images with its stereoscopic navigation cameras and disparity calculations are performed in order to detect elevated objects (such as boulders). Additionally, location and speed data can be extracted from subsequent stereo images by measuring the displacement of objects relative to the rover. In some cases, this is the best source of this type of information as the encoder sensors in the wheels may be inaccurate due to tire slippage.

1.5 Stereo Vision: Detailed definition

A pixel records color at a position. The position is identified by position in the grid of pixels (x, y) and depth to the pixel *z*.

Stereoscopic vision gives two images of the same scene, from different positions. In the diagram on the right light from the point *A* is transmitted through the entry points of a pinhole cameras at *B* and *D*, onto image screens at *E* and *H*.

In the figure 10 the distance between the centers of the two camera lens is *BD* = *BC* + *CD*. The triangles are similar,

- *ACB* and *BFE*
- *ACD* and *DGH*

Therefore displacement *d* = *EF* + *GH* = *BD (BF/AC)* = *k/z*, where,

- *k* = *BD BF*
- *z* = *AC* is the distance from the camera plane to the object.

So assuming the cameras are level, and image planes are flat on the same plane, the displacement in the y axis between the same pixel in the two images is,

$$d = \frac{k}{z}$$

Where *k* is the distance between the two cameras times the distance from the lens to the image.

The depth component in the two images are ***z1*** and ***z2*** , given by,

$$z_2(x, y) = min(\{v : v = z_1(x, y - \frac{k}{z_1(x, y)})\})$$

$$z_1(x, y) = min(\{v : v = z_2(x, y + \frac{k}{z_2(x, y)})\})$$

These formulas allow for the occlusion of voxels, seen in one image on the surface of the object, by closer voxels seen in the other image, on the surface of the object.

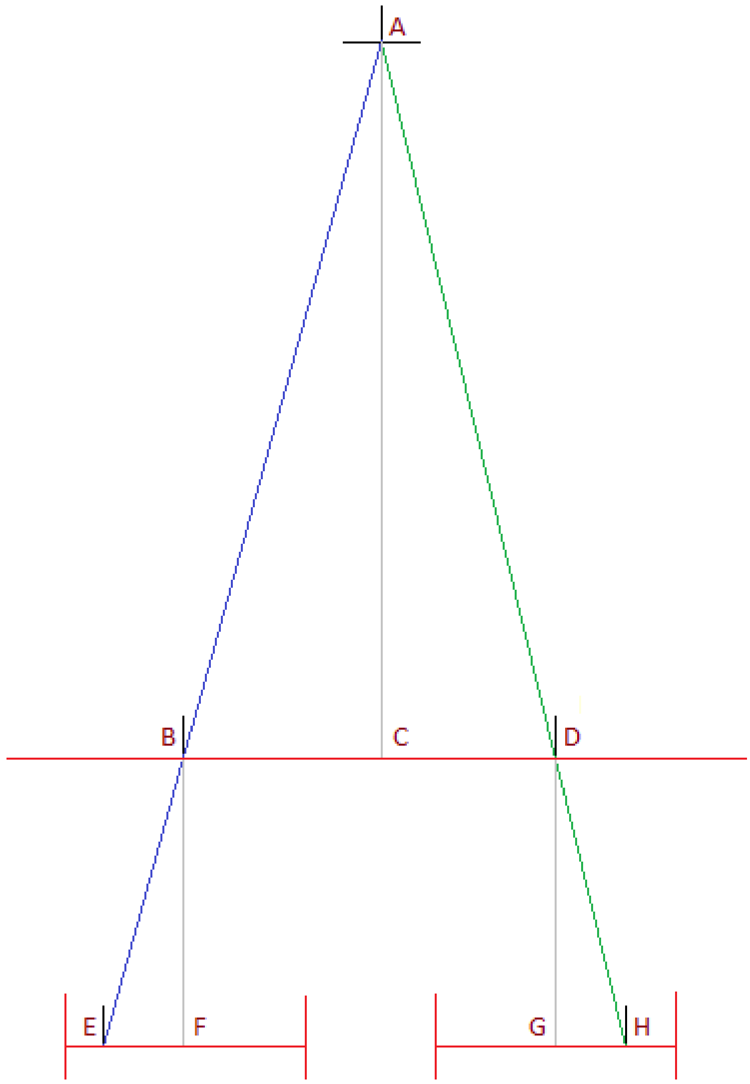

Figure 10.Diagram describing relationship of image displacement to depth with stereoscopic images, assuming flat co-planar images.

1.5.1 Image Rectification

Where the image planes are not co-planar image rectification is required to adjust the images as if they were co-planar. This may achieved by a linear transformation.

The images may also need rectification to make each image equivalent to the image taken from a pinhole camera projecting to a flat plane.

1.5.2 Least squares information measure

The normal distribution is

$$P(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Probability is related to information content described by message length *L*,

$$P(x) = 2^{-L(x)}$$
$$L(x) = -\log_2 P(x)$$

so,

$$L(x,\mu,\sigma) = \log_2(\sigma\sqrt{2\pi}) + \frac{(x-\mu)^2}{2\sigma^2}\log_2 e$$

For the purposes of comparing stereoscopic images, only the relative message length matters. Based on this, the information measure *I*, called the Sum of Squares of Differences (SSD) is,

$$I(x,\mu,\sigma) = \frac{(x-\mu)^2}{\sigma^2}$$

where,

$$L(x,\mu,\sigma) = \log_2(\sigma\sqrt{2\pi}) + I(x,\mu,\sigma)\frac{\log_2 e}{2}$$

Because of the cost in processing time of squaring numbers in SSD, many implementations use Sum of Absolute Difference (SAD) as the basis for computing the information measure. Other methods use normalized cross correlation (NCC).

1.5.3 Smoothness

Smoothness is a measure of how similar colors that are close together are. There is an assumption that objects are more likely to be colored with a small number of colors. So if we detect to 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. It is not inherently a property of an image. For example an image constructed of random dots would have no smoothness, and inferences about neighboring points would be useless.

Theoretically smoothness, along with other properties of the world should be learnt. This appears to be what the human vision system does.

1.6 Stereo Camera Sensors

There is a wide variety of 3D stereo vision sensors for simple to complex applications. A large variety of camera sensors make more difficult the choice and this is the case when before purchasing any stereo vision system has to be calculated a series of features. Some cameras are more sensitive, while others have the ability to let the user to specify the bit – rate, image quality, set the shutter speed, or average illumination in the image.

As an example, for a mobile outdoor robot is preferable to be used a wide field of view to capture a large number of objects that may be moving and at a time, these will get in range of the robot. How many frames per second is required, if the focal length is fixed or variable, how interface the sensor with electronic boards, and a minimum camera resolution are four features which must be taken into account before buying a camera.

Below, it is available a collection of most popular stereo camera sensors with different specifications and designed for different applications.

Bumblebee XB3 and Bumblebee 2

Two stereo vision cameras with complete hardware and software packages. Bumblebee 2 has a resolution of 640x480 at 48 FPS or 1024x768 at 20 FPS while XB3 provide a higher resolution of 1280x960 at 15 FPS.
They offer full field-of-view depth measurements form a single image set. Epipolar lines are aligned to within 0.05 pixels RMS error based on a stereo resolution of 320x240. Cameras easily generate one million 3D points per

second, so there is a real time transformation of images to 3D data. The images and 3D data are perfectly registered. They also offer easy integration with other machine vision techniques. No lasers or projectors and no manual adjustments or in-field calibration are required.

Figure 11. Bumblebee XP3

Kinect 3D

Coming from a virtual gaming world, the sensor Kinect is appreciated among robotics enthusiasts and is used in a wide range of projects. Since it is a 3D sensor with a friendly interface, Kinect is a powerful tool for 3D vision and provide 640x480 pixels resolution at 30 FPS.

Figure 12. Kinect 3D

Surveyvor Stereo Vision System ("SVS")

It is an open-source stereo vision system designed especially for education, research or hobbyist applications. The vision sensor comes with Omnivision OV9655 at 1.3 megapixel sensor and provides clear images in different light conditions.

Figure 13. SVS

MEGA-DCS

Compact stereo camera vision compatible with a wide range of operating systems and provide a maximum resolution of 1280x960 pixels. The system includes standard lenses and SRI Small Vision System software for real time stereo analysis, uncompressed video at megapixel resolution (7.5 fps) or VGA (30 fps), color or monochrome.

Figure 14. MEGA-DCS

PCI nDepth vision system

With a baseline of 6 centimeters, PCI nDepth is a vision system with two wide VGA CMOS digital sensors and a resolution of 752x480 pixels at 60 FPS. All output from the processor is sent via direct memory access (DMA) to the PC.

The undistorted and rectified (calibrated) images, along with the depth image, can also be sent every frame if desired. Available with an automotive quality monochrome image sensor, the stereo vision camera boasts >60dB of dynamic range and near IR enhanced performance for use with non visible near IR illumination. The stereo camera connects to the PCI card, using one standard CAT6 cable. The cable carries all necessary power, data, and control for the stereo camera.

Figure 15. PCI nDepth vision system

Ensenso N35

It is a stereo camera with compact size, USB 2.0 interface, and a 3D image sensor with 1280x1024 pixels resolution at 30 FPS. The Ensenso stereo 3D camera works according to the projected texture stereo vision principle. Each model has two integrated CMOS sensors and a projector that casts a random point pattern onto the object to be captured. The key advantage of the pattern is that it also works in multi-camera mode and can capture images of surfaces that have virtually no texture at all. The compact and robust aluminum housing of the cameras with lockable GPIO connector for trigger and flash underline the suitability of the cameras for industrial use. The cameras are pre-calibrated and come with an MVTec HALCON interface as well as an object-oriented API (C++, C# / .NET).

Figure 16. Ensenso N35

Capella

It is a complete vision system with OpenCV library support. At 30 FPS the camera can record images at 736x480 pixels resolution. Capella is an embedded pixel synchronous Stereo Vision Camera Reference Design for Texas Instruments' (TI) OMAP35x and AM/DM37x processor on Gumstix® Overo® COMs, designed and developed by e-con Systems. The Capella features the Gumstix® Overo® COM, Tobi base board and a camera daughter card e-CAM_9V024_STEREO. The e-CAM_9V024_STEREO delivers pixel-synchronous stereo frames to the OMAP/DM37x processor.

Figure 17. Capella

Minoru 3D Webcam

Perhaps it is the cheapest stereo vision camera that is working as a normal webcam and with 3D vision capabilities.

Figure 18. Minoru 3D Webcam

Scorpion 3D Stinger

Scorpion 3D Stinger for Robot Vision is designed to solve manufacturers' classic challenge: Picking parts from a conveyor belt, a pallet or a crate. Scorpion 3D Stinger captures images, identifies and locates the product and sends the id and 3D location to a picking robot.

Figure 19. Scorpion 3D Stinger

Figure 20. Stereo Camera Sensors

1.7 Single Stereo Camera System

Research on the recovery and recognition of 3-D shapes has been undertaken using a monocular image and multiple views. Depth perception by stereo disparity has been studied extensively in computer vision. The stereo disparity between two images from two distinct viewpoints is a powerful cue to 3-D shape. For the recovery of a 3-D scene from a pair of stereo image of the scene, it is required to establish correspondence [8].

A correspondence algorithm can produce more reliable matches if the underlying images have smaller intensity and geometric differences. Some geometric difference between stereo images is unavoidable, for it is actually the local geometric difference between stereo images that results in the perception of depth. For stereo images acquired by two cameras, the focal lengths and zoom levels of the cameras are often slightly different. Differences in the optical properties of the two cameras cause intensity differences between corresponding points in stereo images. These unwanted geometric and intensity differences should be reduced as much as possible to increase the ability to find correspondences reliably.

Nishimoto and Shirai [9] proposed a single-lens camera system that can obtain stereo images. Stereo images are obtained with a mirror at two different rotational positions. Teoh and Zhang [10] proposed a single-lens stereo camera system. The rotating mirror is made parallel to one of the fixed mirrors and an image is obtained. Then it is made parallel to the other fixed mirror and another image is obtained. Gosthasby and Gruver [11] proposed a single camera system that can obtain images in a single shot and through a single lens. The reversed image should be transformed to appear as if obtained by cameras with parallel optical axes, before carrying out the correspondence and measuring the depth values from the correspondence. In their recent work, Nene and Nayar [12] proposed four stereo systems that use a single camera pointed towards planar, ellipsoidal, hyperboloidal, and paraboloidal mirrors. By use of non-planar reflecting surfaces such as hyperboloids and paraboloids, a wide field of view is easily obtained. However, their stereo system needs a complex mirror mechanism. Doo Hyun Lee, In So Kweon, Roberto Cipolla [13] proposed a novel and practical stereo camera system that uses only one camera and a biprism placed in front of the camera. The equivalent of a stereo pair of images is formed as the left and right halves of a single CCD image using a biprism.

These systems are therefore cheap and extremely easy to calibrate since they require only one CCD camera.

1.8 Stereo Vision: Applications

3D stereo displays find many applications in entertainment, information transfer and automated systems. Stereo vision is highly important in fields such as robotics, to extract information about the relative position of 3D objects in the vicinity of autonomous systems. Other applications for robotics include object recognition, where depth information allows for the system to separate occluding image components, such as one chair in front of another, which the robot may otherwise not be able to distinguish as a separate object by any other criteria.

Scientific applications for digital stereo vision include the extraction of information from aerial surveys, for calculation of contour maps or even geometry extraction for 3D building mapping, or calculation of 3D heliographical information such as obtained by the NASA STEREO project.

Chapter 2: Stereoscopy

Stereoscopy (also called stereoscopics or 3D imaging) is a technique for creating or enhancing the illusion of depth in an image by means of stereopsis for binocular vision. Any stereoscopic image is called stereogram. Originally, stereogram referred to a pair of stereo images which could be viewed using a stereoscope.

Most stereoscopic methods present two offset images separately to the left and right eye of the viewer. These two-dimensional images are then combined in the brain to give the perception of 3D depth. This technique is distinguished from 3D displays that display an image in three full dimensions, allowing the observer to increase information about the 3-dimensional objects being displayed by head and eye movements.

Figure 21. Pocket stereoscope with original test image.

2.1 Stereoscopy: Background

Stereoscopy creates the illusion of three-dimensional depth from given two-dimensional images. Human vision, including the perception of depth, is a complex process which only begins with the acquisition of visual information taken in through the eyes; much processing ensues within the brain, as it strives to make intelligent and meaningful sense of the raw information provided. One of the very important visual functions that occur within the brain as it interprets what the eyes see is that of assessing the relative distances of various objects from the viewer, and the depth dimension of those same perceived objects. The brain makes use of a number of *cues* to determine relative distances and depth in a perceived scene, including:

- Stereopsis
- Accommodation of the eye
- Overlapping of one object by another
- Subtended visual angle of an object of known size
- Linear perspective (convergence of parallel edges)
- Vertical position (objects higher in the scene generally tend to be perceived as further away)
- Haze, desaturation, and a shift to bluishness
- Change in size of textured pattern detail

(All the above cues, with the exception of the first two, are present in traditional two-dimensional images such as paintings, photographs, and television.)

Stereoscopy is the production of the illusion of depth in a photograph, movie, or other two-dimensional image by presenting a slightly different image to each eye, and thereby adding the first of these cues (stereopsis) as well. Both of the 2D offset images are then combined in the brain to give the perception of 3D depth. It is important to note that since all points in the image focus at the same plane regardless of their depth in the original scene, the second cue, focus, is still not duplicated and therefore the illusion of depth is incomplete. There are also primarily two effects of stereoscopy that are unnatural for the human vision: first, the mismatch between convergence and accommodation, caused by the difference between an object's perceived position in front of or behind the display or screen and the real origin of that light and second, possible crosstalk between the eyes, caused by imperfect image separation by some methods.

Although the term "3D" is ubiquitously used, it is also important to note that the presentation of dual 2D images is distinctly different from displaying an image in three full dimensions. The most notable difference is that, in the case of "3D" displays, the observer's head and eye movement will not increase information about the 3-dimensional objects being displayed. Holographic displays or volumetric display are examples of displays that do not have this limitation. Similar to the technology of sound reproduction, in which it is not possible to recreate a full 3-dimensional sound field merely with two stereophonic speakers, it is likewise an overstatement of capability to refer to dual 2D images as being "3D". The accurate term "stereoscopic" is more cumbersome than the common misnomer "3D", which has been entrenched after many decades of unquestioned misuse. Although most stereoscopic displays do not qualify as real 3D display, all real 3D displays are also stereoscopic displays because they meet the lower criteria as well.

Most 3D displays use this stereoscopic method to convey images.

Stereoscopy is used in photogrammetry and also for entertainment through the production of stereograms. Stereoscopy is useful in viewing images rendered from large multi-dimensional data sets such as are produced by experimental data. Modern industrial three-dimensional photography may use 3D scanners to detect and record three-dimensional information. The three-dimensional depth information can be reconstructed from two images using a computer by corresponding the pixels in the left and right images (e.g.,). Solving the Correspondence problem in the field of Computer Vision aims to create meaningful depth information from two images.

2.2 Visual requirements

Anatomically, there are 3 levels of binocular vision required to view stereo images:

1. Simultaneous perception
2. Fusion (binocular 'single' vision)
3. Stereopsis

These functions develop in early childhood. Some people who have strabismus disrupt the development of stereopsis, however orthoptics treatment can be used to improve binocular vision. A person's stereoacuity determines the minimum image disparity they can perceive as depth. It is believed that approximately 12% of people are unable to properly see 3D images, due to a variety of medical conditions. According to another experiment up to 30% of people have very weak stereoscopic vision preventing them from depth perception based on stereo disparity. This nullifies or greatly decreases immersion effects of stereo to them.

2.3 Side-by-Side

Traditional stereoscopic photography consists of creating a 3D illusion starting from a pair of 2D images, a stereogram. The easiest way to enhance depth perception in the brain is to provide the eyes of the viewer with two different images, representing two perspectives of the same object, with a minor deviation equal or nearly equal to the perspectives that both eyes naturally receive in binocular vision.

To avoid eyestrain and distortion, each of the two 2D images should be presented to the viewer so that any object at infinite distance is perceived by the eye as being straight ahead, the viewer's eyes being neither crossed nor diverging. When the picture contains no object at infinite distance, such as a horizon or a cloud, the pictures should be spaced correspondingly closer together.

The principal advantages of side-by-side viewers is the lack of diminution of brightness, allowing the presentation of images at very high resolution and in full spectrum color, simplicity in creation, and little or no additional image processing is required. Under some circumstances, such as when a pair of images are presented for freeviewing, no device or additional optical equipment is needed.

The principal disadvantage of side-by-side viewers is that large image displays are not practical and resolution is limited by the lesser of the display medium or human eye. This is because as the dimensions of an image are increased, either the viewing apparatus or viewer themselves must move proportionately further away from it in order to view it comfortably. Moving closer to an image in order to see more detail would only be possible with viewing equipment that adjusted to the difference.

2.3.1 Freeviewing

Freeviewing is viewing a side-by-side image pair without using a viewing device.

Two methods are available to free view:

• The parallel viewing method uses an image pair with the left-eye image on the left and the right-eye image on the right. The fused three-dimensional image appears larger and more distant than the two actual images, making it possible to convincingly simulate a life-size scene. The viewer attempts to look through the images with the eyes substantially parallel, as if looking at the actual scene. This can be difficult with normal vision because eye focus and binocular convergence are habitually coordinated. One approach to decoupling the two functions is to view the image pair extremely close up with completely relaxed eyes, making no attempt to focus clearly but simply achieving comfortable stereoscopic fusion of the two blurry images by the "look-through" approach, and only then exerting the effort to focus them more clearly, increasing the viewing distance as necessary. Regardless of the approach used or the image medium, for comfortable viewing and stereoscopic accuracy the size and spacing of the images should be such that the corresponding points of very distant objects in the scene are separated by the same distance as the viewer's eyes, but not more; the average interocular distance is about 63 mm. Viewing much more widely separated images is possible, but because the eyes never diverge in normal use it usually requires some previous training and tends to cause eye strain.

• The cross-eyed viewing method swaps the left and right eye images so that they will be correctly seen cross-eyed, the left eye viewing the image on the right and vice-versa. The fused three-dimensional image appears to be smaller and closer than the actual images, so that large objects and scenes appear miniaturized. This method is usually easier for freeviewing novices. As an aid to fusion, a fingertip can be placed just below the division between the two images, then slowly brought straight toward the viewer's eyes, keeping the eyes directed at the fingertip; at a certain distance, a fused three-dimensional image should seem to be hovering just above the finger. Alternatively, a piece of paper with a small opening cut into it can be used in a similar manner; when correctly positioned between the image pair and the viewer's eyes, it will seem to frame a small three-dimensional image.

Prismatic, self-masking glasses are now being used by some cross-eyed-view advocates. These reduce the degree of convergence required and allow large images to be displayed. However, any viewing aid that uses prisms, mirrors or lenses to assist fusion or focus is simply a type of stereoscope, excluded by the customary definition of freeviewing.

Stereoscopically fusing two separate images without the aid of mirrors or prisms, while simultaneously keeping them in sharp focus without the aid of suitable viewing lenses, inevitably requires an unnatural combination of eye vergence and accommodation. Simple freeviewing therefore cannot accurately reproduce the physiological depth cues of the real-world viewing experience. Different individuals may experience differing degrees of ease and comfort in achieving fusion and good focus, as well as differing tendencies to eye fatigue or strain.

2.3.2 Autostereogram

An autostereogram is a single-image stereogram (SIS), designed to create the visual illusion of a three-dimensional (3d) scene within the human brain from an external two-dimensional image. In order to perceive 3D shapes in these autostereograms, one must overcome the normally automatic coordination between focusing and vergence.

2.3.3 Stereoscope and stereographic cards

The stereoscope is essentially an instrument in which two photographs of the same object, taken from slightly different angles, are simultaneously presented, one to each eye. A simple stereoscope is limited in the size of the image that may be used. A more complex stereoscope uses a pair of horizontal periscope-like devices, allowing the use of larger images that can present more detailed information in a wider field of view.

2.3.4 Transparency Viewers

Some stereoscopes are designed for viewing transparent photographs on film or glass, known as transparencies or diapositives and commonly called slides. Some of the earliest stereoscope views, issued in the 1850s, were on glass. In the early 20th century, 45x107 mm and 6x13 cm glass slides were common formats for amateur stereo photography, especially in Europe. In later years, several film-based formats were in use. The best-known formats for commercially issued stereo views on film are Tru-Vue, introduced in 1931, and View-Master, introduced in 1939 and still in production. For amateur stereo slides, the Stereo Realist format, introduced in 1947, is by far the most common.

2.3.5 Head-mounted displays

The user typically wears a helmet or glasses with two small LCD or OLED displays with magnifying lenses, one for each eye. The technology can be used to show stereo films, images or games, but it can also be used to create a virtual display. Head-mounted displays may also be coupled with head-tracking devices, allowing the user to "look around" the virtual world by moving their head, eliminating the need for a separate controller. Performing this update quickly enough to avoid inducing nausea in the user requires a great amount of computer image processing. If six axis position sensing (direction and position) is used then wearer may move about within the limitations of the equipment used. Owing to rapid advancements in computer graphics and the continuing miniaturization of video and other equipment these devices are beginning to become available at more reasonable cost.

Figure 22. A HMD with a separate video source displayed in front of each eye to achieve a stereoscopic effect

Head-mounted or wearable glasses may be used to view a see-through image imposed upon the real world view, creating what is called augmented reality. This is done by reflecting the video images through partially reflective mirrors. The real world view is seen through the mirrors' reflective surface.

Experimental systems have been used for gaming, where virtual opponents may peek from real windows as a player moves about. This type of system is expected to have wide application in the maintenance of complex systems, as it can give a technician what is effectively "x-ray vision" by combining computer graphics rendering of hidden elements with the technician's natural vision. Additionally, technical data and schematic diagrams may be delivered to this same equipment, eliminating the need to obtain and carry bulky paper documents.

Augmented stereoscopic vision is also expected to have applications in surgery, as it allows the combination of radiographic data (CAT scans and MRI imaging) with the surgeon's vision.

2.3.6 Virtual retinal displays

A virtual retinal display (VRD), also known as a retinal scan display (RSD) or retinal projector (RP), not to be confused with a "Retina Display", is a display technology that draws a raster image (like a television picture) directly onto the retina of the eye. The user sees what appears to be a conventional display floating in space in front of them. For true stereoscopy, each eye must be provided with its own discrete display. To produce a virtual display that occupies a usefully large visual angle but does not involve the use of relatively large lenses or mirrors, the light source must be very close to the eye. A contact lens incorporating one or more semiconductor light sources is the form most commonly proposed. As of 2013, the inclusion of suitable light-beam-scanning means in a contact lens is still very problematic, as is the alternative of embedding a reasonably transparent array of hundreds of thousands (or millions, for HD resolution) of accurately aligned sources of collimated light.

Chapter 3: 3D viewers

There are two categories of 3D viewer technology, active and passive. Active viewers have electronics which interact with a display. Passive viewers filter constant streams of binocular input to the appropriate eye.

3.1 Active

3.1.1 Active shutter 3D system

An active shutter 3D system (a.k.a. alternate frame sequencing, alternate image, AI, alternating field, field sequential or eclipse method) is a technique of displaying stereoscopic 3D images. It works by only presenting the image intended for the left eye while blocking the right eye's view, then presenting the right-eye image while blocking the left eye, and repeating this so rapidly that the interruptions do not interfere with the perceived fusion of the two images into a single 3D image.

Modern active shutter 3D systems generally use liquid crystal shutter glasses (also called "LC shutter glasses" or "active shutter glasses"). Each eye's glass contains a liquid crystal layer which has the property of becoming opaque when voltage is applied, being otherwise transparent. The glasses are controlled by a timing signal that allows the glasses to alternately block one eye, and then the other, in synchronization with the refresh rate of the screen. The timing synchronization to the video equipment may be achieved via a wired signal, or wirelessly by either an infrared or radio frequency (e.g. Bluetooth, DLP link) transmitter.

Active shutter 3D systems are used to present 3D films in some theaters, and they can be used to present 3D images on CRT, plasma, LCD, projectors and other types of video displays.


Figure 23. A pair of CrystalEyes shutter glasses

3.1.2 Advantages

• Unlike red/cyan color filter (anaglyph) 3D glasses, LC shutter glasses are color neutral, enabling 3D viewing in the full color spectrum.
• Unlike in a Polarized 3D system, where the screen resolution is halved when the images are combined, the active shutter system retains full resolution by combining the images over time.

3.1.3 Disadvantages

• Flicker can be noticed except at very high refresh rates, as each eye is effectively receiving only half of the monitor's actual refresh rate. However, modern LC glasses generally work in higher refresh rates and eliminate this problem for most people.
• Until recently, the method only worked with CRT monitors; some modern flat-panel monitors now support high-enough refresh rates to work with some LC shutter systems. Many projectors, especially DLP-based ones, support 3D out of the box.
• LC shutter glasses are shutting out light half of the time; moreover, they are slightly dark even when letting light through, because they are polarized. This gives an effect similar to watching TV with sunglasses on, which causes a darker picture to be perceived by the viewer. However, this effect can produce a higher perceived display contrast when paired with LCD displays

because of the reduction in backlight bleed. Since the glasses also darken the background, contrast is enhanced when using a brighter image.

• When used with LCD displays, extreme localized differences between the image to be displayed in one eye and the other may lead to crosstalk, due to LCD panels' pixels sometimes being unable to fully switch, for example from black to white, in the time that separates the left eye's image from the right one. Recent advancements in the panel's response time, however, has led to displays that rival or even surpass passive 3D systems.

• Frame rate has to be double that of a non-3D, anaglyph, or polarized 3D systems to get an equivalent result. All equipment in the chain has to be able to process frames at double rate; in essence this doubles the hardware requirements.

• Despite a progressive fall in prices, and due to the intrinsic use of electronics, they remain more expensive than anaglyph and polarized 3D glasses.

• Because of their integrated electronics and batteries, early shutter glasses were heavy and expensive. However, design improvements have resulted in newer models that are cheaper, lightweight, rechargeable and able to be worn over prescription lenses.

• From brand to brand, shutter glasses use different synchronization methods and protocols. Therefore, even glasses that use the same kind of synchronization system (e.g. infrared) will probably be incompatible across different makers. However, efforts are being made to create a Universal 3D Shutter Glass.

• Some active shutter 3D systems may exhibit depth distortion of screen objects moving horizontally, due to a timing mismatch between the capture (recording) and playback of sequential left-right frames. An object moving laterally in one direction may protrude while in the other direction it may recede into the screen. A static image does not have this distortion.

3.2 Passive

3.2.1 Polarized 3D system

To present stereoscopic pictures, two images are projected superimposed onto the same screen through polarizing filters or presented on a display with polarized filters. For projection, a silver screen is used so that polarization is preserved. The viewer wears low-cost eyeglasses which also contain a pair of opposite polarizing filters. As each filter only passes light which is similarly polarized and blocks the opposite polarized light, each eye only sees one of the images, and the effect is achieved.

3.2.2 Interference filter systems

This technique uses specific wavelengths of red, green, and blue for the right eye, and different wavelengths of red, green, and blue for the left eye. Eyeglasses which filter out the very specific wavelengths allow the wearer to see a full color 3D image. It is also known as spectral comb filtering or wavelength multiplex visualization or super-anaglyph. Dolby 3D uses this principle. The Omega 3D/Panavision 3D system has also used an improved version of this technology. In June 2012 the Omega 3D/Panavision 3D system was discontinued by DPVO Theatrical, who marketed it on behalf of Panavision, citing "challenging global economic and 3D market conditions". Although DPVO dissolved its business operations, Omega Optical continues promoting and selling 3D systems to non-theatrical markets. Omega Optical's 3D system contains projection filters and 3D glasses. In addition to the passive stereoscopic 3D system, Omega Optical has produced enhanced anaglyph 3D glasses. The Omega's red/cyan anaglyph glasses use complex metal oxide thin film coatings and high quality annealed glass optics.

3.2.3 Anaglyph 3D

Anaglyph 3D is the name given to the stereoscopic 3D effect achieved by means of encoding each eye's image using filters of different (usually chromatically opposite) colors, typically red and cyan. Anaglyph 3D images contain two differently filtered colored images, one for each eye. When viewed through the "color-coded" "anaglyph glasses", each of the two images reaches the eye it's intended for, revealing an integrated stereoscopic image. The visual cortex of the brain fuses this into perception of a three-dimensional scene or composition. Anaglyph images have seen a recent resurgence due to the presentation of images and video on the Web, Blu-ray Discs, CDs, and even in print. Low cost paper frames or plastic-framed glasses hold accurate color filters that typically, after 2002, make use of all 3 primary colors. The current norm is red and cyan, with red being used for the left channel. The cheaper filter material used in the monochromatic past dictated red and blue for convenience and cost. There is a material improvement of full color images, with the cyan filter, especially for accurate skin tones.

Video games, theatrical films, and DVDs can be shown in the anaglyph 3D process. Practical images, for science or design, where depth perception is useful, include the presentation of full scale and microscopic stereographic images. Examples from NASA include Mars Rover imaging, and the solar investigation, called STEREO, which uses two orbital vehicles to obtain the 3D images of the sun. Other applications include geological illustrations by the United States Geological Survey, and various online museum objects. A recent application is for stereo imaging of the heart using 3D ultra-sound with plastic red/cyan glasses.

Anaglyph images are much easier to view than either parallel (diverging) or crossed-view pairs stereograms. However, these side-by-side types offer bright and accurate color rendering, not easily achieved with anaglyphs. Recently, cross-view prismatic glasses with adjustable masking have appeared, that offer a wider image on the new HD video and computer monitors.

3.2.3.1 Mechanics

Viewing anaglyphs through spectrally opposed glasses or gel filters enables each eye to see independent left and right images from within a single anaglyphic image. In a red-cyan anaglyph, the eye viewing through the red filter sees red within the anaglyph as "white", and the cyan within the anaglyph as "black". The eye viewing through the cyan filter perceives the opposite. Actual black or white in the anaglyph display, being void of color, are perceived the same by each eye. The brain blends together the red and cyan channeled images as in regular viewing but only green and blue are perceived. Red is not perceived because red equates with white through red gel and is black through cyan gel. However green and blue are perceived through cyan gel.

3.2.3.2 Types

Complementary color anaglyphs employ one of a pair of complementary color filters for each eye. The most common color filters used are red and cyan. Employing tristimulus theory, the eye is sensitive to three primary colors, red, green, and blue. The red filter admits only red, while the cyan filter blocks red, passing blue and green (the combination of blue and green is perceived as cyan). If a paper viewer containing red and cyan filters is folded so that light passes through both, the image will appear black. Another recently introduced form employs blue and yellow filters. (Yellow is the color perceived when both red and green light passes through the filter.)

Anaglyph images have seen a recent resurgence because of the presentation of images on the Internet. Where traditionally, this has been a largely black & white format, recent digital camera and processing advances have brought very acceptable color images to the internet and DVD field. With the online availability of low cost paper glasses with improved red-cyan filters, and plastic framed glasses of increasing quality, the field of 3D imaging is growing quickly. Scientific images where depth perception is useful include, for instance, the presentation of complex multi-dimensional data sets and stereographic images of the surface of Mars. With the recent release of 3D DVDs, they are more commonly being used for entertainment. Anaglyph images are much easier to view than either parallel sighting or crossed eye stereograms, although these types do offer more bright and accurate color rendering, most particularly in the red component, which is commonly muted or desaturated with even the best color anaglyphs. A compensating technique, commonly known as Anachrome, uses a slightly more transparent cyan filter in the patented glasses associated with the technique. Processing reconfigures the typical anaglyph image to have less parallax to obtain a more useful image when viewed without filters.

3.2.3.3 Viewing

A pair of glasses with filters of opposing colors is worn to view an anaglyphic photo image. A red filter lens over the left eye allows graduations of red to cyan from within the anaglyph to be perceived as graduations of bright to dark. The cyan (blue/green) filter over the right eye conversely allows graduations of cyan to red from within the anaglyph to be perceived as graduations of bright to dark. Red and cyan color fringes in the anaglyph display represent the red and cyan color channels of the parallax displaced left and right images. The viewing filters each cancel out opposing colored areas, including graduations of less pure opposing colored areas, to each reveal an image from within its color channel. Thus the filters enable each eye to see only its intended view from color channels within the single anaglyphic image.

Red sharpened anaglyph glasses

Simple paper, uncorrected gel glasses, cannot compensate for the 250 nanometer difference in the wavelengths of the red-cyan filters. With simple glasses, the red filtered image is somewhat blurry, when viewing a close computer screen or printed image. The (RED) retinal focus differs from the image through the (CYAN) filter, which dominates the eyes' focusing. Better quality, molded acrylic glasses frequently employ a compensating differential diopter power (a spherical correction) to balance the red filter focus shift relative to the cyan, which reduces the innate softness, and diffraction of red filtered light. Low power reading glasses worn along with the paper glasses also sharpen the image noticeably.

Anachrome filters

Plastic glasses, developed in recent years, provide both the diopter "fix" noted above, and a change in the cyan filter. The formula provides intentional "leakage" of a minimal (2%) percentage of red light with the conventional range of the filter. This assigns two-eyed "redness cues" to objects and details, such as lip color and red clothing, that are fused in the brain. Care must be taken, however, to closely overlay the red areas into near-perfect registration, or "ghosting" can occur. Anachrome formula lenses work well with black and white, but can provide excellent results when the glasses are used with conforming, "Anachrome friendly" images. By convention, anachrome images try to avoid excess separation of the cameras, and parallax, thereby reducing the ghosting that the extra color bandwidth introduces to the images.

3.2.3.4 Anaglyphic color channels

Anaglyph images may use any combination of color channels. However if a stereoscopic image is to be pursued, the colors should be diametrically opposed. Impurities of color channel display, or of the viewing filters, allow some of the image meant for the other channel to be seen. This results in stereoscopic double imaging, also called ghosting. Color channels may be left-right reversed. Red/Cyan is most common. Magenta/Green and Blue/Yellow are also popular. Red/Green and Red/Blue enable monochromatic images especially Red/Green. Many anaglyph makers purposely integrate impure color channels and viewing filters to enable better color perception, but this results in a corresponding degree of double imaging. Color Channel Brightness % of White: Red-30/Cyan-70, Magenta-41/Green-59 or especially Blue- 11/Yellow-89), the lighter display channel may be darkened or the brighter viewing filter may be darkened to allow both eyes a balanced view. However the Pulfrich effect can be obtained from a light/dark filter arrangement. The color channels of an anaglyphic image require pure color display fidelity and corresponding viewing filter gels. The choice of ideal viewing filters is dictated by the color channels of the anaglyph to be viewed. Ghosting can be eliminated by ensuring a pure color display and viewing filters that mach the display. Retinal rivalry can be eliminated by the (ACB) 3-D 'Anaglyphic Contrast Balance' method patented by that prepares the image pair prior to color channelling in any color.

3.2.4 Chromadepth system

The ChromaDepth procedure of American Paper Optics is based on the fact that with a prism, colors are separated by varying degrees. The ChromaDepth eyeglasses contain special view foils, which consist of microscopically small prisms. This causes the image to be translated a certain amount that depends on its color. If one uses a prism foil now with one eye but not on the other eye, then the two seen pictures – depending upon color – are more or less widely separated. The brain produces the spatial impression from this difference. The advantage of this technology consists above all of the fact that one can regard ChromaDepth pictures also without eyeglasses (thus two-dimensional) problem-free (unlike with two-color anaglyph). However the colors are only limitedly selectable, since they contain the depth information of the picture. If one changes the color of an object, then its observed distance will also be changed.

3.2.5 Pulfrich method

The Pulfrich effect is based on the phenomenon of the human eye processing images more slowly when there is less light, as when looking through a dark lens. Because the Pulfrich effect depends on motion in a particular direction to instigate the illusion of depth, it is not useful as a general stereoscopic technique. For example, it cannot be used to show a stationary object apparently extending into or out of the screen; similarly, objects moving vertically will not be seen as moving in depth. Incidental movement of objects will create spurious artifacts, and these incidental effects will be seen as artificial depth not related to actual depth in the scene.

3.2.6 Over/under format

Stereoscopic viewing is achieved by placing an image pair one above one another. Special viewers are made for over/under format that tilt the right eyesight slightly up and the left eyesight slightly down. The most common one with mirrors is the View Magic. Another with prismatic glasses is the KMQ viewer. A recent usage of this technique is the openKMQ project.

3.3 Other display methods without viewers

3.3.1 Autostereoscopy

Autostereoscopic display technologies use optical components in the display, rather than worn by the user, to enable each eye to see a different image. Because headgear is not required, it is also called "glasses-free 3D". The optics split the images directionally into the viewer's eyes, so the display viewing geometry requires limited head positions that will achieve the stereoscopic effect. Automultiscopic displays provide multiple views of the same scene, rather than just two. Each view is visible from a different range of positions in front of the display. This allows the viewer to move left-right in front of the display and see the correct view from any position. The technology includes two broad classes of displays: those that use head-tracking to ensure that each of the viewer's two eyes sees a different image on the screen, and those that display multiple views so that the display does not need to know where the viewers' eyes are directed. Examples of autostereoscopic displays technology include lenticular lens, parallax barrier, volumetric display, holography and light field displays.

3.3.2 Holography

Laser holography, in its original "pure" form of the photographic transmission hologram, is the only technology yet created which can reproduce an object or scene with such complete realism that the reproduction is visually indistinguishable from the original, given the original lighting conditions. It creates a light field identical to that which emanated from the original scene, with parallax about all axes and a very wide viewing angle. The eye differentially focuses objects at different distances and subject detail is preserved down to the microscopic level. The effect is exactly like looking through a window. Unfortunately, this "pure" form requires the subject to be laser-lit and completely motionless—to within a minor fraction of the wavelength of light—during the photographic exposure, and laser light must be used to properly view the results. Most people have never seen a laser-lit transmission hologram. The types of holograms commonly encountered have seriously compromised image quality so that ordinary white light can be used for viewing, and non-holographic intermediate imaging processes are almost always resorted to, as an alternative to using powerful and hazardous pulsed lasers, when living subjects are photographed.

Although the original photographic processes have proven impractical for general use, the combination of computer-generated holograms (CGH) and optoelectronic holographic displays, both under development for many years, has the potential to transform the half-century-old pipe dream of holographic 3D television into a reality; so far, however, the large amount of calculation required to generate just one detailed hologram, and the huge bandwidth required to transmit a stream of them, have confined this technology to the research laboratory.

3.3.3 Volumetric displays

Volumetric displays use some physical mechanism to display points of light within a volume. Such displays use voxels instead of pixels. Volumetric displays include multiplanar displays, which have multiple display planes stacked up, and rotating panel displays, where a rotating panel sweeps out a volume.

Other technologies have been developed to project light dots in the air above a device. An infrared laser is focused on the destination in space, generating a small bubble of plasma which emits visible light.

3.3.4 Integral imaging

Integral imaging is an autostereoscopic or multiscopic 3D display, meaning that it displays a 3D image without the use of special glasses on the part of the viewer. It achieves this by placing an array of microlenses (similar to a lenticular lens) in front of the image, where each lens looks different depending on viewing angle. Thus rather than displaying a 2D image that looks the same from every direction, it reproduces a 4D light field, creating stereo images that exhibit parallax when the viewer moves.

3.3.5 Wiggle stereography

Wiggle stereoscopy is an image display technique achieved by quickly alternating display of left and right sides of a stereogram. Found in animated GIF format on the web. Online examples are visible in the New-York Public Library stereogram collection.

Chapter 4: Steps towards the implementation of the project

4.1 General description of the purpose and the background of the project

Medical companies have taken inspiration from the film industry to develop systems for 3D surgery that are now being used in several types of operation. On 2009, James Cameron's film Avatar was released onto UK cinema screens. It was a film made specifically for 3D viewing, and its technology was considered as ground breaking and the future of the cinema. The film's secret lay in the use of multiple camera angles during filming. The glasses worn allowed one slightly different view to be projected into each eye, creating the sense of depth. Medical companies taking cue from applications in the film industry began experimenting to transfer the 3D technology into the realm of health.

Although the technology to screen patients' anatomy has been around for several years, the problem has always been on the projection of the stereoscopic image onto the screen. In the early experiences of 3D stereoscopic projection, poor image quality and imaging devices, rendered early 3D stereoscopic experimentation somewhat nauseous for the surgeon. The advent of LED screens, the development of CCDs sensors, and next-gen 3D glasses overcame the aforementioned obstacles, thus rendering surgeries capable to be viewed stereoscopically.

3D surgery reduces error and time taken to do tasks as well as increases the precision of the task. 3D systems, which are capable of presenting three dimensional images of internal organs during surgery, contribute to more accurate diagnosis. In addition, 3D surgery gives greater depth perception of the tissue being operated on, something that was lacking in the flat images provided by previous 2D imaging systems. This supports faster and more accurate surgery. The 3D images help make surgery faster and more precise by giving a realistic view of the tissue being operated on and providing the viewer with the depth perception that is lacking from conventional 2D image. Meanwhile, the high resolution CCD image sensors allow close observation of fine detail in the form of high definition 3D image.

4.2 A Taxonomy of Stereo Imaging Algorithms

The taxonomy of stereo algorithms is based on the observation that stereo algorithms generally perform (subsets of) the following four steps [14, 15]:

1. matching cost computation
2. cost (support) aggregation
3. disparity computation / optimization, and
4. disparity refinement

The actual sequence of steps taken depends on the specific algorithm. For example, local (window-based) algorithms, where the disparity computation at a given point depends only on intensity values within a finite window, usually make implicit smoothness assumptions by aggregating support. Some of these algorithms can cleanly be broken down into steps 1, 2, 3. For example, the traditional sum-of-squared-differences (SSD) algorithm can be described as:

1. the matching cost is the squared difference of intensity values at a given disparity
2. aggregation is done by summing cost over square windows with constant disparity
3. disparities are computed by selecting the minimal (winning) aggregated value at each pixel.

Some local algorithms, however, combine steps 1 and 2 and use a matching cost that is based on a support region, e.g. normalized cross-correlation and the rank transform. (This can also be viewed as a preprocessing step.)

On the other hand, global algorithms make explicit smoothness assumptions and then solve an optimization problem. Such algorithms typically do not perform an aggregation step, but rather seek a disparity assignment (step 3) that minimizes a global cost function that combines data (step 1) and smoothness terms. The main distinction between these algorithms is the minimization procedure used, e.g., simulated annealing, probabilistic (mean-field) diffusion, or graph cuts [16].

In between these two broad classes are certain iterative algorithms that do not explicitly state a global function that is to be minimized, but whose behavior mimics closely that of iterative optimization algorithms [17]. Hierarchical (coarse-to-fine) algorithms resemble such iterative algorithms, but typically operate on an image pyramid, where results from coarser levels are used to constrain a more local search at finer levels.

4.2.1 Matching Cost Computation

The most common pixel-based matching costs include squared intensity differences (SD) and absolute intensity differences (AD). In the video processing community, these matching criteria are referred to as the mean-squared error (MSE) and mean absolute difference (MAD) measures; the term displaced frame difference is also often used.

More recently, robust measures, including truncated quadratics and contaminated Gaussians have been proposed [15]. These measures are useful because they limit the influence of mismatches during aggregation.

Other traditional matching costs include normalized cross-correlation, which behaves similar to sum-of-squared-differences (SSD), and binary matching costs (i.e., match / no match), based on binary features such as edges or the sign of the Laplacian. Binary matching costs are not commonly used in dense stereo methods, however.

Some costs are insensitive to differences in camera gain or bias, for example gradient-based measures and non-parametric measures such as rank and census transforms. Of course, it is also possible to correct for different camera characteristics by performing a preprocessing step for bias-gain or histogram equalization. Other matching criteria include phase and filter-bank responses. Finally, Birchfield and Tomasi have proposed a matching cost that is insensitive to image sampling [18]. Rather than just comparing pixel values shifted by integral amounts (which may miss a valid match), they compare each pixel in the reference image against a linearly interpolated function of the other image.

The matching cost values over all pixels and all disparities form the initial disparity space image (DSI) $C0(x, y, d)$. The initial DSI can easily incorporate information from more than two images by simply summing up the cost values for each matching image m, since the DSI is associated with a fixed reference image r. This is the idea behind multiple-baseline SSSD and SSAD method. This idea can be generalized to arbitrary camera configurations using a plane sweep algorithm.

4.2.2 Aggregation of Cost

Local and window-based methods aggregate the matching cost by summing or averaging over a support region in the DSI $C(x, y, d)$. A support region can be either two-dimensional at a fixed disparity (favoring fronto-parallel surfaces), or three-dimensional in x-y-d space (supporting slanted surfaces). Two-dimensional evidence aggregation has been implemented using square windows or Gaussian convolution (traditional), multiple windows anchored at different points, i.e., shiftable windows, windows with adaptive sizes, and windows based on connected components of constant disparity. Three-dimensional support functions that have been proposed include limited disparity difference, limited disparity gradient, and Prazdny's coherence principle.

Aggregation with a fixed support region can be performed using 2D or 3D convolution,

$$C(x,y,d) = w(x,y,d) * C_0(x,y,d),$$

or, in the case of rectangular windows, using efficient (moving average) box-filters. Shiftable windows can also be implemented efficiently using a separable sliding min-filter. A different method of aggregation is iterative diffusion, i.e., an aggregation (or averaging) operation that is implemented by repeatedly adding to each pixel's cost the weighted values of its neighboring pixels' costs.

4.2.3 Disparity computation and optimization

Local Methods. In local methods, the emphasis is on the matching cost computation and on the cost aggregation steps. Computing the final disparities is trivial: simply choose at each pixel the disparity associated with the minimum cost value. Thus, these methods perform a local "winner-take-all" (WTA) optimization at each pixel. A limitation of this approach (and many other correspondence algorithms) is that uniqueness of matches is only enforced for one image (the reference image), while points in the other image might get matched to multiple points.

Global Optimization. In contrast, global methods perform almost all of their work during the disparity computation phase and often skip the aggregation step. Many global methods are formulated in an energy-minimization framework. The objective is to find a disparity function d that minimizes a global energy,

$$E(d) = Edata(d) + \lambda Esmooth(d).$$

The data term, Edata(d), measures how well the disparity function d agrees with the input image pair. Using the disparity space formulation,

$$Edata(d) = \sum_{(x,y)} C(x, y, d(x, y)),$$

where C is the (initial or aggregated) matching cost DSI.

The smoothness term Esmooth(d) encodes the smoothness assumptions made by the algorithm. To make the optimization computationally tractable, the smoothness term is often restricted to only measuring the differences between neighboring pixels' disparities,

$$Esmooth(d) = \sum_{(x,y)} \rho\big(d(x,y) - d(x+1,y)\big) + \rho(d(x,y) - d(x,y+1)),$$

where ρ is some monotonically increasing function of disparity difference. (An alternative to smoothness functional is to use a lower-dimensional representation such as splines.)

The terms in Esmooth can also be made to depend on the intensity differences, e.g.,

$$\rho_d(d(x, y) - d(x+1, y)) \cdot \rho_I (\| I(x, y) - I(x+1, y) \|),$$

where ρ_I is some monotonically decreasing function of intensity differences that lowers smoothness costs at high intensity gradients. This idea encourages disparity discontinuities to coincide with intensity/color edges and appears to account for some of the good performance of global optimization approaches.

Once the global energy has been defined, a variety of algorithms can be used to find a (local) minimum. Traditional approaches associated with regularization and Markov Random Fields include continuation, simulated annealing, highest confidence first, and mean-field annealing.

More recently, max-flow and graph-cut methods have been proposed to solve a special class of global optimization problems. Such methods are more efficient than simulated annealing and have produced good results.

Dynamic Programming. A different class of global optimization algorithms are those based on dynamic programming. Dynamic programming can find the global minimum for independent scanlines in polynomial time. Dynamic programming was first used for stereo vision in sparse, edge-based methods. More recent approaches have focused on the dense (intensity-based) scanline optimization problem. These approaches work by computing the minimum-cost path through the matrix of all pairwise matching costs between two corresponding scanlines. Partial occlusion is handled explicitly by assigning a group of pixels in one image to a single pixel in the other image.

Problems with dynamic programming stereo include the selection of the right cost for occluded pixels and the difficulty of enforcing inter-scanline consistency, although several methods propose ways of addressing the latter. Another problem is that the dynamic programming approach requires enforcing the monotonicity or ordering constraint. This constraint requires that the relative ordering of pixels on a scanline remain the same between the two views, which may not be the case in scenes containing narrow foreground objects.

Cooperative Algorithms. Finally, cooperative algorithms, inspired by computational models of human stereo vision, were among the earliest methods proposed for disparity computation. Such algorithms iteratively perform local computations, but use nonlinear operations that result in an overall behavior similar to global optimization algorithms. In fact, for some of these algorithms, it is possible to explicitly state a global function that is being minimized.

4.2.4 Refinement of Disparities

Most stereo correspondence algorithms compute a set of disparity estimates in some discretized space, e.g., for integer disparities (exceptions include continuous optimization techniques such as optic flow or splines). For applications such as robot navigation or people tracking, these may be perfectly adequate. However for image-based rendering, such quantized maps lead to very unappealing view synthesis results (the scene appears to be made up of many thin shearing layers). To remedy this situation, many algorithms apply a sub-pixel refinement stage after the initial discrete correspondence stage. (An alternative is to simply start with more discrete disparity levels.)

Sub-pixel disparity estimates can be computed in a variety of ways, including iterative gradient descent and fitting a curve to the matching costs at discrete disparity levels. This provides an easy way to increase the resolution of a stereo algorithm with little additional computation. However, to work well, the intensities being matched must vary smoothly, and the regions over which these estimates are computed must be on the same (correct) surface.

Besides sub-pixel computations, there are of course other ways of post-processing the computed disparities. Occluded areas can be detected using cross-checking (comparing left-to- right and right-to-left disparity maps). A median filter can be applied to “clean up” spurious mismatches, and holes due to occlusion can be filled by surface fitting or by distributing neighboring disparity estimates [13, 96].

4.2.5 Other methods

Generally, stereo algorithms first enumerate all possible matches at all possible disparities, then select the best set of matches in some way. This is a useful approach when a large amount of ambiguity may exist in the computed disparities. An alternative approach is to use methods inspired by classic (infinitesimal) optic flow computation. Here, images are successively warped and motion estimates incrementally updated until a satisfactory registration is achieved.

A univalued representation of the disparity map is also not essential. Multi-valued representations, which can represent several depth values along each line of sight, have been extensively studied recently, especially for large multiview data set. Many of these techniques use a voxel-based representation to encode the reconstructed colors and spatial occupancies or opacities. Another way to represent a scene with more complexity is to use multiple layers, each of which can be represented by a plane plus residual parallax. Finally, deformable surfaces of various kinds have also been used to perform 3D shape reconstruction from multiple images.

4.3 Hardware analysis

4.3.1 General description of how the hardware of the project is implemented

For the implementation of the project the following items were used:

- 2 CCD cameras
- 1 cube beam splitter
- 1 relay lens
- 1 lens
- 1 pair of 3D glasses
- 2 Triple band pass filters

Two charged-coupled device (CCD) image sensors are located to provide the left and the right image respectively. These two image signals are processed to generate high-resolution anaglyph 3D image. This image is then displayed on a monitor and viewed through 3D glasses to provide realistic three-dimensional images.

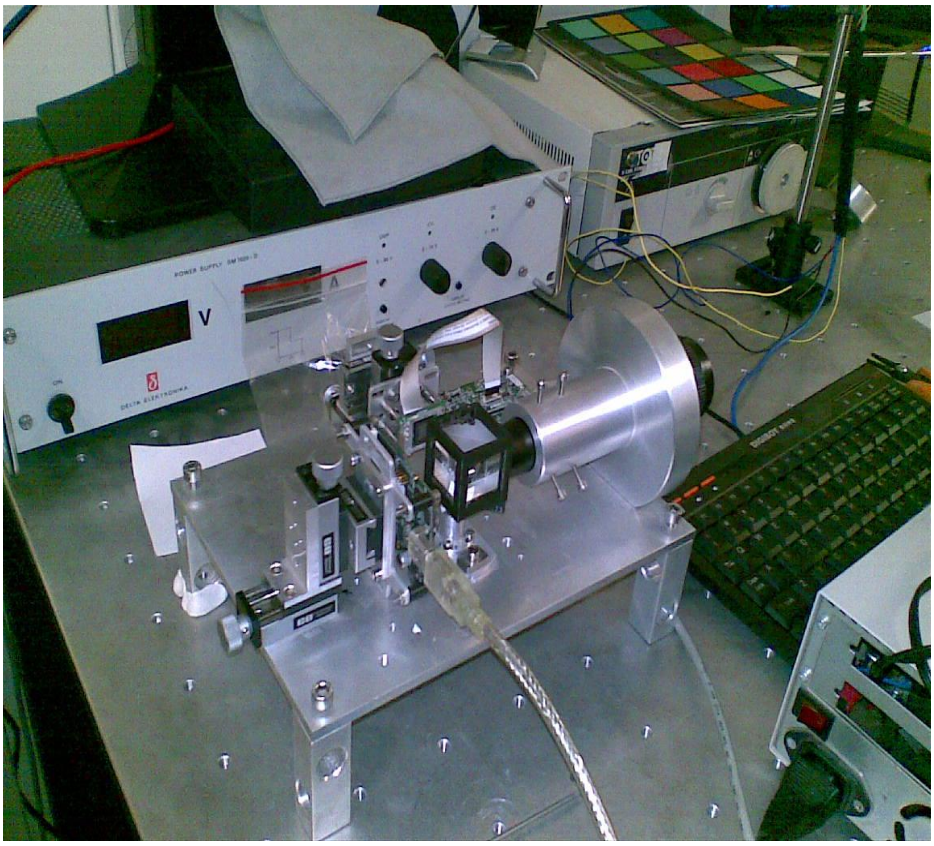

Figure 24. Garida Stereo

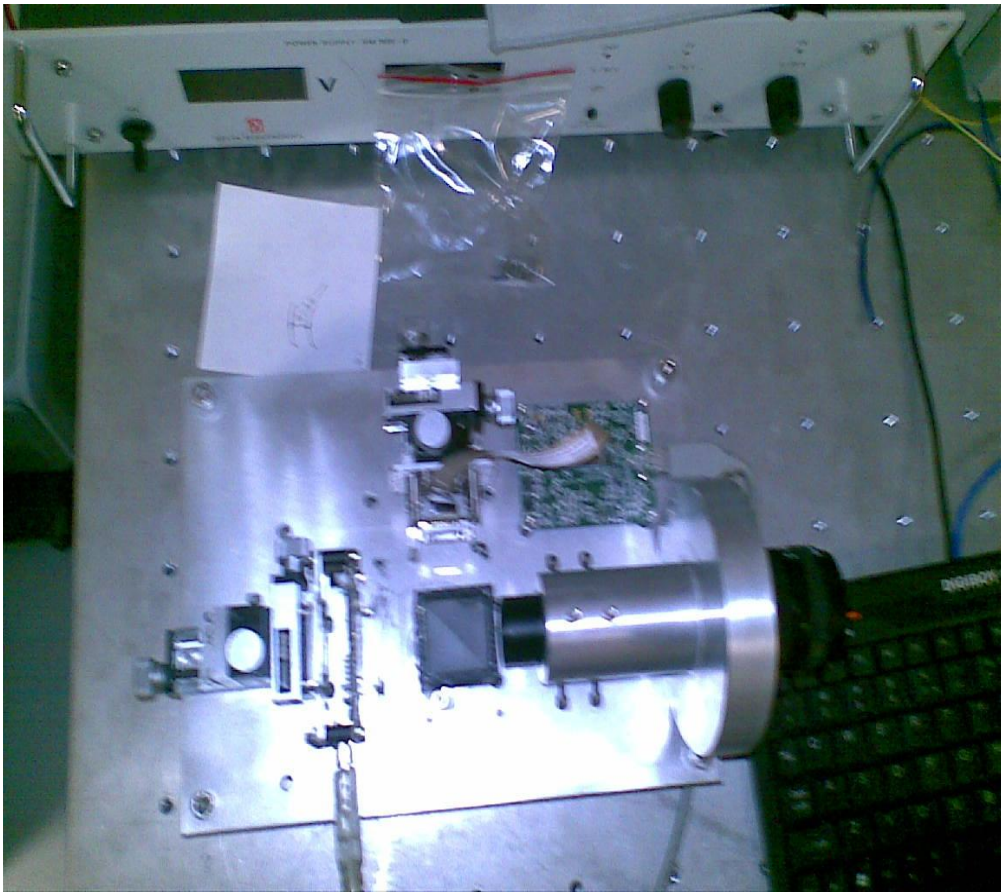

Figure 25. Garida Stereo

4.3.2 Charged-Couple Device

The CCD camera which is used for this implementation was the Dragonfly 2 Model. The most important features of the camera are the following:

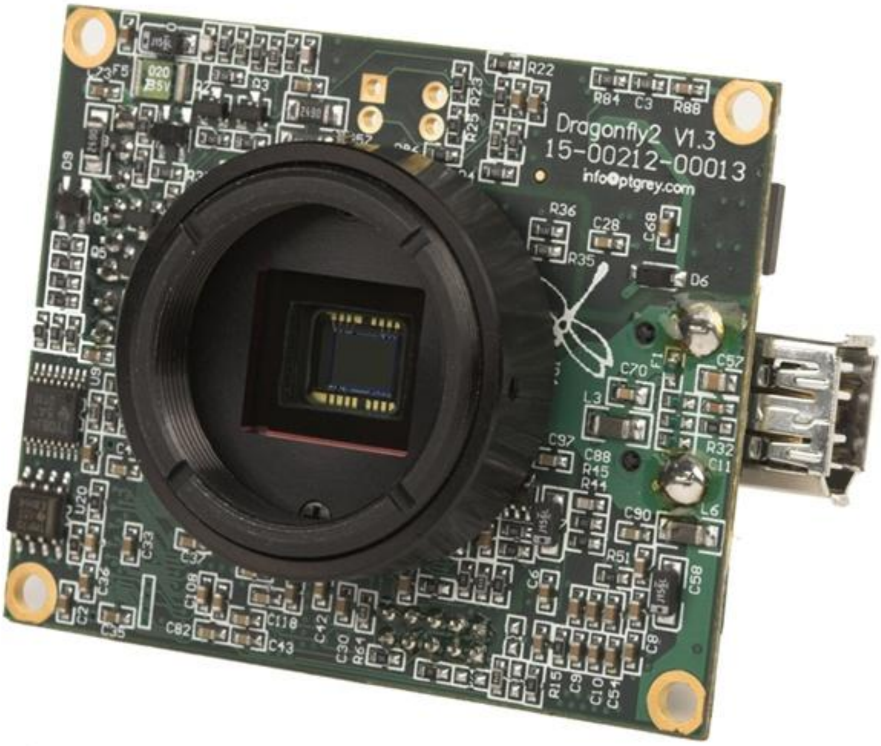

Figure 26. Dragonfly®2 camera

Camera Specifications

Overview	OEM board-level camera
Imaging Sensor	Sony 1/3" progressive scan CCD
Sensor Size	Diagonal 6 mm
Sensor Active Pixels	640(H) x 480(V)
Sensor Chip Size	5.79mm(H) x 4.89mm(V)
Sensor Unit Cell Size	7.4um(H) x 7.4um(V)
A/D Converter	Analog Devices 12-bit A/D converter
Video Data Output	8 and 16-bit digital data
Standard Resolutions	640x480, 320x240, 160x120
Frame Rates	60, 30, 15, 7.5, 3.75, 1.875 FPS
Partial Image Modes	Pixel binning and region of interest modes
Interfaces	6-pin IEEE-1394 for camera control and video data transmission 4 general purpose digital input/output (GPIO) pins
Voltage Requirements	8-32V via IEEE-1394 cable or GPIO connector
Power Consumption	Less than 2W

Gain	Automatic/Manual/One-Push Gain modes 0dB to 24 dB
Shutter	Automatic/Manual/One-Push Shutter modes 0.01ms to 66.63ms @ 15FPS Extended Shutter Modes 0.01ms to 7900ms @ 15FPS
Gamma	0.50 to 4.00
Trigger Modes	DCAM v1.31 Trigger Modes
Signal to Noise Ratio	Greater than 60 dB
Dimensions	63.5mm x 50.8mm x 13.15mm (bare board w/o case or lens holder)
Mass	25 grams (bare board w/o case or optics
Camera Specification	IIDC 1394-based Digital Camera Specification v1.31
Emissions Compliance	Complies with CE rules and Part 15 class A of FCC
Operating Temperature	Commercial grade electronics rated from 0^0 – 45^0 C
Storage Temperature	-30^0 - 60^0 C
Operative Relative Humidity	20 to 80% (no condensation)
Storage Relative Humidity	20 to 95% (no condensation)

Image Acquisition

- 12-bit analog-to-digital converter
- Multiple Dragonfly2's on the same 1394 bus automatically sync
- Faster standard frame rates plus pixel binning and ROI support
- Bulp-trigger mode, multiple triggered exposures before readout
- Overlapped trigger input, image acquisition and transfer

Image Processing

- On-camera conversion to YUV411, YUV422 and RGB formats
- On-camera control of sharpness, hue, saturation, gamma, LUT
- Horizontal image flipping (mirror image)
- Pixels contain frame-specific info (e.g. shutter, 1394 cycle time)

- Continuous static image for testing and development

Camera and Device Control

- Apply settings (e.g. shutter, gain) to all cameras on the same bus
- On-board DC output for use by an auto iris lens
- Auto and one-push white balance for easy color balancing
- Reports the temperature near the imaging sensor
- Monitors sensor voltages to ensure optimal image quality
- Fine-tune frame rates for video conversion (e.g. PAL @ 24FPS)
- Increased drive strength, configurable strobe pattern output
- Provides serial communication via GPIO TTL digital logic levels
- Non-volatile storage of camera default settings and user data
- General purpose input/output pins for external device control
- Firmware upgradeable in field via IEEE-1394 interface.

Generally, in selection of the appropriate sensors of stereo systems, high resolution in order to achieve the expected accuracy, low noise and good image quality, greater viewing angles up to cover the environment, good geometric stability, and camera controlling, synchronization, and low weight are important parameters.

CCD cameras are computer vision cameras that are used more in monitoring and control applications. The input to the camera is, as we know, the incoming light, which enters the camera's lens and hits the image plane. In a CCD camera, the physical image plane is the CCD array, a n x m rectangural grid of photosensors, each sensitive to light intensity. Each photosensor can be regarded as a tiny, rectangular black box which converts light energy into a voltage. The output of the CCD array is usually a continuous electric signal, the video signal, which we can regard as generated by scanning the photosensors in a CCD array in a given order (e.g., line by line) and reading out their voltages. The video signal is sent to an electronic device called a frame grabber, where it is digitalized into a 2-D, rectangular array of N x M integer values and stored in a memory buffer. At this point, the image can be conveniently represented by a N x M matrix, E, whose entries are called pixels, with N and M being two fixed integers expressing the image size, in pixels, along each direction. Finally, the matrix E is transferred to a host computer for processing.

The main advantage of this type of cameras is that they are offering synchronized images, they can be controlled and we can achieve high dynamic range and low noise images. These cameras are offered in various sizes and formats. Generally, the following features of these cameras make them different from each other:

- functioning in outside environment (indoor/outdoor)
- imaging in the day and night (equipped with IR)
- different focal length and viewing angle
- different resolution and quality
- ability the lens to mount on

4.3.3 Cube beam splitter

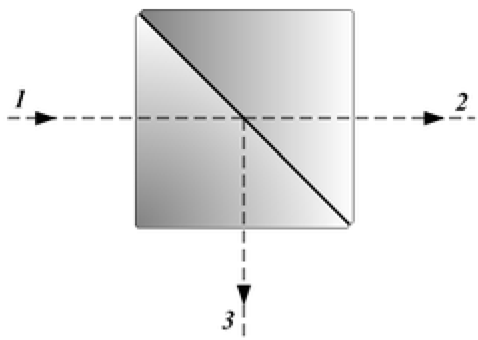

Figure 27. Schematic illustration of a cube beam splitter

Cube beam splitters are a type of beam splitter used in many life science or laser applications and biomedical instrumentation. Cube beam splitters are used to split incident light into two separate components. Cube beam splitters are durable, easy to mount beam splitters that feature equal optical path lengths. In its most common form, a cube, it is made from two triangular glass prisms which are glued together at their base using polyester, epoxy, or urethane-based adhesives. The thickness of the resin layer is adjusted such that (for a certain wavelength) half of the light incident through one "port" (i.e., face of the cube) is reflected and the other half is transmitted due to frustrated total internal reflection. Instead of glass, crystalline media can be used, which can be birefringent. This allows the construction of various types of polarizing beam splitter cubes such as Wollaston prisms and Nomarski prisms, where the two output beams emerge from the same face, and the angle between these beams is typically between 15° and 45°, i.e., much smaller than shown in Figure 28.

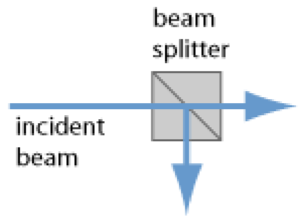

Figure 28. Beam splitter

Antireflection coatings on the entry and exit faces of the cube minimize loss and reduce ghost reflections (though they are still present). Cube beam splitters eliminate beam displacement without being fragile. They are easy to mount and mechanically durable, but the presence of an interface can limit power handling if epoxy is used for bonding. For best spectral performance and transmitted wavefront, cube beam splitters should be used with collimated or near-collimated light, as convergent or divergent beams will contribute unwanted spherical aberration to an optical system. If one prism is marked with a dot, this indicates the coated prism. For best performance, the optical beam should traverse this prism first.

Polarizing beam splitters, such as the Wollaston prism, use birefringent materials, splitting light into beams of differing polarization. Beam splitters cubes can be used not only for simple light beams, but also for beams carrying images, e.g. in various types of cameras and projectors. In our project, the cube beam splitter is used to split the image into two separate images and to project them into the CCDs.

4.3.4 Relay Lens

Relay Lenses are used to extend imaging systems or invert images for use in a range of imaging applications including microscopy, inspection, or life sciences. Relay Lenses use arrays or imaging lenses placed behind eyepieces to extend the length of imaging systems, such as in microscopes, endoscopes, or where proximity to a viewed object is not feasible. Relay Lenses may also be used to invert an image behind the eyepiece of an imaging system for proper viewing. Relay Lenses are available in a wide range of options suited to a number of relay needs. They may be made of one or more conventional lenses or achromatic doublets, or a long cylindrical gradient-index of refraction lens (a GRIN lens). Relay lenses operate by producing intermediate planes of focus. For example, an objective lens such as an SLR lens produces an image plane where the image sensor would usually go. If you place a lens with focal length *f* a distance 2*f* from that image plane and then put an image sensor 2*f* beyond the lens, that lens will relay the first image to the second with 1:1 magnification. Ideally, this second image plane will be identical to the first, so you could put a sensor there and record exactly the same image. If a longer distance is needed, this can be repeated.

Figure 29. Relay Lenses

4.3.5 Lens

The opening or aperture of a lens, also known as the "iris", greatly affects the amount of light reaching the sensor. The f-number of the lens is the quotient of the focal length of the lens and the diameter of the opening. The higher the f-number, the smaller the opening will be, and vice versa. A lower f-number means that more light will reach the sensor. The aperture also affects the depth of field, that is, how much of the scene that can be in focus at the same time. A wide-open lens will have a very shallow depth of field. Objects slightly closer to or further from the camera than the set focus point will be out of focus. By increasing the f-number (and thereby closing the light opening), the depth of field increases, and the objects can be brought back into focus.

4.4 Camera Calibration

Calibration of a stereo camera system involves the estimation of the intrinsic parameters of each camera. In order to optimally capture a scene for streaming, it's imperative that the cameras used to capture the scene are properly calibrated. This will result in clear, properly exposed images with good white balance, focus, contrast, luminance, shutter speed, gain.

The camera calibration problem is to relate the locations of pixels in the image array to points in the scene. Since each pixel is imaged through perspective projection, it corresponds to a ray of points in the scene. The camera calibration problem is to determine the equation for this ray in the absolute coordinate system of the scene. The camera calibration problem includes both the exterior and interior orientation problems, since the position and orientation of the camera and the camera constant must be determined to relate image plane coordinates to absolute coordinates, and the location of the principal point, the aspect ratio, and lens distortions must be determined to relate image array locations (pixels coordinates) to positions in the image plane. The camera calibration problem involves determining two sets of parameters: the extrinsic parameters for rigid body transformation (exterior orientation) and the intrinsic parameters for the camera itself (interior orientation).

We can use an initial approximation for the intrinsic parameters to get a mapping from image array (pixel) coordinates to image plane coordinates. Suppose that there are n rows and m columns in the image array and assume that the principal point is located at the center of the image array:

$$c_x = (m-1)/2$$

$$c_y = (n-1)/2$$

The image plane coordinates for the pixel at grid location [i,j] are

$$x' = \tau_X d_x(j - c_x)$$

$$y' = -d_y(i - c_y)$$

where d_x and d_y are the center-to-center distances between pixels in the *x* and *y* directions, respectively, and τ_x is a scale factor that accounts for distortions in the aspect ratio caused by timing problems in the digitizer electronics. The row and column distances, *dx* and *dy*, are available from the specifications for the CCD camera and are very accurate, but the scale factor τ_x must be added to the list of intrinsic parameters for the camera and determined through calibration. Note that these are uncorrected image coordinates, marked with a tilde to emphasize that the effects of lens distortions have not been removed. The coordinates are also affected by errors in the estimates for the location of the principal point (c_x,c_y) and the scale factor τ_x.

We must solve the exterior orientation problem before attempting to solve the interior orientation problem, since we must know how the camera is positioned and oriented in order to know where the calibration points project into the image plane. Once we know where the projected points should be, we can use the projected locations p'_I and the measured locations p_i, to determine the lens distortions and correct the location of the principal point and the image aspect ratio. The solution to the exterior orientation problem must be based on constraints that are invariant to the lens distortions and camera constant, which will not be known at the time that the problem is solved.

4.4.1 Binocular Stereo Calibration

There are several tasks in developing a practical system for binocular stereo:

1. Calibrate the intrinsic parameters for each camera.
2. Solve the relative orientation problem.
3. Resample the images so that the epipolar lines correspond to image rows.
4. Compute conjugate pairs by feature matching or correlation.
5. Solve the stereo intersection problem for each conjugate pair.
6. Determine baseline distance.
7. Solve the absolute orientation problem to transform point measurements from the coordinate system of the stereo cameras to an absolute coordinate system for the scene.

There are several ways to calibrate a binocular stereo system. To start, each camera must be calibrated to determine the camera constant, location of the principal point, correction table for lens distortions, and other intrinsic parameters. Once the left and right stereo cameras have been calibrated, there are basically three approaches to using the cameras in a stereo system.

The first approach is to solve the relative orientation problem and determine the baseline by other means, such as using the stereo cameras to measure points that are a known distance apart. This fully calibrates the rigid body transformation between the two cameras. Point measurements can be gathered in the local coordinate system of the stereo cameras. Since the baseline has been calibrated, the point measurements will be in real units and the stereo system can be used to measure the relationships between points on objects in the scene. It is not necessary to solve the absolute orientation problem, unless the point measurements must be transformed into another coordinate system.

The second approach is to solve the relative orientation problem and obtain point measurements in the arbitrary system of measurement that results from assuming unit baseline distance. The point measurements will be correct, except for the unknown scale factor. Distance ratios and angles will be correct, even though the distances are in unknown units. If the baseline distance is obtained later, then the point coordinates can be multiplied by the baseline distance to get point measurements in known units. If it is necessary to transform the point measurements into another coordinate system, then solve the absolute orientation problem with scale, since this will accomplish the calibration of the baseline distance and the conversion of point coordinates into known units without additional computation.

The third approach is to solve the exterior orientation problem for each stereo camera. This provides the transformation from the coordinate systems of the left and right camera into absolute coordinates. The point measurements obtained by intersecting rays will automatically be in absolute coordinates with known units, and no further transformations are necessary.

4.4.2 White Balance

White balance is a name given to a system of color correction to deal with differing lighting conditions. Setting a camera's white balance removes unrealistic color in a scene so that objects that appear white are rendered white in a video frame. Proper camera white balance has to take into account the "color temperature" of a light source, which refers to the relative warmth or coolness of white light. Our eyes are very good at judging what is white under different light sources, but digital cameras often have great difficulty with auto white balance (AWB) — and can create unsightly blue, orange, or even green color casts. At its simplest – the reason we adjust white balance is to get the colors in our images as accurate as possible. The Dragonfly2 supports white balance. Adjusting the white balance by modifying the relative gain of R, G and B in an image enables white areas to look "whiter". Taking some subset of the target image and looking at the relative red to green and blue to green response, the general idea is to scale the red and blue channels so that the response is 1:1:1.

The Dragonfly2 also implements Auto and One_Push white balance. One of the uses of one_push / auto white balance is to obtain a similar color balance between different cameras that are slightly different from each other. Theoretically, if different cameras are pointed at the same scene, using one_push / auto will result in a similar color balance between the cameras.

One_push is similar identical to auto white balance, except One_Push only attempts to automatically adjust white balance for a set period of time before stopping. The white balance of the camera before using One_Push/Auto must already be relatively close. However, if the camera is already close to being color balanced, then it will work (it may only be a small change). One_push only attempts to automatically adjust white balance for a set period of time before stopping. It uses a “white detection” algorithm that looks for “whitish” pixels in the raw Bayer image data. One_push adjusts the white balance for a specific number of iterations; if it cannot locate any whitish pixels, it will gradually look at the whitest objects in the scene and try to work off them. It will continue this until it has completed its finite set of iterations. Auto is continually adjusting white balance. It differs from one_push in that it works almost solely off the whitest objects in the scene.

We used color targets, such as the ColorChecker, which can be captured by the cameras, and the resulting images’ output can be compared to the original chart, or to reference measurements, to test the degree to which image acquisition reproduction systems and processes approximate the human visual system.

Figure 30. Color Checker

4.4.3 Shutter Speed

The shutter speed is the amount of time that the camera takes to register an image. A camera's shutter speed setting controls the amount of light in a frame and to allow a camera to capture fast motion properly. With a lot of light entering the camera, faster shutter speeds are possible. As light decreases, the shutter speed slows down because the sensor needs more time to "collect" enough light to form an image. With a sufficiently low shutter speed, anything moving in the scene will appear blurred in the image, since the moving object's position changes during the capture. This is called "motion blur" (or "ghosting" in certain applications), and will have a negative effect on both image quality and usability of our video. Faster shutter speeds reduce blurriness but also limit the amount of light that gets into the video.

The Dragonfly2 supports automatic, manual and one_push control of the CCD shutter time. Shutter times are scaled by the divider of the basic frame rate. For example, dividing the frame rate by two (e.g. 15 FPS to 7.5 FPS) causes the maximum shutter time to double (e.g. 66ms to 133ms).

The time between the end of shutter for consecutive frames will always be constant. However, if the shutter time is continually changing (e.g. shutter is in Auto mode being controlled by Auto Exposure), the time between the beginning of consecutive integrations will change. If the shutter time is constant, the time between integrations will also be constant. The Dragonfly2 will continually expose and read image data off of the sensor under the following conditions:

2. The camera is not in asynchronous trigger mode. When in async trigger mode, the camera simply clears the sensor and does not read the data off the sensor.

It is important to note that the camera will continue exposing images even when isochronous data transfer is disabled and images are not being streamed to the PC.

4.4.4 Gain

Many cameras employ an internal boost of the image signal; this is called gain. To enable capture in low light without sacrificing shutter speed or depth of field, the weak sensor signal is electronically amplified, resulting in a brighter image. A side effect of this is that tiny imperfections in the capture are also amplified and are produced as image noise. This noise degrades image quality and generally increases the bandwidth needed for the video stream. That's why we disable gain by setting it to 0. The Dragonfly 2 supports automatic, manual and one-push gain modes.

4.5 Why use anaglyph 3d?

Among many three-dimensional (3D) displaying techniques, e.g., glasses dependent or auto-stereoscopic, anaglyph is the least expensive way to make the 3D visual experience achievable on ordinary monitors or even prints, with no special hardware but only cheap colored glasses. Despite of its many inherent deficiencies, its resurgence has been seen recently thanks to the abundance of 3D content which is more easily accessible nowadays than ever.

Like other 3D displaying techniques, anaglyph can provide a slightly different view to each of two eyes. From the disparity between the two views and other visual cues, the human visual system can generate the stereoscopic representation of spatial relationships in the scene. An anaglyph image is formed by superimposing two views (i.e., left and right images of a stereo image pair) in different colors. When perceived through colored glasses, the two images will be separated from the composite one to feed each eye. The separation is based on color filtering. Fig.31 illustrates how anaglyph works. As shown in Fig.31, a red- cyan color anaglyph image is generated by combining the red channel $\{R'_l\}$[2] of the left image and the blue and green channels $\{G'_r, B'_r\}$ of the right image together, i.e., $\{R'_l, G'_r, B'_r\}$. Wearing red-cyan colored glasses, the left eye perceives $\{R'_l\}$ since the red lens blocks most of the green and blue colors. And the right eye perceives $\{G'_r, B'_r\}$ due to the opposite color filtering property of the cyan lens.

Anaglyph images mainly suffer from three deficiencies, i.e., color distortions, retinal rivalry, and ghosting effect. Color distortions cause the colors perceived through the glasses to be quite different from those of the original scene. This can be observed from Fig. 31 by comparing the original stereo pair with the simulated perception of the two eyes. Retinal rivalry occurs when the same object appears to have different colors in the two eyes. This phenomenon can be observed from Fig. 31 by comparing the simulated perception of the two eyes. Retinal rivalry is distracting, and causes visual fatigue and other side effects. As the color gamuts of our eyes behind the colored glasses do not overlap, retinal rivalry is inevitable. However, by manipulating the lightness, retinal rivalry can be significantly reduced. Color anaglyph does not attempt to control the retinal rivalry effect in its formation. It is evident from Fig. 31 that fishes perceived by the left eye are much brighter than those of the right eye. Under this circumstance the retinal rivalry will be extremely severe. Ghosting is a common problem for most 3D displaying techniques. In anaglyph, ghosting is caused by the imperfect filtering of the light wavelengths, so that the unwanted image leaks through the colored glasses mixing with the intended one. It is often more obvious in regions where there is a striking contrast and a large binocular disparity.

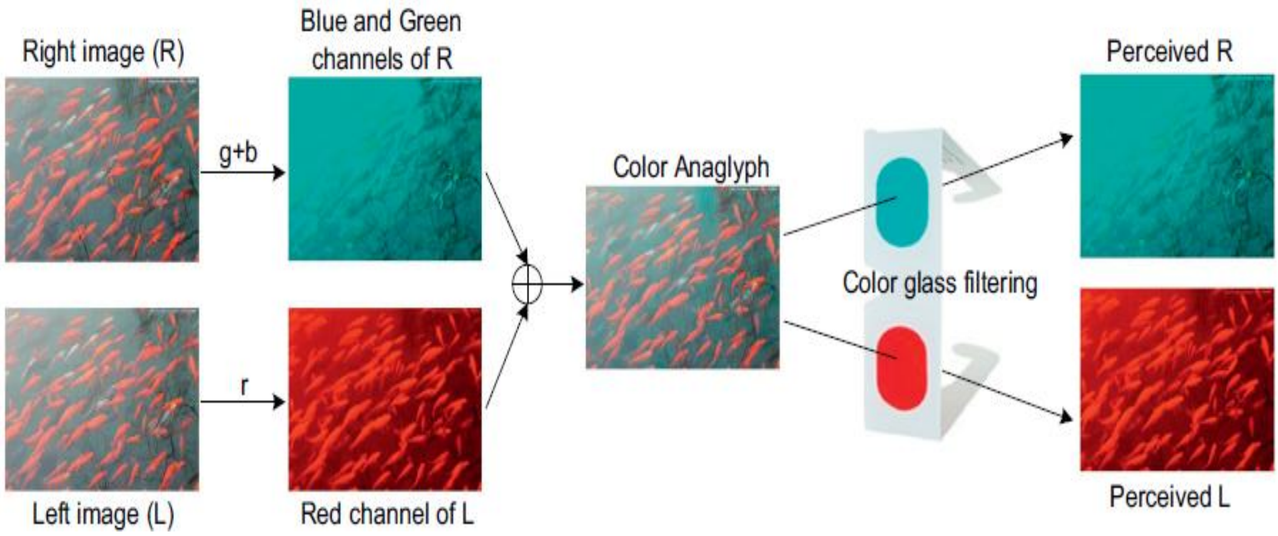

Figure 31. An illustration of anaglyph 3D displaying using color anaglyph.

Dubois [19] proposed an anaglyph generation method, which takes into account the spectral distributions of display primaries and transmission functions of the colored glasses. In general, the principle of this anaglyph generation method is to minimize the color distortions, i.e., differences between the perception of the original stereo pair and that of the anaglyph image through the colored glasses, in the CIE 1931 XYZ color space.

McAllister et al. [20] proposed an anaglyph method, namely LAB anaglyph, minimizing the color distortions in the CIELAB color space. The CIELAB color space exhibits several major advantages: (a) it is regarded as a uniform colorspace, i.e., the Euclidean distance in the CIELAB colorspace correlates well with the perceptual color distance; (b) it incorporates chromatic adaptation transform and non-linear response compression to more accurately simulate the perception of the human visual system (HVS); (c) it provides means for transforming tri-stimulus values to several color appearance attributes, i.e., lightness, saturation, hue, etc.

Woods et al. [21] discussed in detail the source of the anaglyph ghosting. In [22], three methods were proposed to reduce ghosting, i.e., stereo pair registration, color component blurring, and depth manipulation. In [23], the authors proposed to inhibit ghosting by controlling the amount of saturation. In [24], ghosting reduction was implemented in a sequential process, i.e., analyzing differences between the left and right images, detecting the ghosting area, and eventually adjusting the intensities of the ghosting area. Tran [25] proposed ghosting reduction methods which relied on explicit knowledge of properties of the display device and the colored glasses. They can serve as post-processing components for existing anaglyph algorithms. Recently, Sanftmann et al. [26] defined a model to quantify the perceived luminance through the colored glasses. Five parameters of the model can be captured by simple subjective tests. Illustration was given on how to use the model together with several simple anaglyph methods to reduce the ghosting artifacts.

4.5.1 Anaglyph methods comparison

In this section several anaglyph methods are compared. The formulas to calculate the anaglyph's RGB values r_a, g_a and b_a from the RGB values of the original left (r_1, g_1, b_1) and right images (r_2, g_2, b_2) are presented. The formulas must be applied for each pixel.

True Anaglyphs or Red/Blue Monochrome Mode

$$\begin{pmatrix} r_a \\ g_a \\ b_a \end{pmatrix} = \begin{pmatrix} 0{,}299 & 0{,}587 & 0{,}114 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} r_1 \\ g_1 \\ b_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0{,}299 & 0{,}587 & 0{,}114 \end{pmatrix} \cdot \begin{pmatrix} r_2 \\ g_2 \\ b_2 \end{pmatrix}$$

Figure 32. True Anaglyph

Each R, B output is simply the same average of the three color channels. Instead of merely adding the R, G, B values and dividing by three, the proportions used are the accepted values for converting a color image to its grayscale equivalent and are based on the radiometric sensitivity of the eye.

The left eye sees a grayscale red image and the right eye sees a grayscale blue image. All color information has been lost, the image is a bit dark but the stereo effect is good as there is little ghosting.

Gray Anaglyphs or Red/Cyan Monochrome Mode

$$\begin{pmatrix} r_a \\ g_a \\ b_a \end{pmatrix} = \begin{pmatrix} 0{,}299 & 0{,}587 & 0{,}114 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} r_1 \\ g_1 \\ b_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0{,}299 & 0{,}587 & 0{,}114 \\ 0{,}299 & 0{,}587 & 0{,}114 \end{pmatrix} \cdot \begin{pmatrix} r_2 \\ g_2 \\ b_2 \end{pmatrix}$$

Figure 33. Gray Anaglyph

Each R,G,B output is simply the same grayscale average. So, the left eye sees a grayscale red image and the right eye sees a grayscale cyan image. All color information has been lost, the image is brighter than red/blue anaglyphs and the stereo effect is good but with more ghosting than red/blue mode.

Half Color Anaglyphs

$$\begin{pmatrix} r_a \\ g_a \\ b_a \end{pmatrix} = \begin{pmatrix} 0{,}299 & 0{,}587 & 0{,}114 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} r_1 \\ g_1 \\ b_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} r_2 \\ g_2 \\ b_2 \end{pmatrix}$$

Figure 34. Half Color Anaglyph

All of the blue and green information is presented to the right eye but only a grayscale representation to the left eye. This produces less retinal rivalry but the color reproduction is not as good as Color Anaglyph mode.

Color Anaglyphs

$$\begin{pmatrix} r_a \\ g_a \\ b_a \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} r_1 \\ g_1 \\ b_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} r_2 \\ g_2 \\ b_2 \end{pmatrix}$$

Figure 35. Color Anaglyph

In this method, which is the one we used to generate 3d images, all of the red information of the left image is presented to the left eye and all of the blue and green information is presented to the right eye. Color reproduction can be quite good but is more prone to retinal rivalry effects.

Optimized Anaglyph Mode

$$\begin{pmatrix} r_a \\ g_a \\ b_a \end{pmatrix} = \begin{pmatrix} 0 & 0{,}7 & 0{,}3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} r_1 \\ g_1 \\ b_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} r_2 \\ g_2 \\ b_2 \end{pmatrix}$$

Figure 36. Optimized anaglyph

In the above example, all of the blue and green information is presented to the right eye, but none of the left image red color data has been used in deriving the output red channel. Instead, 30% of the green channel and 70% of the blue channel are used. This would eliminate rivalry caused by the red component of the image but color reproduction is not accurate.

4.6 Results

In this section, we present several stereo pair images, which are converted to three-dimensional (3D) images, using the Color Anaglyph method. As discussed in previous sections, the two viewpoints should be on the same horizontal line and not too far apart, or the brain cannot fuse the stereo pair. It is not essential for a 3D impression that the stereo pair should the same distance apart as human eyes (65mm adult average). In order to perceive the stereoscopic effect we shift the right-eye image to the left while holding the left-eye image fixed.
It is necessary that you're wearing 3D red cyan glasses to view the following pictures.

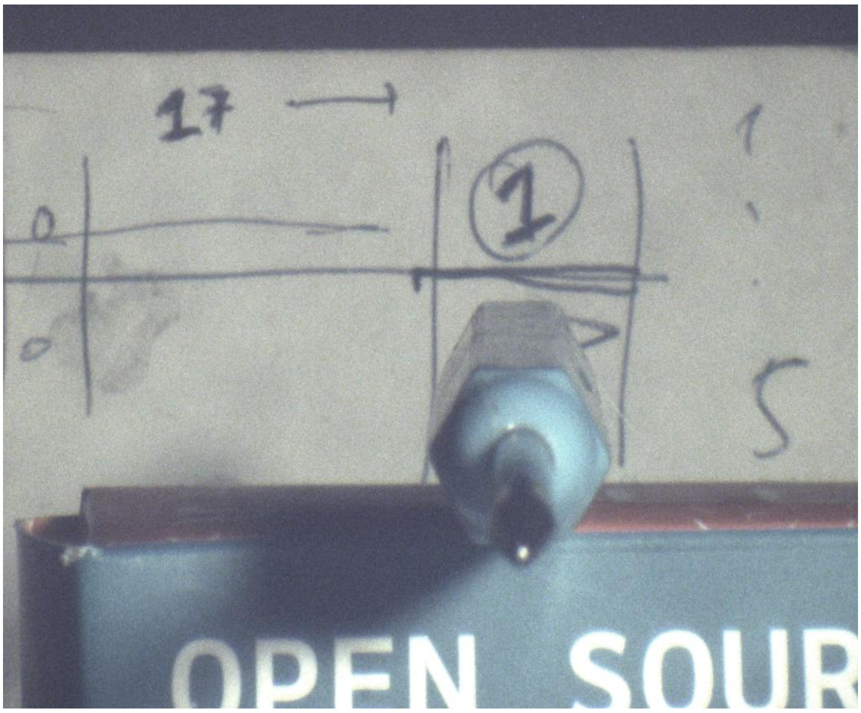

Figure 37. Reference image

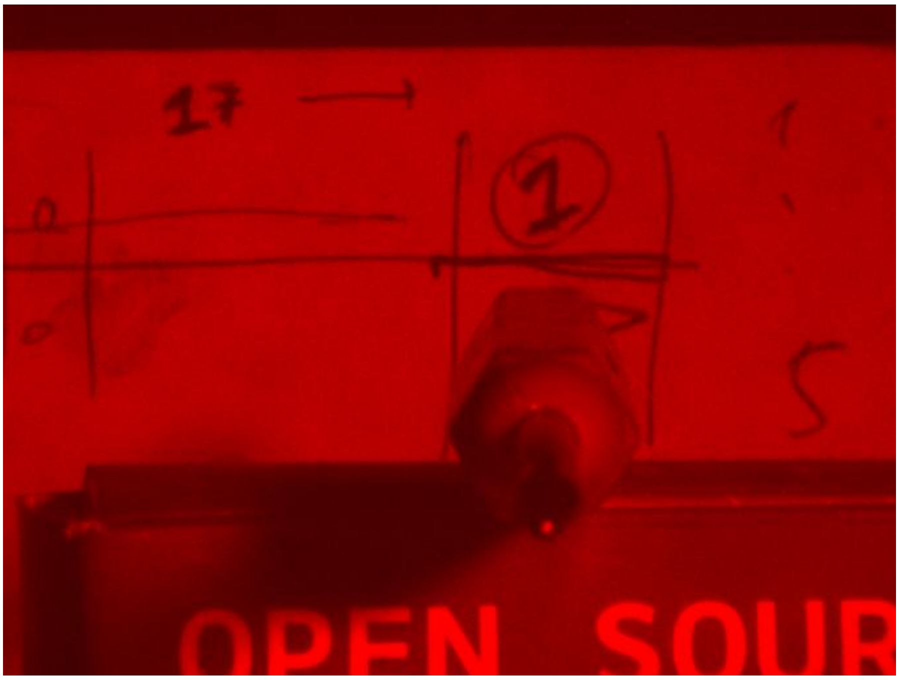

Figure 38. Remove cyan from reference image

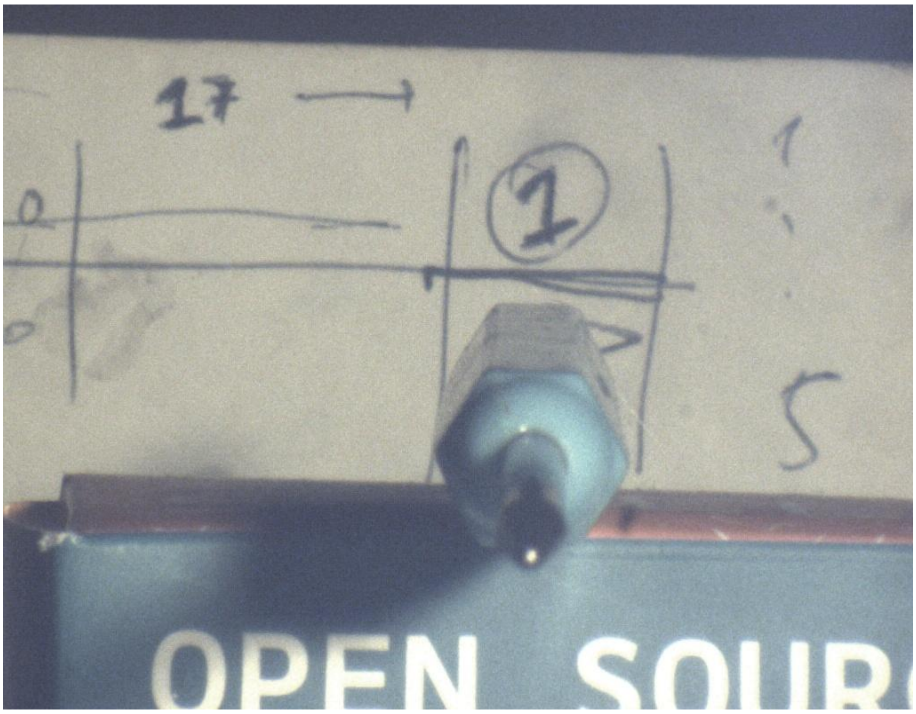

Figure 39. Target Image

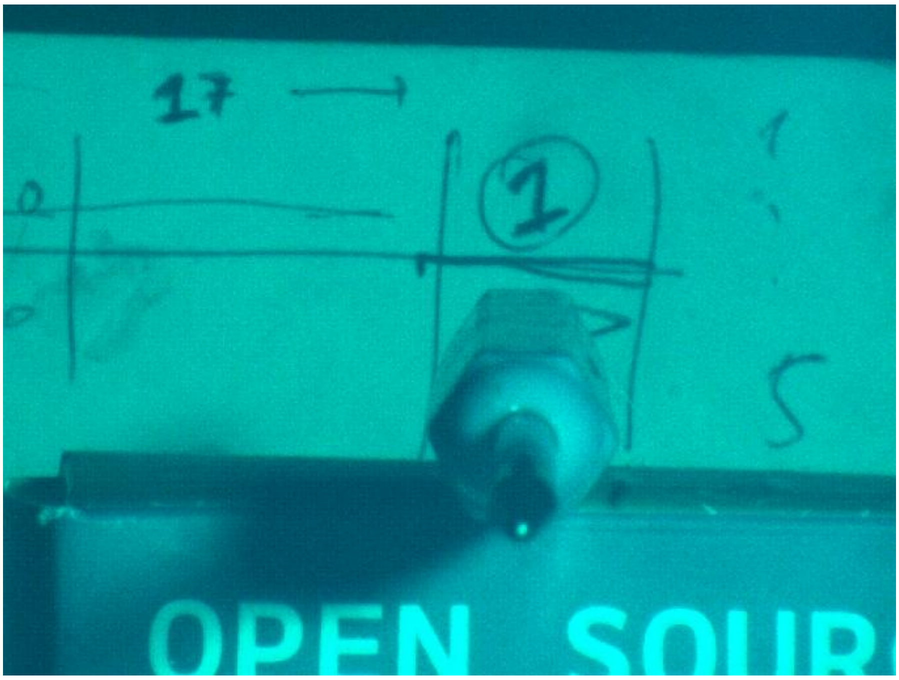

Figure 40. Remove red from target image

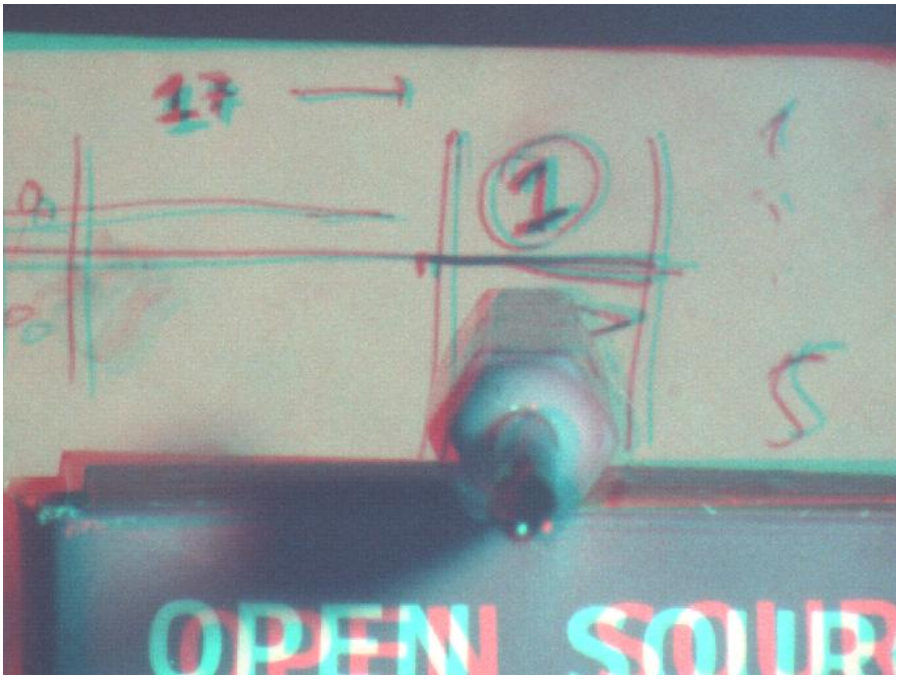

Figure 41. Anaglyph 3D image

Figure 42. 3D image using triple bandpass filters

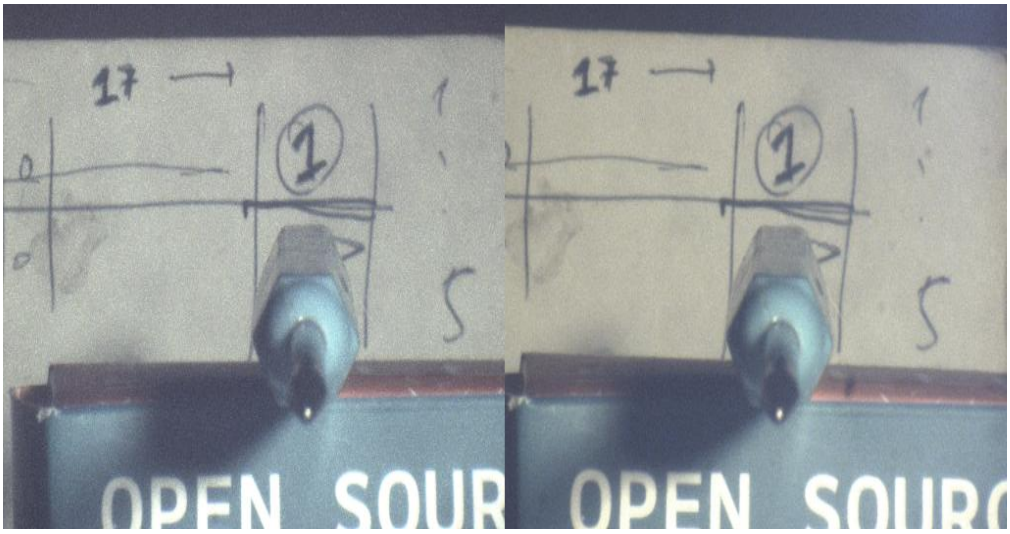

Figure 43. Reference and Target image

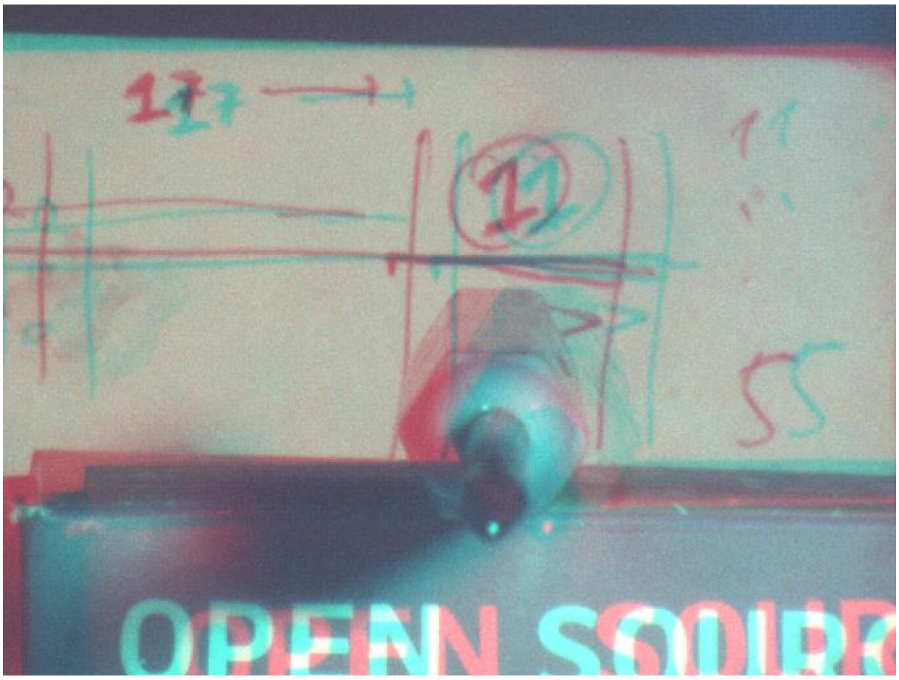

Figure 44. Anaglyph 3D image

Figure 45. 3D image using triple bandpass filters

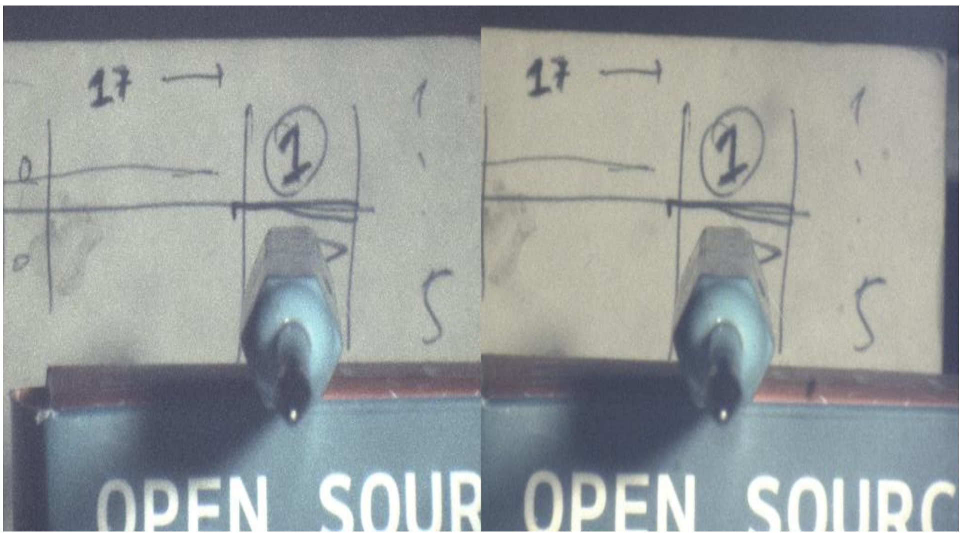

Figure 46. Reference and Target Image

Figure 47. Anaglyph 3D image

Figure 48. 3D image using triple band pass filters

It seems that the use of triple band pass filters contribute to the improvement of the stereoscopic effect.

It is possible to construct 3D anaglyph images, if we place the cyan lens of the red-cyan glasses in front of the one CCD, and the red lens in front of the other. If we compare them with real 3D images, we note that there is not enough difference between them, as far as the stereoscopic effect is concerned.

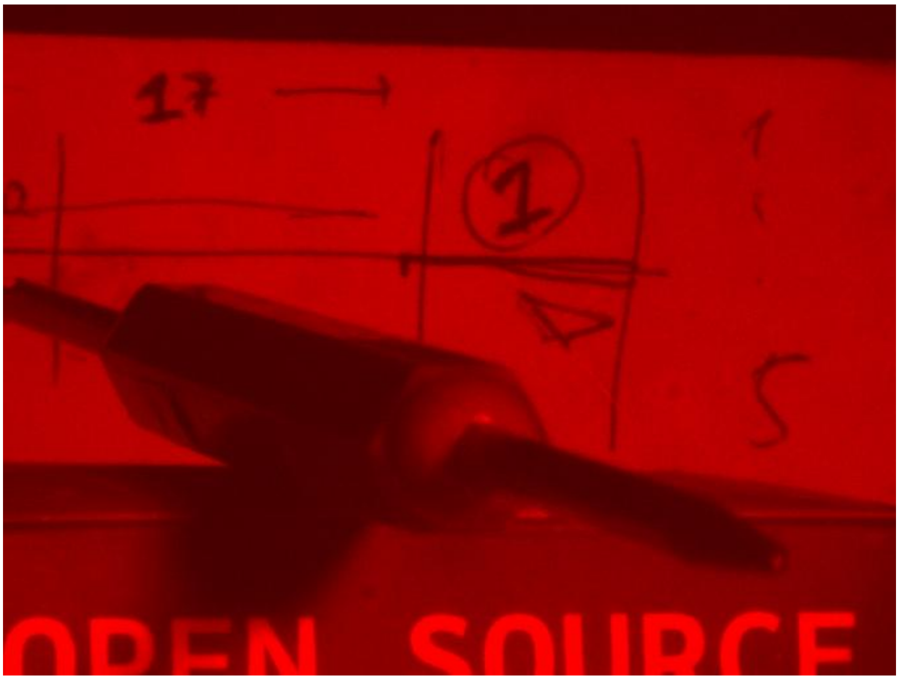

Figure 49. Red lens in front of the CCD

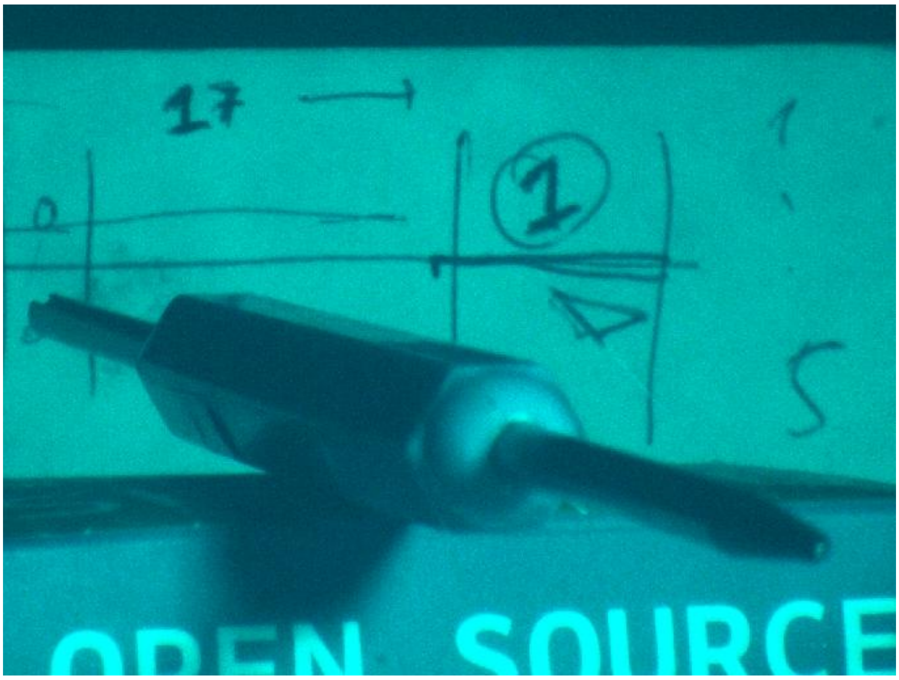

Figure 50. Cyan lens in front of the CCD

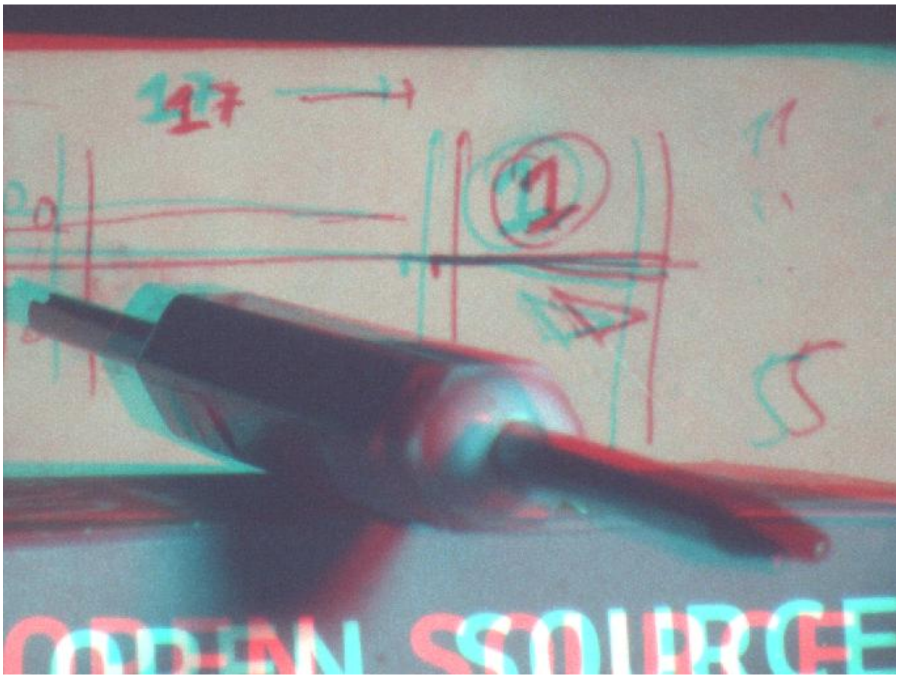

Figure 51. A “constructed” anaglyph image

The real 3D image from the two pictures without placing in front of the CCDs the red and the cyan lens of the red-cyan glasses is the following.

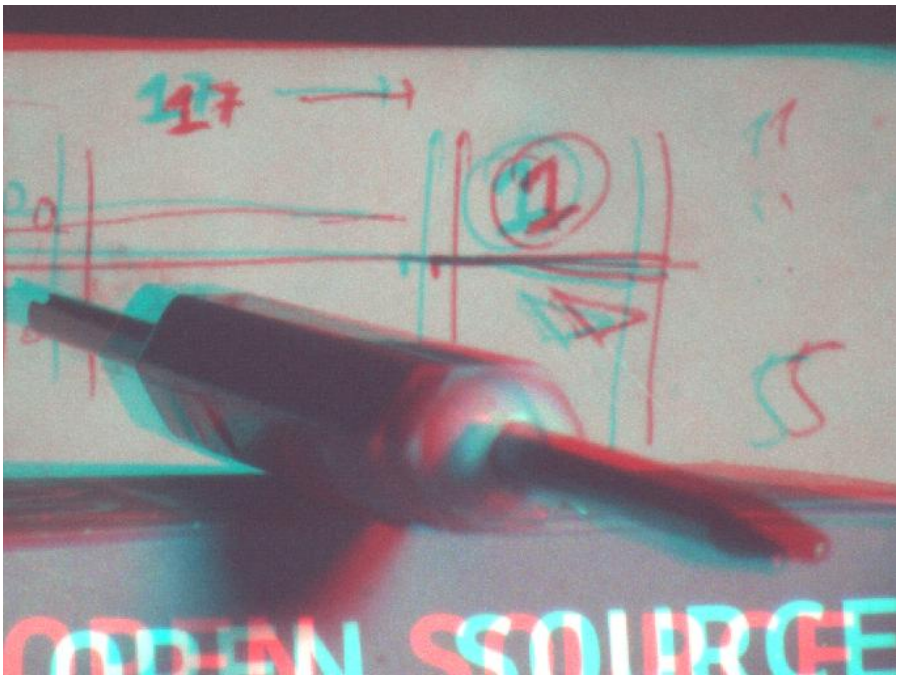

Figure 52. Anagplyph 3D image

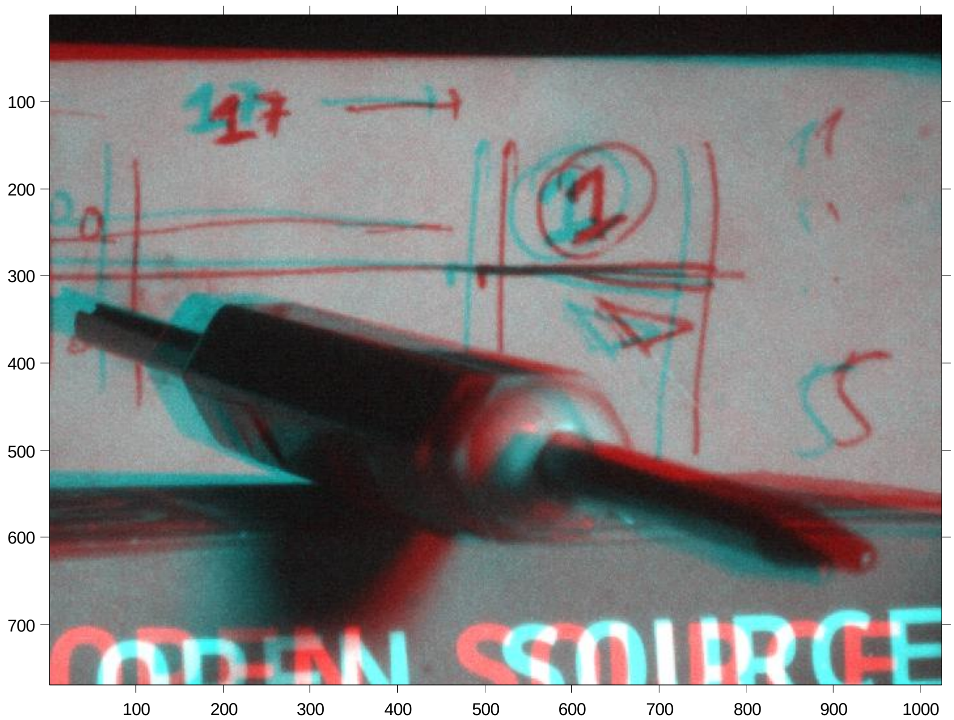

Figure 53. 3D image using triple band pass filters.

Chapter 5: Conclusions – Future Work

In this thesis we presented a real time stereoscopic imaging system for guiding surgery. We showed that there is a large variation between individuals about which factor is considered more important in the evaluation of visual quality of 3D images and 3D video. The most important thing about the 2D images is that the two viewpoints should be on the same horizontal line and not too far apart, or the brain cannot fuse the stereo pair. Also, you might think that the cameras should be the same distance apart as human eyes (65 mm adult average) but that is not essential for a 3D impression.

As far as the future work is concerned significant improvement in the stereoscopic effect will take place if we combine two images of the same scene focused at different depths (focused near and far). In addition, 3D optical microscopy is another promising area that we can use the system.

References

[1] Bradski, Gary and Kaehler, Adrian. Learning OpenCV: Computer Vision with the OpenCV Library.

[2] Howard, I. P.; Rogers, B. J. Binocular vision and stereopsis. 1995

[3] Howard, I. P.; Rogers, B. J. Perceiving in Depth. Volume 3. 2012

[4] Barry, Susan. Fixing My Gaze: A Scientist's Journey into Seeing in Three Dimensions. 2009

[5] Dhanraj Vishwanath. Toward a new theory of stereopsis. 2014

[6] Qian, N., Binocular Disparity and the Perception of Depth, 1997.

[7] Linda G. Shapiro and George C. Stockman. Computer Vision. 2001

[8] S. T. Barnard and M. A. Fischler, Computational Stereo, Computing Surveys, 2002.

[9] Y. Nishimoto and Y. Shirai, A feature-based stereo model using small disparties, 2007.

[10] W. Teoh and X. D. Zhang, An inexpensive stereoscopic vision system for robots. Proc. Int. Conf. Robotics, 2004.

[11] A. Goshtasby and W. A. Gruver, Design of a Single-Lens Stereo Camera Sstem, Pattern Recognition, 1993.

[12] S. Nene and S. Nayar, Stereo with Mirrors, 1998.

[13] Doo Hyun Lee, In S Kweon, Roberto Cipolla, A Biprism-Stereo Camera System

[14] D. Scharstein. View Synthesis Using Stereo Vision 2000.

[15] D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. 1998.

[16] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. 2001.

[17] C. L. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. 2000.

[18] S. Birchfield and C. Tomasi. Depth discontinuities by pixelto-pixel stereo. 1998.

[19] E. Dubois, A projection method to generate anaglyph stereo images, 2001,

[20] D.F. McAllister, Y. Zhou, S. Sullivan, Methods for computing color anaglyphs, Stereoscopic Displays and Applications, 2010.

[21] A.J.Woods, T.Rourke, Ghosting in anaglyphic stereoscopic images, Stereoscopic Displays and Virtual Reality Systems 2004.

[22] I. Ideses, L. Yaroslavsky, Three methods that improve the visual quality of colour anaglyphs. 2005

[23] W. Alkhadour, et al., Creating a Color Anaglyph from a Pseudo-Stereo Pair of Images, 2009.

[24] A.J.Chang, et al., Ghosting reduction method for color anaglyphs, 2008

[25] V. Tran, New Methods for Rendering Anaglyphic Stereoscopic Images on CRT Displays and Photo-Quality Inkjet Printers, 2005.

[26] H. Sanftmann, D.Weiskopf, Anaglyph stereo without ghosting, 2011