UNIVERSITY OF DUBLIN, TRINITY COLLEGE

Monitoring 3D camera rigs for film production

by

Guan Qun Chen

A thesis submitted in partial fulfillment for the degree of Master of Science in Computer Science

in the
School of Computer Science and Statistics
Department of Computer Science

September 2010

Declaration

I declare that the work described in this dissertation is, except where otherwise stated, entirely my own work and has not been submitted as an exercise for a degree at this or any other university.

Signed: _____

Date: _____

Permission to lend and/or copy

I agree that Trinity College Library may lend or copy this dissertation upon request.

Signed: _____

Date: _____

UNIVERSITY OF DUBLIN, TRINITY COLLEGE

Abstract

School of Computer Science and Statistics
Department of Computer Science

Master of Science in Computer Science

by Guan Qun Chen

Filming in 3D has gained a tremendous momentum over the last year. The use of two cameras rigged side by side instead of one has brought a number of challenges to movie production. The movie industry has been actively looking at adjusting the tools of its traditional production pipeline to make shooting 3D movies no more complicated than shooting in 2D.

A stereo diagnostic system was developed for helping camera operators and production staffs to prevent avoidable mistakes in the stereoscopic 3D movie production. A feature keypoint based analysis estimates the two camera images in order to detect the colour disparity, aperture synchronization and horizontal level of the front parallel stereo camera setup. The stereo corresponding feature keypoins were located and extracted by using the scale invariant feature transformation (SIFT). The colour disparity was analyzed in two different colour spaces RGB and CIELAB with each feature keypoint. Three different methods were implemented to analyze the aperture synchronization. The camera horizontal level diagnostic tool was an additional function for detecting front parallel stereo camera setup only. Experiment results show that the colour difference detection system perform better in the CIELAB colour space with the corresponding feature point. The aperture synchronization detection produce better result without helping of the SIFT feature keypoint detection. From the ROC experiment we can estimate the threshold value for the each diagnostic tool parameter. Being able to prevent these unbalanced photometry problems greatly helps camera crew in the stereoscopic 3D movie production.

Acknowledgements

First I would like to thank my family for all their support, friends, all my classmates and every people that helped me in the departments of Electronic Engineering and Computer Science over the last year.

I would like to thank Yue Wang, Paul Flanagan, Andrew Scott, Peng Gao, Stefan Weber, Sofiane Yous to participate in my experiments.

I would like to thank my project supervisor Prof. Anil Kokaram, for giving me the freedom to think and to solve problems independently, for his understanding, encouragement and patience.

Most of all, I would like to thank Dr. Francois Pitie, for the incredible amount of help he has given me. Without his support, advice and technical insight this project would not have been possible.

Contents

List of Figures

To my grandfather...

Chapter 1

Introduction

1.1 Purpose of Research

Filming in 3D has gained a tremendous momentum over the last year. The use of two cameras rigged side by side instead of one has brought a number of challenges to film production. The movie industry has been actively looking at adjusting the tools of its traditional production pipeline to make shooting 3D movies no more complicated than shooting in 2D.

However, some fundamental differences remain. For instance shooting in 3D requires a dedicated calibration of the camera rig to control the 3D effect. In addition of setting up the f-stop, exposure levels and focal distance, it is now essential before each scene to adjust 3D specific parameters such as the interocular distance between both cameras and the convergence angle. In practice, a dedicated technician, called a stereographer, is present on set to adjust the camera rig before each shot. The role of the stereographer is also to check that all objects in the scene are within some acceptable distance range to the camera. This is essential as any object that is too close to the cameras will give eye strain to the viewers.

While the role of stereographers is irreplaceable in checking the 3D composition of picture, a stereo diagnostic system is designed for automatically morning some of the simple but tedious aspects of the camera setup. For instance, the colour difference of both cameras could be automatically calculated, stereo calibration techniques could be used to check whether the cameras are properly aligned or whether the one of the stereo image is out of focus. Any sort of real-time feedback about these problems would be of great help for the camera crew and would prevent avoidable mistakes.

The primary object of this diagnostic system was to combine various types of image processing methods and algorithm together; to make it be able to help the camera crews, to detect unnoticeable problem before they starting to shoot the movie. This type of diagnostic tools set is becoming more and more popular with the fast growing stereoscopic filming industry.

To complete this diagnostic tools kit, it was necessary to become familiar with several image processing methods and algorithm such as, Scale Invariant Feature Transform, various types of colour space, Euclidean distance, Histogram, average gradient and Matlab image process tools. After us familiar with these image processing methods, the different possible techniques had been explored to combine them together and seek an appropriate way to implement the diagnostic system for using in the real situation.

1.2 Stereo vision

In the last decades stereo vision has been one of the most studied tasks of computer vision and many proposals have been made in literature on this topic. The purpose of this chapter is to discuss prevalent stereo vision systems developed for stereoscopic creation and representation.

1.2.1 A little history of the Stereoscopic

In the 1850, Frenchman Joseph D'Almeida discovered anaglyptic 3D by using the red/-green filters to implement colour separation. The first 3D anaglyptic film was created by William Friese-Green in 1889, which first went on show to the public in 1893. These anaglyptic films became extremely popular during the 1920s. The films used a single film with the green image emulsion on one side of the film and the red image emulsion on the other. Traditional stereoscopic photography uses a pair of two dimensional photo graphs to create a 3D illusion.

3D movies have existed in some form since 1890. In 1950s, 3D movies were played in some American theatre with cheap red-and-cyan glasses, this was a golden time for 3D movies, and in the 1980s it became popular again. As the IMAX theatres which are especially built for high quality 3D movies developed worldwide and Disney themed-venues 3D movies filmed, 3D movies became more and more successful from 2000 to 2009. Since the significant success of 3D movie of Avatar in December 2009, 3D movie and 3D imaging technology are leading the trends of future movie market.

1.2.2 Human Three Dimension Perception

The way the brain works to interpret images plays a role in the difference between high and low resolution. Gestalt psychologists explain that the brain perceives dots that are very close together as a single, continuous image. Monitors and printers exploit this phenomenon to render images [1]. There is a certain disparity through the human eyes to see the object, and the brain is using the discrepancy between the two images captured from left and right eye, to assess depth and distance. In real life, our eyes see everything in three dimensional, not a flat image. When the scientists found that the characteristics of the human eye, there is a subsequent three movie technology, through the eyes of three-dimensional human eye's visual difference will be taken after the synthesis of overlapping images, reduced to a depth of three-dimensional and multi-layered three-dimensional effect reached an unprecedented look and real feeling.

For best understanding, a simple example is the easiest way to explain the human perception. Hold up a pencil in front of your face, focus on far side background and by closing each eye in turn, and see how difference the pencil is positioned in front of you:

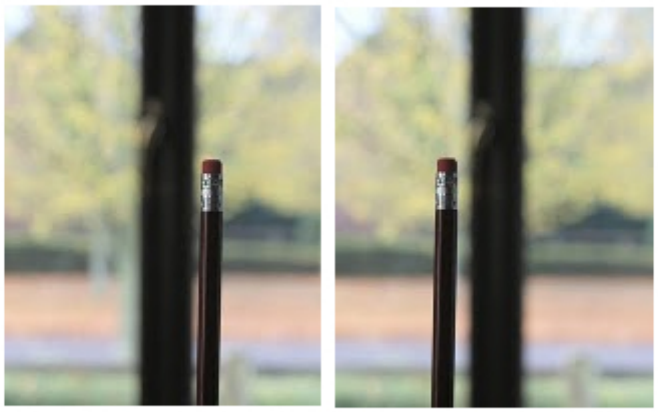

FIGURE 1.1: Image from left eye Image from right eye

From the figure 1.1 above that when we look at a specific object, then the two eyes to the image is not the same, each eye can see different parts of the object and the overlapping part, thus creating three-dimensional sense. If we focus on the pencil, you will notice that the images of the background are shifted from each other. From the figure 1.2, conversely, if we focus on the background, then the pencil appears double.

Our brain can roughly calculate the distance between the pencil and background to determine which one is close to you, which one is far away from you.

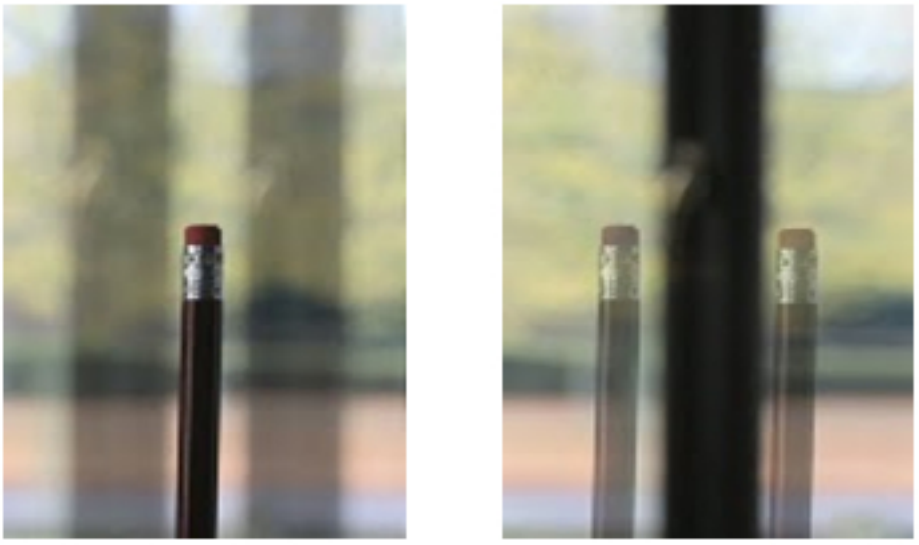

FIGURE 1.2: focus on the pencil focus on the far side

Human vision system uses several different cues to determine relative depths in a perceived scene [2].

Some of these cues are:
•Stereopsis
•Accommodation of the eyeball (eyeball focus)
•Occlusion of one object by another
•Subtended visual angle of an object of known size
•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 presently used in traditional two-dimensional images such as paintings, photographs and television. Stereoscopy imaging is the enhancement of the illusion of depth in a photograph, movie, or other two-dimensional image by presenting a slightly different image to each eye [2].

1.2.3 Three-Dimensional(3D)Imaging Technology

3D imaging which also named as stereoscopic 3D image, which is a method to re-create the illusion of depth in an image or record three-dimensional visual information. In the brain the easiest way to enhance depth perception is to provide the eyes of the viewer with two slightly different images, but representing the same object, with a minor deviation exactly equal to the perspectives that both eyes naturally receive in binocular vision. Trying to avoid the eyestrain and distortion, the image from left eye and the right eye should be presented, so that any object at infinite distance seen by the viewer could perceive while it is oriented 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.

1.2.4 Imaging Methods

FIGURE 1.3: Fujifilm FinePix Real 3D W1.

For the stereoscopic 3D movie production, it's just like human vision system, to produce stereo 3D movies; the images are required to be shot by special 3D cameras or camera rigs. At the creation stage, the left and right views are normally captured by two cameras

in a stereoscopic setup. Several other acquisition techniques exist. One of the methods requires two cameras simultaneously take a one-shot, or take a photo by a stereo camera such as the Fujifilm FinePix Real 3D W1. It also can use single camera takes two photos at different positions, but only works on still object. There are several professional level stereoscopic movie camera rigs:

FIGURE 1.4: Professional level stereoscopic 3D movie camera rigs.

1.2.5 Viewing Method

3D cinema is like to create a portal or a window can be seen through the objects fly out of or go deep insight the screen, these all rely on the 3D projectors and 3D glasses technology. A pair of slightly different images is synchronously projected onto the screen with two filmstrips by two projectors. These two images are blurry if you watch them with your naked eyes. To achieve a stereo effect, a polarizing lens, serving as a polarizer, need to be installed before each projector to produce orthogonal polarization directions and lights. These two beams of polarized lights are projected superimposed onto the screen and reflected to the audience, still keeping the same polarization directions. When the audience watch through their polarized glasses, each eye only sees the respective polarized light images, i.e., the left eye only sees the images from the left projector and vice versa. After that, the eyes converge the left- and the right-eye images on the retinas to form a stereoscopic 3D effect with the help of the brain nerves. Thus continuous motion pictures are presented to the audience, giving them a strong feeling that they were on the scene themselves with objects and views jumping onto them or imbedded inside the screen. And this is the principle of stereoscopic films.

There are two categories of 3 glasses technology, active and passive. Active glasses have electronics which interact with a display, such as liquid crystal shutter glasses and display glasses. One of the passive 3D glasses use a method called complementary colour

anaglyphs. This works by using a filter to block certain colours from each eye. The most commonly seen glasses utilise a red and cyan lens to pass red to one eye and blue and green to the others. Polarization is another method of displaying 3D content via passive glasses. It works by using lenses that block certain wavelength of visible light. For example, linear polarized glasses use vertical polarization on one lens and horizontal polarization on the other.

In general, there are other 3D viewing methods such as, Freeviewing.Obviously Freeviewing is viewing a 3D image without 3D glasses. Freeviewing includes two methods: the parallel view method and the cross-eyed view method [3].

1.3 Statement of Problem

Although stereoscopic technology helps 3D movies developed and brings human a lot of joy of viewing, it has some problems as well. There are researches prove stereoscopic 3D images could affect people's health[4]. The causes of visual fatigue from stereoscopic technology are diverse, such as the distance between the target object, the foreground and the background on the scene, the magnitude of parallax image of parallax images displayed, Convergence angle and Stereo position, which are all prevailing in the aided and non-aided stereoscopic display devices. Multiple and binocular viewpoint display may also generate distortion of stereo vision and cause visual fatigue. Watching unaligned 3D images may cause moderate eye fatigue, feel heavy in eyes, smart eyes, and difficulty in focusing distance object, headache and nausea. 3D images course more serious problems than 2D images..

An ideal stereo camera rig composes of two identical cameras mounted in adjustable rig, and separated by an interocular distance.however in fact as the cameras can be exactly identical and can never be mounted perfectly in parallel. Misalignment of the cameras and discrepancies between the internal parameters of the camera give rise to distortions at the creation stage. If stereoscopic content is being transmitted across a channel, additional distortions may be introduced due to common compression artefacts such as blocking, blurring and ringing, or due to packet loss occurring during transmission.

Also the stereoscopic rig requires that a reliable extraction of certain features such as edges or points from both images, also it is necessary to match the corresponding features between images. This method has some disadvantages, such as it can be hard to find and locate accurate features in each image. So that when using cameras to take images, there are several important considerations: how close the object to the camera, how far the image will be viewed from and physically what distance is Object in the scene from

the two cameras. The distance from the image which people are intending to view to the cameras will result in the depth of image. More images will pop out from the screen, when a person gets further from the screen, vice versa the flatter the image will appear on the screen, when the closer a person is to the screen.

Finally, distortions may also be introduced at the restitution stage, caused by artefacts which depend on the stereoscopic display technology as well as the actual scene being displayed. One example of such a distortion is ghosting, the phenomenon whereby one eye can see part of the image from the view intended for the other eye [5].

1.3.1 Monitoring equipment limitation

For the stereoscopic camera rig, the several settings must pay attention such as the image resolution, white balance, contrast, and aperture settings all have to be the same from the both left and right camera; otherwise it may cause visual uncomforted or visual fatigue. The best way to eliminate these problems is to use two identical cameras same camera brand, same model and even has the same branch number. However the things do not always turn out the way you want, even the cameras are the same in every aspect, there are always some unnoticeable problems.

The causes of visual fatigue from stereoscopic technology are diverse, such as the distance between the target object, the foreground and the background on the scene, the magnitude of parallax image of parallax images displayed, Convergence angle and Stereo position, which are all prevailing in the aided and non-aided stereoscopic display devices. Multiple and binocular viewpoint display may also generate distortion of stereo vision and cause visual fatigue.

What's the point to detect the unnoticeable problem? Films are not always shot under the perfect light condition such as in the studio, outdoor scenes are always indispensable. The unnoticeable image discrepancies are becoming significant problems, because of the poor performance of LCD monitors under the sunlight condition.The monitors are not able to provide high fidelity reproduction of the scenes. It causes the picture discrepancies are not easily to catch by the camera crews.

1.3.2 Colour balance

For a normal digital camera consumer, getting the right colour in the photo is one of the easiest things to do, just click the shutter button, if you always satisfied with the automatic settings on the camera. It can provide the quality photos that almost close to what user wants. For a professional photographer or a film maker, trying to get the

prefect colour that precisely meets their desire, it can be almost the hardest thing in the photography. However in the stereoscopic film industry, it can be even more difficult. The colour of the two pictures are not only required to meet the photographers desire, but also required matching each other from the side by side stereo camera rig. The best way to avoid this problem is to use two identical cameras same camera brand, same model and even has the same branch number. However even the cameras are the same in every aspect, there are always some external problems such as, the manufacturing process of CCD may varies, component outsourcing, software microcode may change over the production of a given model, and also the human error.

Colour difference can be caused by several factors, for example the two cameras set differently polarised, or maybe the slight differences between the physical characteristics such as the scratch marks on the camera lens or the dust entering the camera body laying on the image sensor. Colour difference directly affects the quality of the stereoscopic image. Not only it reduces the 3D effect, but also it can cause the viewer uncomfortable to enjoy the 3D movie. Correcting the colour difference onsite is a time consuming process and it also requires considerable experienced stereographer. The diagnostic system was designed for solving these problems; it is able to help less experienced camera crew to disclosure the problem. It has been specifically designed to detect the colour differences that are present between two stereoscopic images

1.3.3 Stereo camera rigs misalignment

The binocular stereo imaging system of human requires that following certain rules of cameras alignment must be obeyed to ensure the quality of stereo images.

Effective Camera Distance (as abbreviated as ECD) is the distance between the left and right cameras. This distance should be equal to 1/30 of TD, i.e., ECD=1/30*TD, based on experiences. The camera points to an infinite distance and the TD is the boundary. If the object lies just on the boundary, then it is located at the same position on the images shot by the left and right cameras, so there is no parallax; if this object is in front of the boundary, then it has negative parallax, and vice versa, it has positive parallax. All the objects with positive parallax will appear behind the screen when projected, and the ones with negative parallax will appear in front of the screen. The objects behind the screen look more comfortable than those in front of the screen. Experiences have shown that if objects are placed in front of the screen for a long time, it will lead to discomfort, like dizziness and nausea, so it should be avoided.

Target Distance (abbreviated as TD) is the distance between the object and your camera, also referring to the position of the focus point. ECD: between 1/2 TD to 3/2 TD, i.e.,

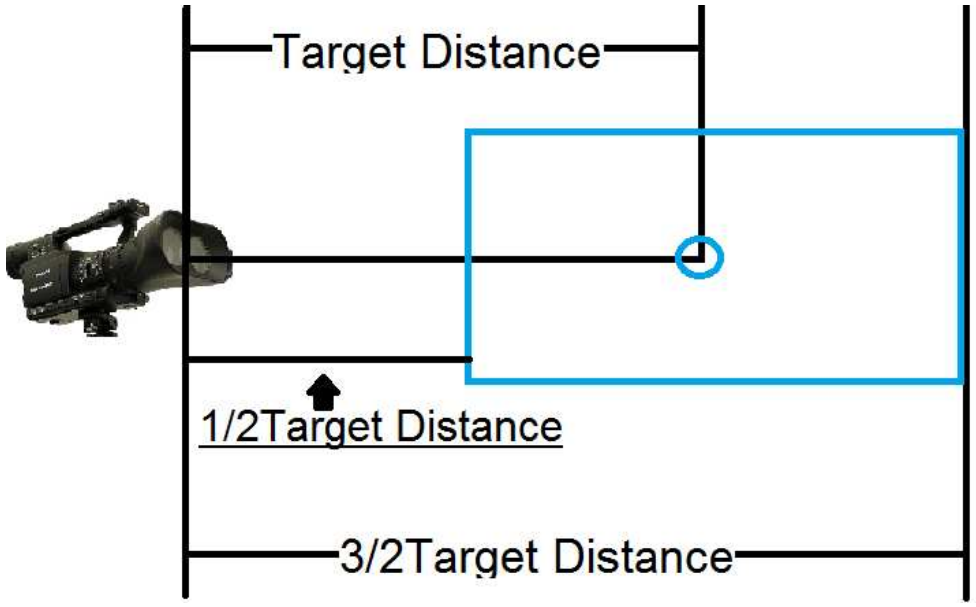

FIGURE 1.5: figure illsutrtes the Effective Camera Distance

1/2*TD < ECD< 3/2*TD. It is not hard to imagine that when one object (e.g. your nose) is very close to your eyes, then you have to try hard to turn your eye balls in order to focus on this object, known as cross eyes that will lead to discomfort of the brain.

Max Image Separation (abbreviated as MPS) is the horizontal distance between the left and right images of the same object when viewing stereo effect. The maximum effective value of this distance equals to 1/30 of viewing Distance (abbreviated as VD), i.e. MPS=1/30*VD.

1.3.4 Limitation of Stereoscopic 3-D cinema

For the 3-D cinema not only require special project, but also it has several problem need to solve: such as resolution, focus and color balance. In the 3-D cinema, when any pair of 2-D images lost detail of light or shadow, or any kind of low resolution image will seriously affect 3-D reconstruction. The audiences not only lose their 3-D sense, but also they feel painful to watch.

There is a certain length range of limits in the 3-D cinema that people cannot sense of 3-D information from screen. The normal reading range about 30 centimeters, the angle between the left and right eye is 12 in this condition, the eyes are not easy to get fatigue, vision clarity and feel comfortable. And in this comfortable position the two eyes have the biggest different image so this is also the strongest three-dimensional sense position. In the human nature, when people are looking at object far away, the perspective of the

eyes is getting smaller, 3-D sense is getting weaker. Vice versa when people are looking at something too close to them, it only takes a while people start to feel headache. This limit also happens in the cinemas, usually in range from 100 m to 200 m. The range of the limits has been defined as depth resolution or depth range.

It is necessary to seek solutions to solve these issues of viewing comfort in the stereo 3D movie post-production. One important reason for solving these problems that seriously affects 3D depth cues. There are some many golden rules for creating quality 3D contents, unfortunately, only small amount of specialized people know, not many people understand these rules from the outside of the 3D community. This reason directly affects the 3D movie production for those less experience operators, such as computer graphic artists, designers and even the people outside of movie industry like doctors and biologists. Although it is almost impossible to ask every people to master these rules of 3D in a short period time, an elaborate designed assistance system may help less experience user to work with 3D content.

Chapter 2

Literature Review

2.1 Related work

Many works have been done on the correction stereoscopic disparity from the camera geometric, camera position as a source of information. Not much work has been done with detecting colour difference based on corresponding feature between the image pairs and detecting aperture settings for both cameras. There are some of the research groups addressing the problem related to our goal, which will be discussed in this chapter.

The colour difference detection method is described for example in Gadia [6], and Pedersen [7]. In the Hardeberg [8], it introduces a list of state of the art of image difference metrics, includes the CIELAB and S-CIELAB. In Gadia [6], the authors propose perceptually-based spatial colour computational models, which are inspired by Retinex theory [9]. The approaches are first to apply a pre-filter to the stereo image pairs, then to perform an unsupervised spatial colour correction to each single pixel separately. Furthermore they prove the approaches can prevent the missing hues and local contrasts caused by global transformation, which cause a serious problem in the stereoscopic visualization. From Gabriele [10], the authors propose a method to inspect simple pixel value difference in the Log-Compressed OSA-UCS space.

For the aperture synchronization detection, a few methods are discussed in the Wu [11], the authors propose the idea to perform the blur measurement, which is to derive the Point spread function from the line spread function of the bur image. This method is faster and cheaper than the most of methods depend on the Fast Fourier Transform (FFT). The other blur measurement method introduces in Marziliano [12], they apply the Sobel filter to detect the edge in the target image, then set specific threshold of the gradient to remove noise, compute the local maximum and local minimum for each

corresponding edge in the image, detect the image blur by calculate the edgewidth. Finally, the blur measurement value is obtained by:

Blur measure = $\frac{Sumofalledgewidths}{Numberofedges}$

2.2 State of the art Stereoscopic 3D Assistance Systems

The purpose of this chapter is to introduce 4 state of the art assistant systems for stereoscopic 3D film production in IBC exhibition, Amsterdam from 10 — 14 September 2010 [13]. These systems are the SONY MPE-200 Multi Image Processor, the Stereoscopic Analyzer (STAN), Cel-Scope 3D Stereoscopic Analyser and Silicon Imaging SI-3D.

2.2.1 SONY MPE-200

In the first half of 2010, Sony launched multi-image processing system MPE-200 that can adjust the non-synchronism of two cameras for effective shooting and stereo film production.

Figure 2.1 MPE-200 is equipped with a high-performance micro-processor Cell Broadband Engine and calibrates very easily by calculating the resolution ratio at a high speed and displaying minor deviations like the hues and optical axis with 2 cameras. Therefore, when shooting stereo images, you no longer have to adjust while watching the images with eyes and waste much time. MPE-200 makes shooting with two cameras very easy and saves time to set up cameras parameters.

FIGURE 2.1: SONY MPE-200 multi-image processing system

Meanwhile, the minor deviation between the left and right images can also be tuned; therefore it not only helps to shoot clear stereo images with less deviation, but also supports calibrating all parameters during post-production. It has side by side, top and bottom modes, etc. to combine the stereo image pair into one output signal.

2.2.2 The Stereoscopic Analyzer (STAN)

Figure 2.2, the Stereoscopic Analyzer (STAN) is a real-time analysis and correction system for stereoscopic 3D post-production and 3D live events [14]. It is developed by the Fraunhfer Heinrich Institute, Berlin, in cooperation with KUK Film Production, Munich. It provides optimal advice to camera crew and production staff to create excellent quality of stereo 3D images. The STAN is combination of software and hardware stereoscopic 3D assistance system, it is capable of capturing and analyzing stereoscopic 3D contents in real-time. It provides feature-based scene analysis to match the corresponding feature points on the both left and right video source to prevent the stereo camera misalignment; the stereo baseline is monitored all the time by using the actuators. It also detects the colour disparity and camera geometric. Not only these functions, but also the optimal Inter-axial distance is produced by using the near and far clipping plane.

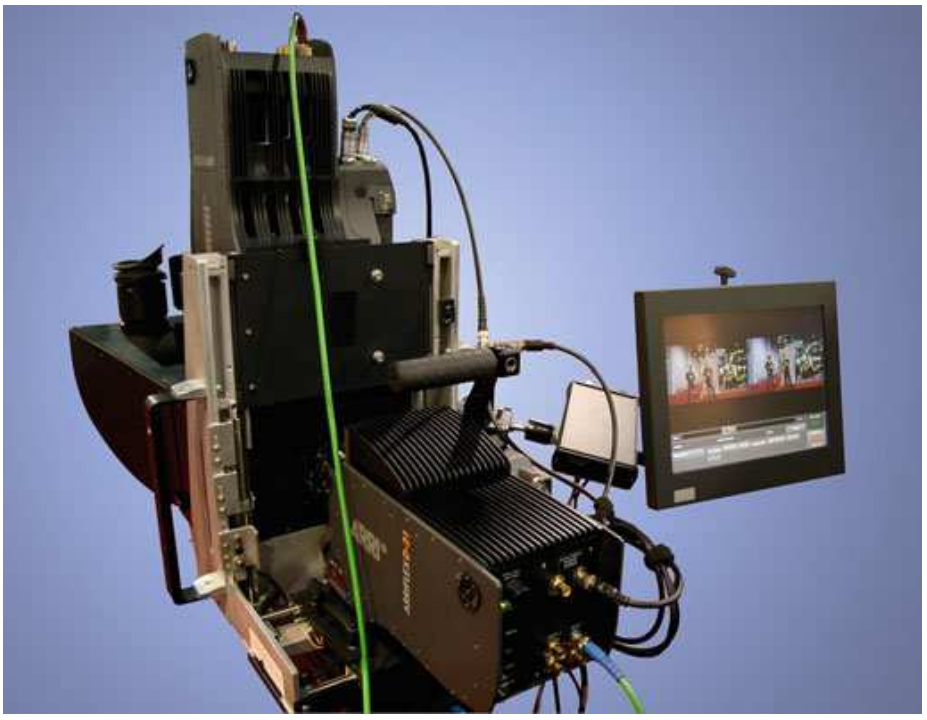

FIGURE 2.2: Stereoscopic Analyzer attached to a mirror stereo rig with two ARRIFLEX D-21 camera

2.2.3 Cel-Scope 3D Stereoscopic Analyser

Cel-Scope 3D is another stereoscopic 3D post-production monitoring system. It is developed by Cel-soft. It is a complete software solution and monitor movie production at low cost. It only requires a Microsoft Windows based PC with sets of video capture devices. Although it does not require any specific hardware and expensive equipment, it

monitors every possible aspect of the stereo 3D movie production, such as stereo camera alignment diagnostic, colour disparity and depth budget. A depth histogram is used to analysis the depth disparity pixel range, in order to monitor full range of values during the production. The stereo camera misalignment is monitored by the vertical disparities histogram. Cel-Scope3D supports multiple colour space to detect the colour disparity. In the camera synchronization, it intends to use genlock from a pair of image input device. The Cel-Scope3D is a low cost fully diversity multi-function software solution for the stereoscopic 3D movie post-production.

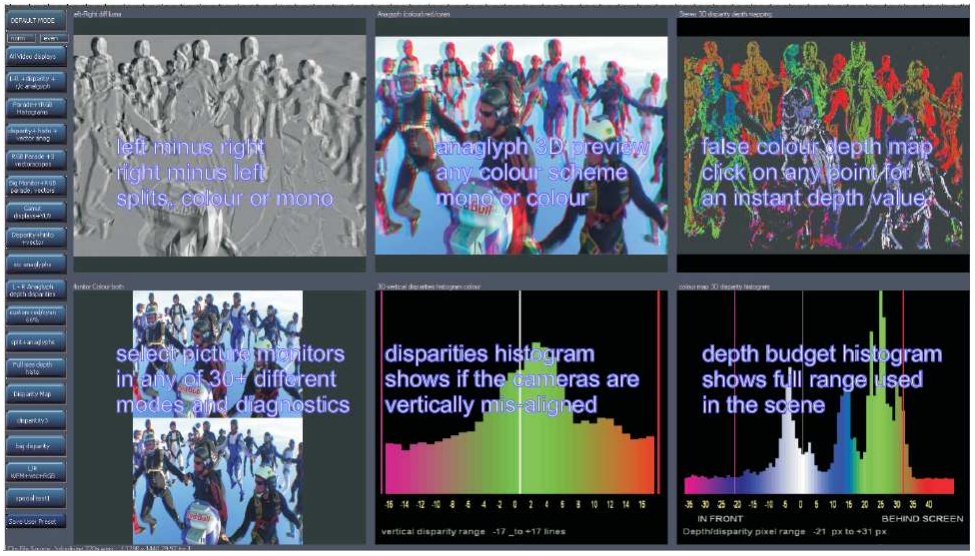

FIGURE 2.3: Cel-Scope3D Graphical user interface

2.2.4 Silicon Imaging SI-3D system

"The SI-3D camera system streamlines the entire stereo-3D content acquisition and post production process;" states Ari Presler, CEO of Silicon Imaging.

The SI-3D is combination solution for stereoscopic 3D movie production. It equips with two remote SI-2K Mini cameras, which are capable of recording in high definition RAW 3D video content. The SI-3D system has a remarkable rig design, it allows the camera to be setup in side by side configuration or with the beam splitter two cameras in the 90 degree angle, one camera is horizontal setup and the other one is vertical setup behind the beam splitter. The system also comes with a touch screen interface, with the state of the art software package SiliconDVR. Usually, the SI-3D system takes two camera operators; the first operator is in charge of taking the shots and controlling the frame, the next operator stands behind him to monitor and diagnostic problem frame of the Stereo 3D content. The figure 2.4 illustrates touch screen interface with SiliconDVR

FIGURE 2.4: SI-3D Touchscreen Interface

software. The SI-3D system is integrated various tools together, such as the stereo focus adjustment, 3D effects, stereo camera alignment false colour zebras, dual histograms, parallax shifts, anaglyph mixing and wiggle displays. The SI-3D system has been widely tested and used in the stereoscopic 3D movie production.

2.3 Summary

One thing in common is that these assistant systems are Jumbo pack, they have everything that you need for the stereoscopic 3D production, also they are quite expensive, even the known low cost Cel-Scope 3D system cost a lot more than normal consumers can afford. This project is designed and developed in the small scale, the state of the art stereoscopic 3D assistance systems are the reference for the diagnostic system development, some of the functionalities and algorithms are inspirited by these described assistance systems. Our diagnostic system concentrates in the colour disparity, aperture synchronization detect and camera horizontal level detection between the stereoscopic image pairs.

In the next chapter methodology, the entire diagnostic system is described in fine details and any relevant algorithms are analyzed. In the result and discussion chapter, the system performance is estimated and evaluated. The conclusion and recommend future are discussed in the final chapter.

Chapter 3

Methodology

In this chapter, we discuss the overview of the stereo diagnostic system and give an explanation of the each section of the system and algorithm used.

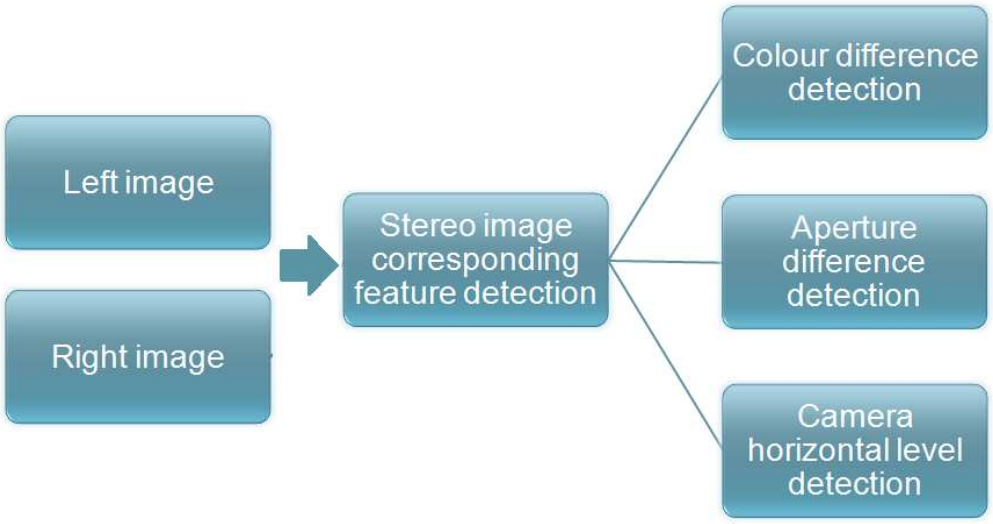

FIGURE 3.1: The diagnostic system workflow

Figure 3.1 is the diagnostic system workflow. The first process was the stereo corresponding feature kepoint detection, it was based on David Lowe's Scale Invariant Feature Transform (SIFT), it provides reliable feature matching techniques between different images, the SIFT capable to the features that are invariant to image rotation, scale and illumination. After the experiment, SIFT was proved to be a perfect solution to detect the invariant features between two stereoscopic image pair.

After the feature keypoints had been detected and located between the two stereo image pair, these feature keypoints were able to be extracted and analysis the colour information in the appropriate colour space, previously in this project the RGB colour space

was used to analysis the colour information. After fully investigation [8, 15, 16], Second colour space CIELAB was chosen to analyze colour disparity. In the result chapter, the comparison of two colour space will be discussed.

Three different sizes of image were analyzed in the aperture synchronization detection stage. The optimal method will be introduced in the result chapter. The camera horizontal level diagnostic tool was an additional function for detecting front parallel stereo camera setup only.

3.1 Scale invariant Feature Transform (SIFT)

The stereoscopic image is composed of left and right two images, although the perfect matched stereoscopic image pair is highly similar to each other, the content of the two images are not absolutely the same. So comparison of the global information is an inexpedient option, however to detect the invariant feature keypoints of the two images, then compare keypoints between each other is a more efficient method to diagnostic the difference.

The diagnostic system was based on David Lowe's Scale Invariant Feature Transform (SIFT), it is a reliable feature matching method that for detecting the distinctive invariant features from the different images. SIFT is able to detect the features that will not change with the image rotation and the scale, this makes it becomes an excellent choice for diagnostic the quality of the stereoscopic photography, and it also can combine with other toolkits to provide good results in stereo matching with the distortion such as a pair of images in different colour temperature, images taken with different aperture settings, whether the images are at the same horizontal level.

The main purpose of SIFT is to improve the Harris corner detector, which is not scale-invariant problem. To be able to detect scale-invariant feature, basically it requires searching stable features in every possible scales. However this is only in theory that practically is not achievable. SIFT uses sample scale space method with a reasonable sampling frequency that it is able to detect scale-invariant feature. Following are the major steps for feature extracting.

3.1.1 Scale-space extreme detection

To obtain the rotation invariance feature, David G. Lowe proposes that first analysis image in scale space L(x, y, σ), which is a convolution image I(x, y) with a Gaussian Kernel. Scale space is generated by using difference of Gaussian (DoG) function. There

are two main reasons for using the difference of Gaussian function, first is that DoG is an efficient and convenient filter, second the stable performance of the DoG as the same as the Laplacion of Gaussian function. DoG is very similar to the Gaussian filter; the size of kernel is controlled by the σ. The scale can be seen as the σ in DoG, where given two different scale of the Gaussian filter, after subtracting these two scales then apply the filter of the original image.

$$D(x, y, \sigma) = (g(x, y, k\sigma) - g(x, y, \sigma)) * I(x, y) \tag{3.1}$$

$$= L(x, y, k\sigma) - L(x, y, \sigma) \tag{3.2}$$

Where $g(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-x^2/2\sigma^2}$ and (x, y) is the, σ is the Scale-space factor, size of the kernel.

Figure 3.2 below is an example of applying the scale space[17]. As a result, the blurred image is obtained.

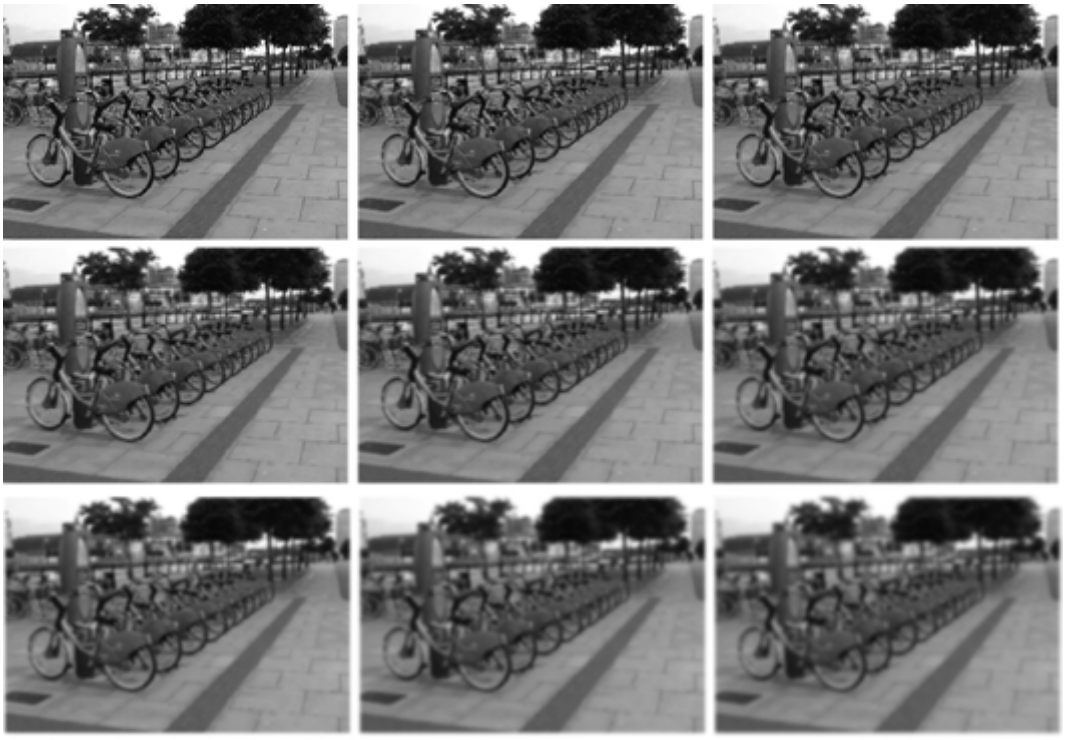

FIGURE 3.2: Scale space sample.

Figure 3.3, after doubles the scale space goes into another octave, sample rate also reduces to half; this is equivalent to 512x512⟶256x256, then the process repeat again, until the end.

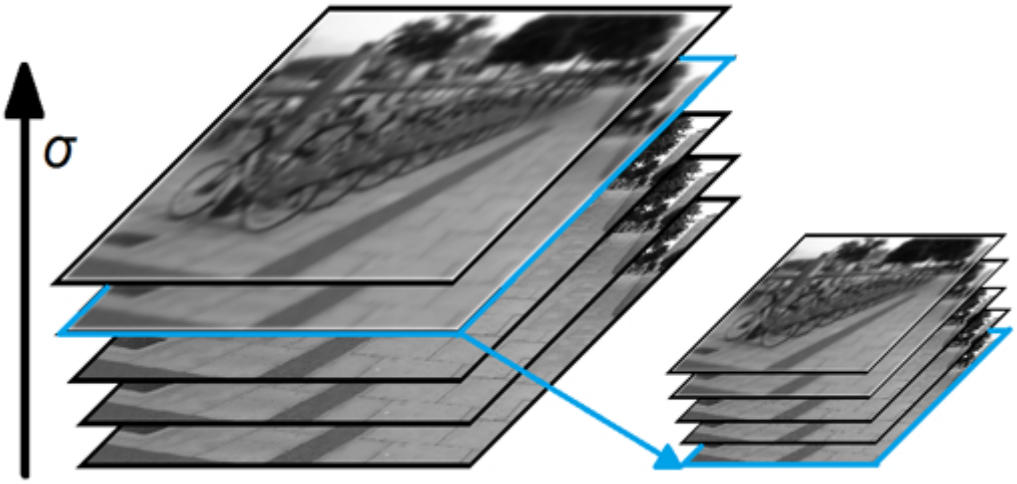

FIGURE 3.3: Two octaves of a Gaussian scale-space image pyramid with s =2 intervals. The first image in the second octave is created by down sampling the second to last image in the previous.

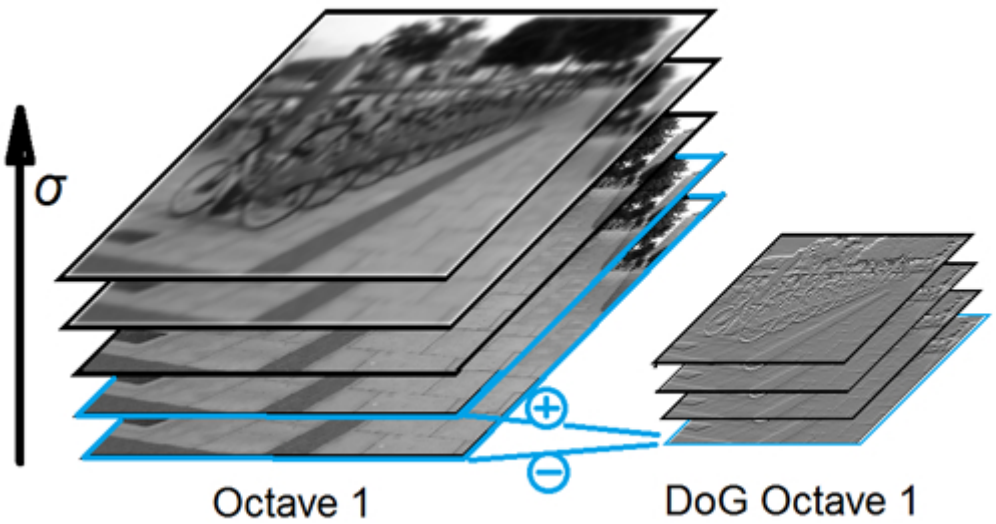

FIGURE 3.4: The difference of two adjacent intervals in the Gaussian scale-space pyramid create an interval in the difference-of-Gaussian pyramid (shown in blue).

This way is similar to create a pyramid. The frequency is double, if we applied the Laplacian function every time, the sample will be too sparse. For trying to get the desire dense sampling, each octave (O) has to be divided into several sub levels (S).The next group of octave is from sampling the last octave.

Figure 3.5, the extrma is a pixel (x, y), which is the feature candidate. For generating the extrema, the process requires comparing each pixel to its 8 neighbours in the current image, also to 18 (9x2) neighbours in scale above and below, the exrame must be the local maximum or minimum. There are at least 3 images (levels) are required for ensure that the extrema is on the both scale space and two dimensional space.

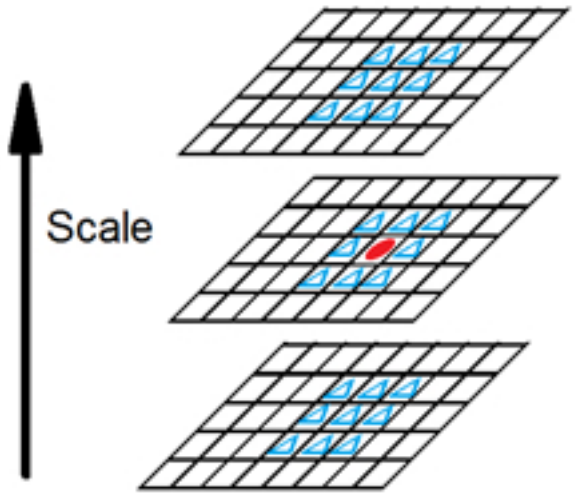

FIGURE 3.5: Extrema:the local maxima or minima value for identifying the potential interest feature.

3.1.2 Keypoint localization

After identified the potential interest feature, a filter process is required to reject the low contrast points and edge points, not every feature point is stable,due to the DoG function has a strong response along the edges, and these undesired keypoints reduce the matching accuracy and noise resistance.

For eliminating the low contrast point, a 3D quadratic function is used to determine the sub-pixel maximum. This is a Taylor expansion:

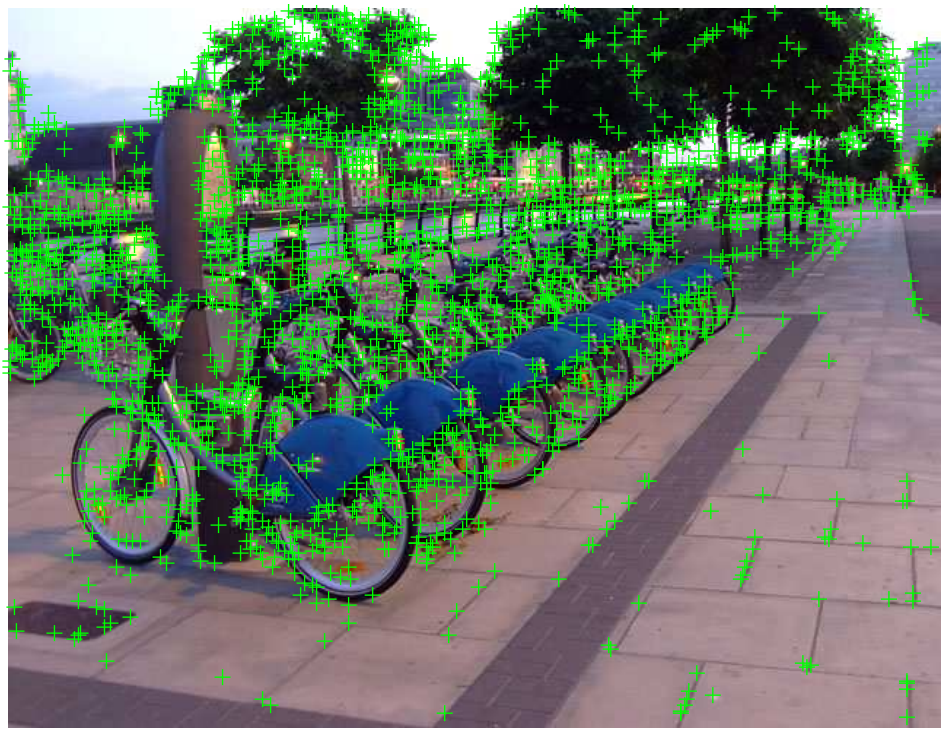

FIGURE 3.6: Sample image P(205).jpg shows too many keypoints (2969 keypoints).

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \tag{3.3}$$

$$\hat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X} \tag{3.4}$$

Where D is result of the DoG, X is the potential interest feature, and according to the Taylor expansion with D and X, offset X can be determined. This offset X can be treated as the real location of the local extrma of the sub-pixel. Then the offset X is substituted into the Taylor expansion, if the absolute value of the answer is less than 0.03 (|D(X)| <0.03), discard this keypoint as a low contrast point.

For eliminating edge responses, Duo to the DoG function has a strong edge response along the edges; an inappropriate DoG function has a large principal curvature across the edge and has a small principal curvature perpendicular to the edge. The two principal curvatures are always perpendicular to each other, they can be calculated by using a 2X2 Hessian matrix:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{3.5}$$

The principal curvature D and the eigenvalues H are proportional with each other. Assume α is the maximum eigenvalues,β is the minimum eigenvalues. Then:

$$Tr(H) = D_{xx} + d_{yy} = \alpha + \beta \tag{3.6}$$

$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \tag{3.7}$$

Assume $\alpha = \gamma\beta$,

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(\gamma\beta+\beta)^2}{\gamma\beta^2} = \frac{(\gamma+1)^2}{\gamma} \tag{3.8}$$

$\frac{(\gamma+1)^2}{\gamma}$ is at minimum when $\alpha = \beta$, and as the γ increases as the ratio increases. To check whether the ratio of the principal curvature is below a threshold value γ, only needs to check:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(\gamma+1)^2}{\gamma} \tag{3.9}$$

Use $\gamma = 10$ to discard the keypoint along the edges.[17] Figure 3.7 shows the keypoints of the sample image P (205).jpg after eliminating the low contrast keypoints and edge keypoints. Before applying the filter process, according to the Figure 3.6, there were 2969 keypoints detected, only 1007 keypoints remained, 1962 unstable keypoints rejected.

3.1.3 Orientation assignment

This step is the preparation for matching. The major orientation is required to achieve the rotation invariant detection. So that it is necessary to calculate the value and the orientation of gradient for each keypoint by using the method called the orientation histogram. In concept, the orientation histogram generates the gradient from a region around the keypoint. The majority orientation in the region is decided as the major orientation. The central point of each neighboring pixel of its weight, it is a Gaussian distribution and then multiplied by the gradient of the point size of future decisions.

FIGURE 3.7: Sample image P(205).jpg after reject low contrast and edge keypoints (1007 keypoints).

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \tag{3.10}$$

$$\theta(x,y) = \alpha \tan 2((L(x,y-1) - L(x,y-1))/(L(x+1,y) - L(x-1,y))) \tag{3.11}$$

The pixel difference is used to calculate the gradient and magnitude, where L is used for each key point in the scale space. In the actual calculations, samples around the center keypoint and uses the orientation histogram to determine gradient direction.

Orientation histogram with 36 bins, which is covered 10 degrees each. The gradient magnitude of the each sample is weighted by using the orientation histogram. In the orientation histogram, to detect the highest peaks and local peaks that is within 80

3.1.4 Keypoint descriptor

Previous process determined the keypoint location, scale and orientation. Next step is to descript the keypoints for the original image. Figure 3.9 and 3.9, the gradient magnitude and orientation around the each keypoint requires to be sampled and weighted, use the scale of the keypoint as the level of the Gaussian blur. Assign weight to magnitude by using a Guassian weighted function descriptor window.

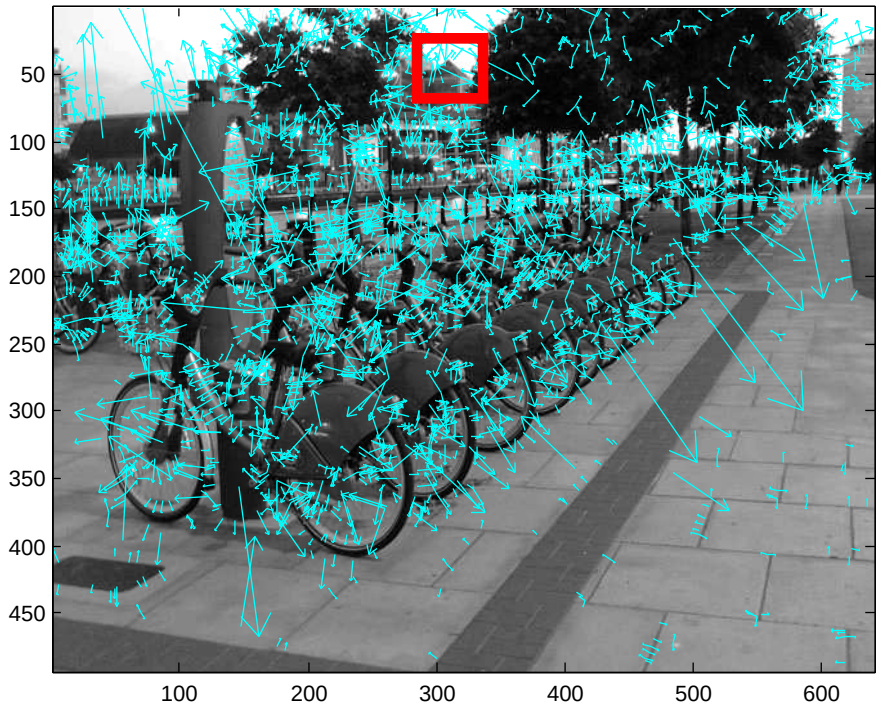

FIGURE 3.8: Orientation sample of P (205).jpg. The red box area is an orientation example in the Figure 3.8.

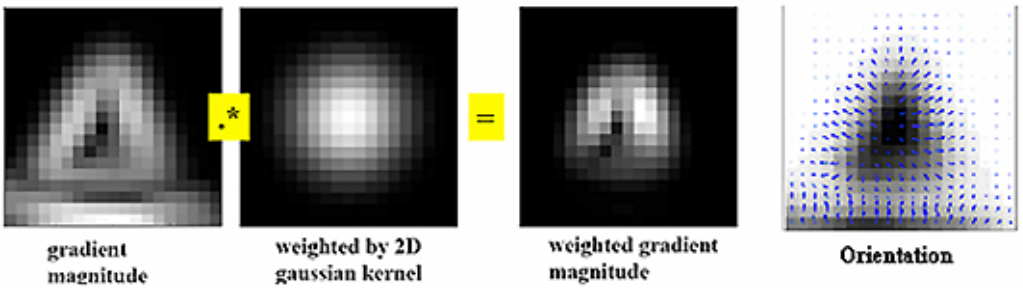

FIGURE 3.9: Orientation.

In the figure 3.11, the blue circle indicates the Gaussian weighted region. As the pixels are closer to the keypoint, they provide greater the contribution of gradient direction information. In the Figure A keypoint is described by the 2 2 of 4 sub points, each sub point with 8 direction vector information.

The reason for using the orientation histogram for feature point descriptor is because the gradient, which is more robust method to deal with illumination. When the illuminations changes on the image that is like every single pixel multiplies to a constant, this has the same affect on the gradient that multiplies to a constant. The constant will be cancelled by the normalization.

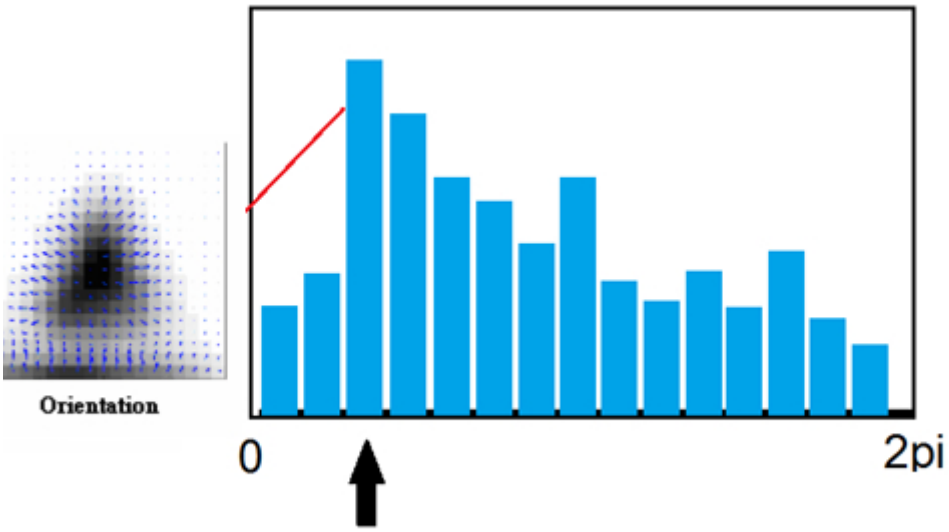

FIGURE 3.10: Orientation histogram.

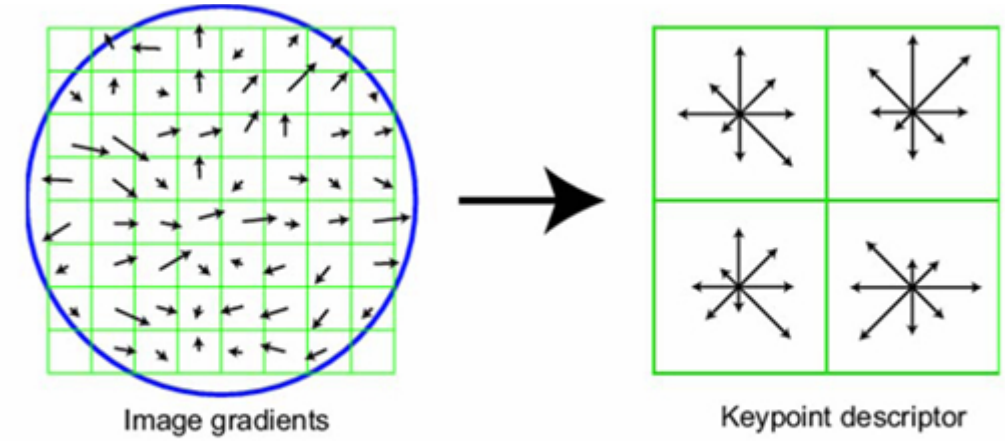

FIGURE 3.11: keypoint descriptor.

After generate the SIFT feature vector of the two images, Euclidean distance is used to determine the similarity measurement according to the location, scale and orientation of the each keypoint. The best match is found the smallest Euclidean distance of its neighbour for each keypoint.

3.2 Colour disparity detection

3.2.1 RGB colour space

From the figure 3.12, once the invariant feature keypoints was located by the Scale Invariant Feature Transform, next colour difference diagnostic process can be proceed. Numbers of small image were cropped according to these locations from the original

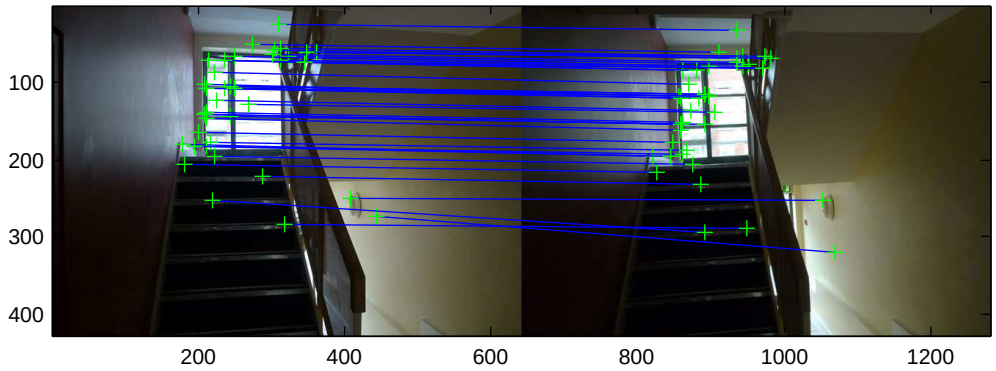

FIGURE 3.12: Image P(291).jpg and P (292).jpg SIFT match example,41 matches.The green crosses illustrate the matched SIFT descriptors with blue lines connects together.

image. The size of the image was depended on the scale of the SIFT descriptor,shows in the figure 3.13 or fixed in 50 x 50 pixels, shows in the figure 3.14. The histogram value was computed from red, green and blue channel for each cropped image. Once the histogram values of each channel were generated, the histogram intersection distance was calculated to determine the colour similarity of the two images.

In the nature, most of the visible spectrum can be mixed by different proportions of red, green and blue. The RGB colour space contents these three primary colours. It was original designed for reproducing colour in the electronic system, such as CRT monitor [18]; each colour has intensity value from 0-255. There are total 16581375 (255x255x255) colours can be reproduced on the display. Most of image devices are using RGB colour space, like TV, computers and digital cameras, also including the stereoscopic cameras. The colour in the RGB colour space can be easy described by using the colour histogram.

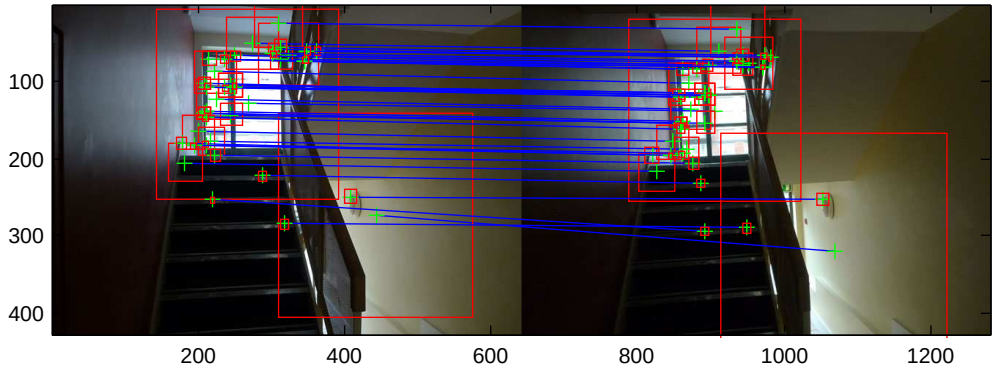

FIGURE 3.13: Image P(291).jpg and P (292).jpg SIFT match example. The red rectangle boxes illustrate the area of the image is cropped to diagnostic colour difference.The size of box is depended on the scale of the SIFT descriptor

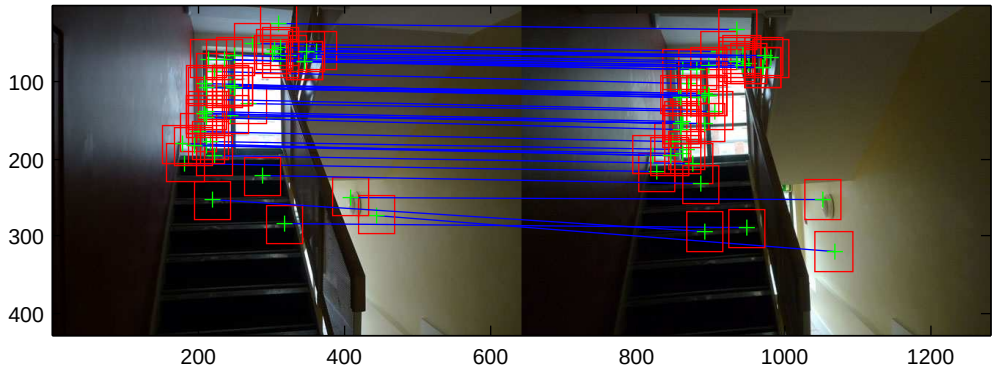

FIGURE 3.14: Image P(291).jpg and P (292).jpg SIFT match example. The red rectangle boxes illustrate the area of the image is cropped to diagnostic colour difference,the size of box is fixed by 50 x 50 pixels.

3.2.2 Colour histogram

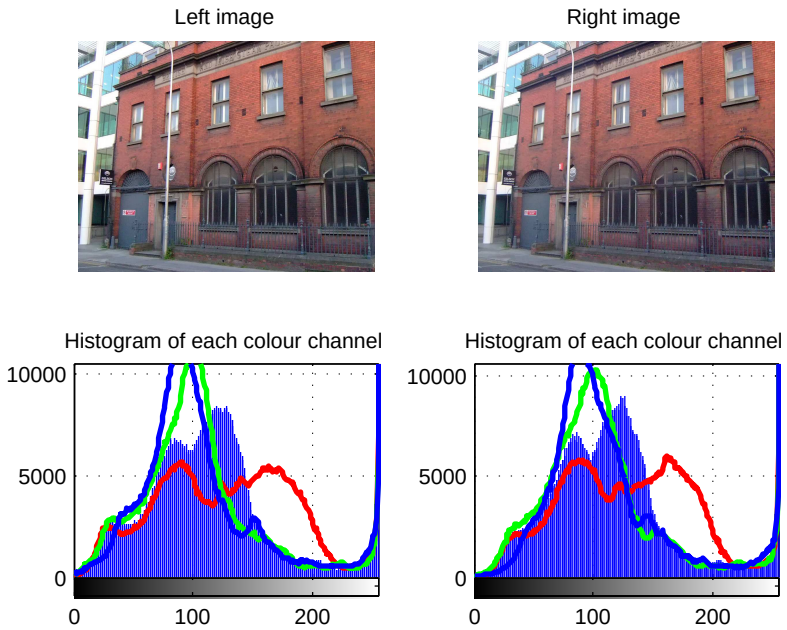

FIGURE 3.15: Colour histogram of stereo image pair sample.

The colour histogram is composed the number of pixels in each colour channel on a colour image. In the sample figure 3.15, a perfect matched stereoscopic image placed side by side. The histograms for these two images were almost the same. The large peak represented a large number of pixels had the brightness value for the gray scale image. The peak should always be in the middle, which means the image was not too dark or too bright. If the peak was located on the right side that indicated there were too any

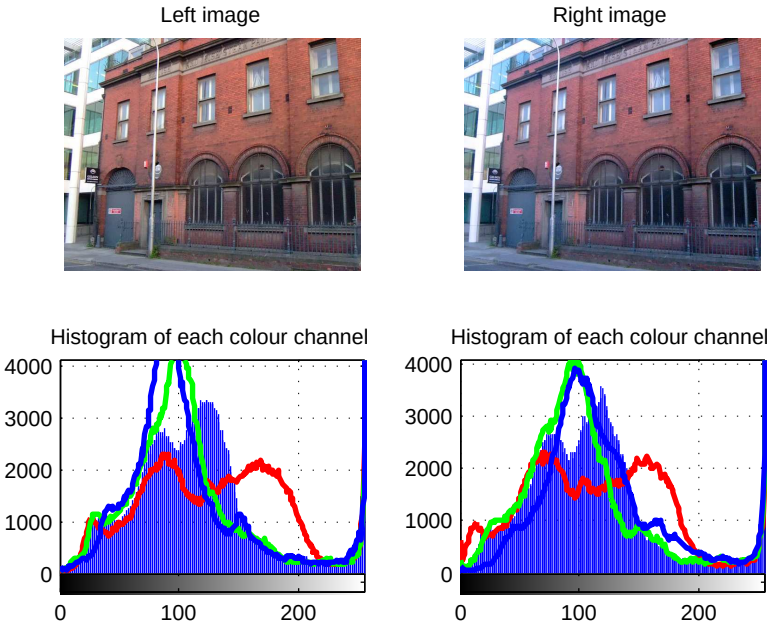

FIGURE 3.16: Colour histogram of stereo image pair sample.

pixels with those brightness values. The image would look very bright. The other three solid lines indicated the intensity values of each red, green and blue channel.

Next figure 3.16,sample stereo pair had colour difference, the right image was bluer than the left image, the histogram of blue channel was shifted to right, it was hard to describe how much difference by just look at the histrogram curve. For comparing the colour difference, a distance measurement was required to numerically indicate the difference. In the project, the histogram intersection distance was used.

3.2.3 Histogram intersection distance

The histogram intersection distance was proposed by M. J. Swain and D. H. Ballard in 1991 [19]. The histogram intersection distance of Left image (L) and Right image (R) is given by [20]:

$$d_his(R, L) = 1 - \frac{\sum_{i=0}^{N-1} min(R_i, L_i)}{min(|R|, |L|)} \tag{3.12}$$

where L and R represent the value of the each histogram. The intersection distance does not count the colour that not present in the right image, this way can reduce the background colour interference. The sum of all the bins has the minimum distance between each histogram. The result is normalized by the histogram with minimum bins.

If the colour of the two images is mostly like each other, the sums of the fewest bins are almost the same as the sum of the minimum of the two histograms, which means the histogram intersection distance is 0, if the colour of two images is dissimilar; the value of distance increases until it reaches 1, which represents the colour of two images is completely different.

FIGURE 3.17: Matched stereo image pair P (85).jpg and P (86).jpg. Modified stereo image pair P(529).jpg and P (530).jpg.

Here is an example of the histogram intersection distance.

From the figure 3.17matched stereo image pair P(85).jpg and P (86).jpg, the next pair on the bottom is the Photoshop modified stereo pair, the colour of P(530).jpg is clearly different from the P (529).jpg. The matched stereo image pair P (85).jpg and P (86).jpg The modified (applied the blue filter) stereo image pair P (529).jpg and P (530).jpg, applied the blue filter. The difference is not distinguished by using this method.

	Histogram distance	Red channel	Green channel	Blue channel
P(85-86).jpg	0.5	0.513889	0.52	0.530612
P(529-530).jpg	0.542716	0.61767	0.543253	0.625
Difference	-0.042716	-0.103781	-0.023253	-0.094388

3.2.4 CIELAB colour space

For calculating the colour difference in CIELAB colour space, the procedure was similar in the RGB colour space. After the invariant feature keypoints was located by the

Scale Invariant Feature Transform, the sample pixel was inspected, which location was according to SIFT location parameter. The colour information was extracted from each pixel in the RGB colour space. Then these colour information were converted into LAB colour space. The Euclidean distance was calculated from these LAB colour information to determine the colour similarity between the two images.

After review several papers about CIELAB [8, 15, 16], the idea of using CIELAB instead of RGB had Excited. The RGB colour space is depends on equipment; the value could be varied from the different equipment.CIE L*a*b* (CIELAB) is the most complete colour space specified by the International Commission on Illumination (Commission Internationale d'Eclairage, hence its CIEinitialism), which is completely independent from any device. It is designed for human perception. It describes all the visible colours to the human eye.

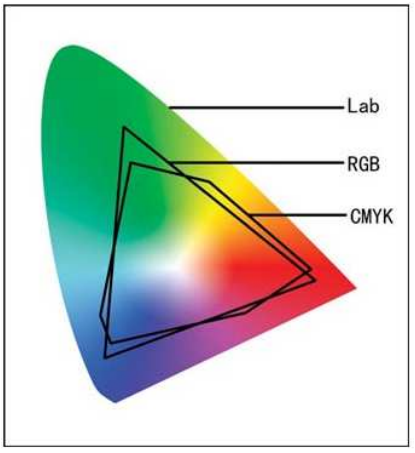

FIGURE 3.18: CIELAB RGB CMYK comparison

According to the fig above, LAB and CMYK colour is richer than RGB, because the LAB colour gamut area is greater than RGB.

3.2.5 RGB convert Lab

RGB to LAB Conversion All the test images were taken in the RGB colour space; the colour information had to be converted before performing the LAB colour difference calculation. There was no direct conversion from RGB to LAB, the XYZ colour space had to be used as an intermediate model.

RGB convert to LAB function was adopted from Mark Ruzons RGB2LAB.m Matlab function [21].The function took RGB value of red, green and blue from each channel or a single M x N x 3 matrices image, produced the image in the CIELAB color space. The

range of RGB values could be either between 0 and 1 or between 0 and 255. The LAB Values for L are in the range [0,100], a and b were approximately in the range [-110,110].

Mark Ruzons conversion was based on ITU-R Recommendation BT.709 using the D65 white point as a reference. The error in transforming RGB → Lab → RGB is approximately 10^{-5}.

3.2.6 Eulidean distance

Human eye has a good ability to identify the colour difference, but it has a very poor ability to tell how much the difference. CIELAB is designed to be a device independent model therefore it is able to accurately measure the colour difference in the small area between the target image and the original image. CIELAB is composed of three channels; the first channel L indicts the brightness. a channel a mangers colour range from red to dark green; channel b mangers colour range from blue to yellow.It is intended for equal perceptual disparity for equal difference in each channel L, a and b. The brightness can be measured and represented by ΔL between the left image (L1) and right image (L2) ΔL = L1-L2. And if the colour difference between two images are from blue turns into yellow, Δb can be used to represent, Δb = b1-b2. Δa is used to represent the colour difference from red to green, Δa = a1-a2.

The Euclidean distance is most common way to represent the distance. It examines the root of square differences between two objects. The total colour difference ΔE is indicted by Euclidean distance of the L,a and b . The formula is

$$\Delta E = \sqrt{(L1 - L2)^2 + (a1 - a2)^2 + (b1 - b2)^2} \tag{3.13}$$

Colour disparity detection system detected if the ΔE was higher than the threshold value, it would continue to calculate the individually channel for L, a, b in the CIELAB colour space. The range of L is 0-100, the value of L represents the level of brightness between the two images, if value of a is positive, which represents redder, negative represents greener than the target image. If the value of b is positive, which represents yellower, negative represents bluer than the target image.

3.3 Aperture synchronization detection

The aperture in the camera is similar to the iris in the human eye, which controls the amount of light go in through. The different aperture setting creates different level of blur on the photo, which it refers the depth of field. Large depth of field represents the most objects of the photo is in focus, picture displays sharply. Small (shallow) depth of filed represents the only part of object in the photo is in focus, and the rest is blurring.

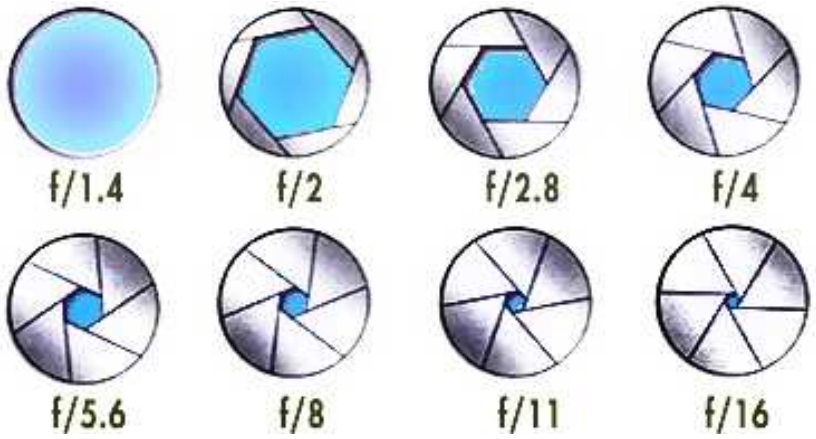

FIGURE 3.19: Illsion the different size of the aperture. The image is from TECHtata .

From the figure 3.19,pay attention to the aperture measurement, large aperture setting has a small value, so the $f/1.8$ is much larger aperture than f/22. The smaller aperture setting leads to large depth of field such as f/22. The large aperture setting leads to small depth of field such as f/1.8.Clearly shows in the figure 3.20.

FIGURE 3.20: Illsion the different size of the aperture cause different depth of field. The image is from edwinsetiawan

3.3.1 Average gradient

For the stereoscopic 3D movie if the two cameras had different aperture setting that would cause serious problem for the viewers.The aperture synchronization detection was intended to avoid the two camera rigs at different aperture setting. It was achieved by calculating the difference of the average gradient between the left and right images.

The aperture synchronization detection was achieved by calculating the average gradient of two images. Three methods were experimented in this procedure. The first method was to calculate the average gradient value of the corresponding feature keypoint in fixed size 50 x 50 pixels, the second method was to calculate in the scale based size of the corresponding feature keypoint. The last method was to compute the average gradient of the entire image, the global information. Then the average gradient value of left and right image was compared to determine different blurring between the two images. The optimal method will be discussed in the result chapter.

To calculte the average gradient, first a template image was created from the original image, and then shifted each pixel by 1 respectively on the x-axis and the y-axis. The difference of each pixel was calculated between the template image and original image.the boundery was set to 0, which did not require calculating. The average gradient of each pixel was calculated by using:

$$AVEGRAD = \frac{\sqrt{\frac{(\Delta X^2 + \Delta Y^2)}{2}}}{No.of pixels} \tag{3.14}$$

The average gradient defines as a contrast measurement in the image, which is represented as the slope of a straight line joining two density points on the sensitometric curve. It can be used sensitively to reflect the image contrast on the small details and presentation skills; it also can be used to evaluate image blurring. In the image, the direction of the gray scale of a large rate of change, it is also a large gradient. Therefore, the average gradient value can be used to measure image clarity, it also reflects the image contrast of small details and texture transformation feature.

3.4 Stereo camera alignment

3.4.1 Horizontal position calibration

Due to huge amount of work had been done in the camera geometric area. The camera horizontal level diagnostic tool was an additional function for detecting front parallel stereo camera setup only in this project.

After obtained the matched SIFT feature point pairs between the left and right image, the horizontal level of the each pair was contrasted. In the figure 3.21, according to the location on the y-axis of the feature points, the angle theta was calculated with the trigonometric function. The angle theta = $\tan^{-1}$ (a/b).

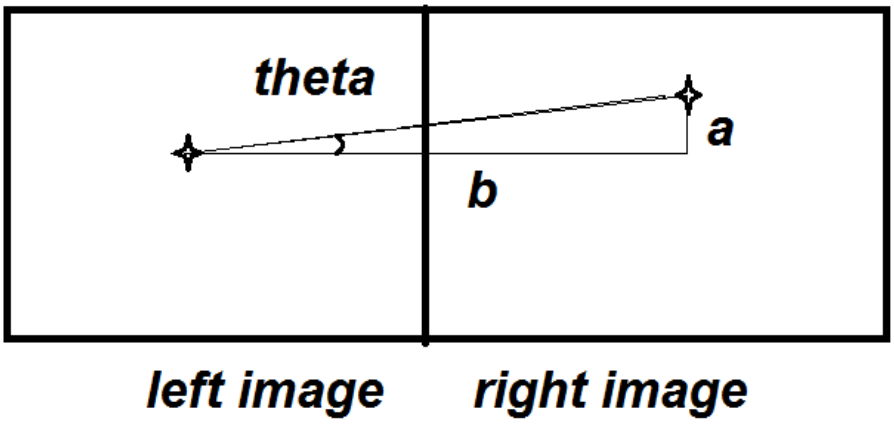

FIGURE 3.21: The figure illustrates the right triangle in the trigonometric function.

The median theta value was taken from every matched SIFT feature point pairs as a final result of the horizontal level detection. In the later ROC experiment, the threshold value was estimated for separating the camera alignment or misalignment.

3.5 Image database setup

The Fujifilm 3D W1 digital stereoscopic camera,the Panasonic LX3 digital camera and the Nikon D50 digital SLR camera were used for setting up the image database. The Fujifilm 3D W1 digital camera was a native stereoscopic camera; the image was taken from it that can be considered as a perfect matched stereoscopic image. The Panasonic LX3 digital camera was an entry-level professional compact digital camera, which provided ability of full manual control and also the AF/AE lock function was extremely useful for taking stereoscopic image or panoramic image. The Nikon D50 digital SLR camera was a professional single lens reflex digital camera. There were 132 image pairs

from Fuji 3D W1 camera, 84 image pairs from Panasonic LX3 camera, contains variety problems of the image pairs, 108 image pairs from Nikon D50 SLR camera, they were taken intentionally at different aperture setting for test the aperture diagnostic system. There were total 394 stereo image pairs (788 images) in the database for testing. All the images were resized into 640 x 500 for reducing the algorithm computation time.

Capturer equipment:	Image name	Number of Stereo pairs
Fuji 3D W1 stereo digital camera:	P (1).jpg - P (264).jpg	132 stereo pairs
Panasonic LX3 digital camera:	P (265).jpg-P (432).jpg	84 stereo pairs
Nikon D50 digital SLR camera:	P (433).jpg-P (540).jpg	108 stereo pairs
Total images:	788 images	394 stereo pairs

For comparing the performance of the different stages in the diagnostic system, 124 of 248 images (the right image only) from Fuji 3D W1 had been deliberately modified in the Photoshop as the problem images.

For the RGB and CIELAB colour difference diagnostic system, there were 24 image were applied a red colour filter, 30 images applied blue filter and 70 images applied green filter. The three colour filters were all in the CIELAB colour space with concentration was 25 in the Photoshop.

Photoshop Colour changed	Image name	No.of stereo pairs
Red colour filter parameter: L=71.37 a=32 b= 120	P (541).jpg- P (588).jpg	24
Blue colour filter parameter: L=45, a=18, b=-123	P (589).jpg-P (648).jpg	30
Green colour filter parameter: L=76.47, a=-69, b=82	P (649).jpg-P (788).jpg	70

For the aperture diagnostic system, there were total 108 stereo image pairs taken from Nikon D50 digital SLR camera, 108 image pairs were disjoined into 2 groups, Group 1: the left images were taken with the aperture setting at f/1.8 and right images were at f/2.8. Group 2: the left images were taken with the same aperture setting at f/1.8 and right images were changed to f/5.6. In addition, there were 124 image pairs from the Fuji 3D camera were modified in the Photoshop, they were applied lens blur 4, Gaussian noise 4 at image size 1024 then resize to 640x500 to simulate the blur image caused by camera shake .

Nikon D50 digital SLR camera:	Image name	No.of stereo pairs
Group1 (f/1.8 VS f/2.8):	P (433).jpg - P (486).jpg	54
Group1 (f/1.8 VS f/5.6):	P (487).jpg - P (540).jpg	54
Photoshop lens blur + Gaussian noise		
Lens blur 4, Gaussian noise 4	P (541).jpg- P (788).jpg	124

For the horizontal position difference diagnostic system, the 124 images were crop different position from original size of the image, then resize the both left and right images to 640 x 500 to create horizontal shift intentionally. The first 24 images crop the original image from button left leave 1.69 cm space upper at the top right corner. The group crop from top left corner leave 3.2 cm space at the button right corner. The third group crop from button right corner leave 4.73 cm space at the top left corner.

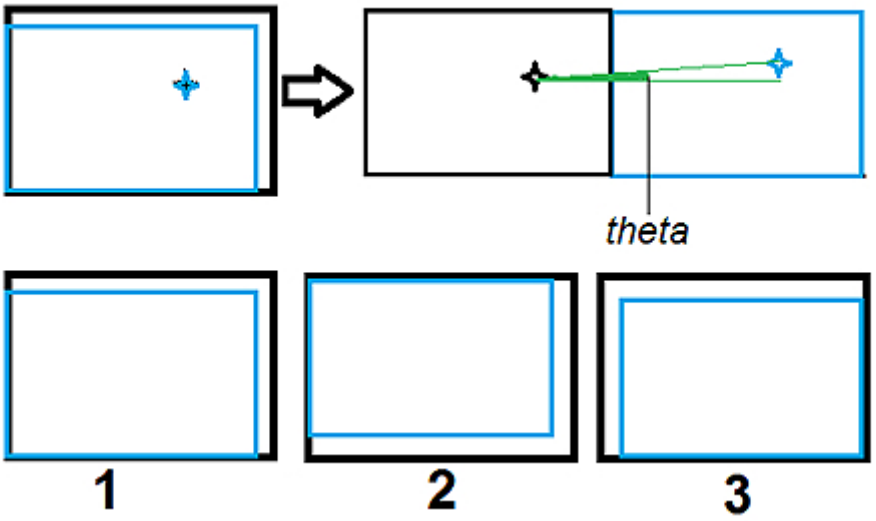

FIGURE 3.22: Horizontal level shift illustration. The star represents the same feature on the images, after the shift creates a horizontal difference, an angle theta. The next three graphs illustrate the crop steps.

Photoshop horizontal shifted	Image name	No.of stereo pairs
Group 1 Crop from button left corner	P (541).jpg- P (588).jpg	24
Group 2 Crop from top left corner	P (589).jpg-P (648).jpg	30
Group 3 Crop from button right corner	P (649).jpg-P (788).jpg	70

Preparation for the ROC test, first the problem images and the matched images need to be separated. A standard rule is taken on value of 1 was represented image pairs with problem, 0 was represented the image pairs without problem.

Data source	Image name	No.of pairs	ROC
Fuji 3D w1:	P (1).jpg-P (264).jpg	132	0
Panasonic LX3 :	P (265).jpg-P (432).jpg	84	1 or 0
Nikon D50 :	P (433).jpg-P (540).jpg	108	1 or 0
Photoshop modified	P (541).jpg-P (788).jpg	124	1

For the different stages in the diagnostic system, corresponding Photoshop modified images were used. The image pairs from LX3 contained combination of problems were rated one by one corresponding to the different ROC test.

Chapter 4

Results and Discussion

In this chapter the diagnostic tools set performance is presented in each stage of the process. The evaluation focus on that the various stages are:

1. Contrast the colour disparity of each feature area or point in the RGB colour space and CIELAB colour space.
2. Calculate the average gradient of the images to determine the aperture difference.
3. According to the location of corresponding feature point to detect the camera horizontal level detection

The experiment was processed with total 394 stereo image pairs (788 images); each stage of detection was processed with specific setup database, which described in the chapter 3 database setup. The Receiver Operating Characteristic (ROC) was used to estimate the threshold value for detecting the mismatch image pairs, and accuracy of the diagnostic tools set performance.

4.1 Receiver Operating Characteristic test

The receiver operating characteristic (ROC) is a common method to estimate the threshold between positive and negative experimental results. ROC was originally designed for estimate military radar accuracy, now it has widely been used in the clinical medicine experimental area. In ideal world, there is a threshold value to perfectly separate the positive and negative; however in reality, the two fractions always overlap.[22]

When perform threshold estimation, ROC test generates an ROC curve to separate the positive (stereo image pairs have problem) and negotiate (matched stereo image pairs).

		True class	
		p	n
Hypothesized class	Y	True Positives	False Positives
	N	False Negatives	True Negatives
Column totals:		P	N

$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

$$\text{precision} = \frac{TP}{TP+FP} \qquad \text{recall} = \frac{TP}{P}$$

$$\text{accuracy} = \frac{TP+TN}{P+N}$$

$$\text{F-measure} = \frac{2}{1/\text{precision}+1/\text{recall}}$$

FIGURE 4.1: ROC table [23].

For the possible threshold between problem images and matched images, the ROC curve indicates the sensitivity and the specificity.

Sensitivity is the number of images with problem that the test correctly indentifies as the problem image pair (positive) = TP / (TP + FN)

Specificity is the number of images without problem that that correctly indentifies as the perfect matched image pair (negative) =TN / (FP +TN)

There is a tradeoff between the sensitivity and specificity. If a high threshold is set, the specificity of the test increases, but decreases the sensitivity, that means the colour difference diagnostic tool wont mistakenly diagnose the problem in many matched images, but it will miss some of images that have the problem. If a low threshold is set the sensitivity increases and specificity decreases, that means the colour difference diagnostic tool will indentify all of the images that have problem, but it will also indentify the problem in most of matched images [23].

The original ROC curve is a graph that has the 100% of false positive rate (1-specificity) as the horizontal axis, 100% of true positive rate (sensitivity) as the vertical axis.

The ROC curve is plotted the cutoff point, which is the accuracy (ACC) = (TP+TN)/(P+N). For the best prediction results, the ROC curve should appear in the upper left corner. If the ROC curve is close to the random classifier (the black line), it is a useless test that means the chance of detect problem is like flipping a coin, head or tail. [22]

In this project, the mirrored ROC curve has been used for testing; it appears more useful and easy to understand. The x axis changes to True negative rate (specificity) instead of the original True positive rate (1-specificty). The upper left corner represents the problem images detected and curve goes all the way down to the lower right corner,

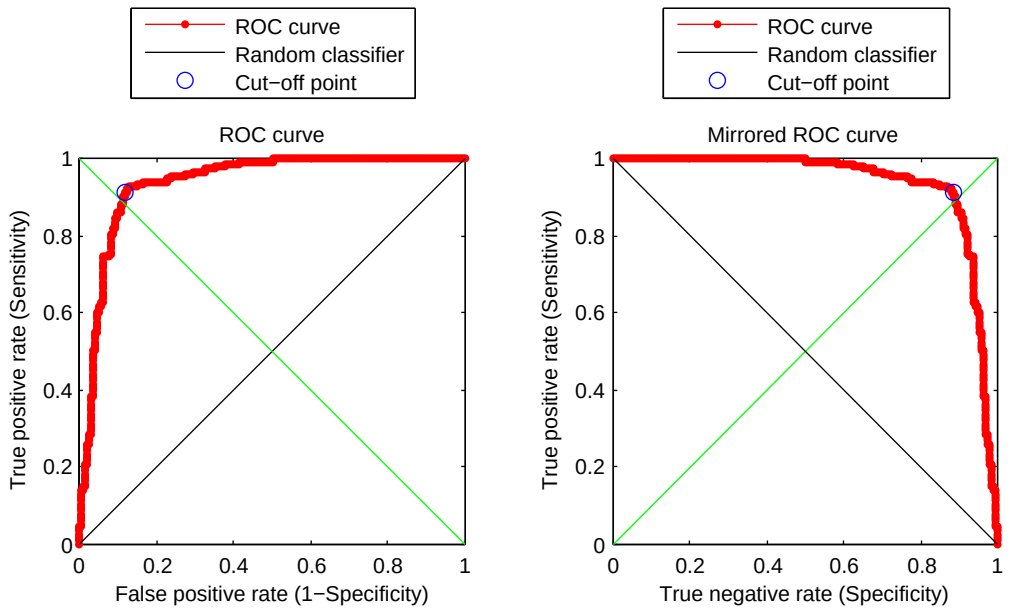

FIGURE 4.2: ROC curve.

which represents the number of the image without problem. This way is clearer to indicate which part of the test goes worry. The threshold value is marked with a blue circle on the curve, which is the balance point for the best sensitivity and specificity. On the ROC curve, the threshold value is the closest value to the green line (0, 1).

4.2 Colour difference detection system

First, for each left and right image pair, the corresponding feature points were located. This was done by the SIFT feature keypoint detection, if the any two feature points had similar 4 parameters (position, scale, orientation and local image structure), the two feature points was consider to be the same object on the both left and right image. A small sample image was cropped from each keypoint. Two approaches were investigated in the RGB colour space: a fixed size sample image in 50 x 50 pixels and a variant size sample image, which size was according to the SIFT scale parameter. Then histogram was performed in the each sample image, the colour difference was calculated by using histogram intersection. The number of matched SIFT feature points were depended on the content of the stereo image pairs, which may vary same as the sample image. The median value was chosen to represent the final result of the colour difference in the RGB colour space.

For the CIELAB colour space, previous corresponding feature steps were the same as in the RGB colour space, the only difference was that the colour difference detection in

the CIELAB colour space was computed with single pixel instead of the small sample image. The Euclidean distance was used to measure the colour difference between the two pixels on the left and right image. The median value was chosen as the final result to indicate the colour disparity from the all matched feature points.

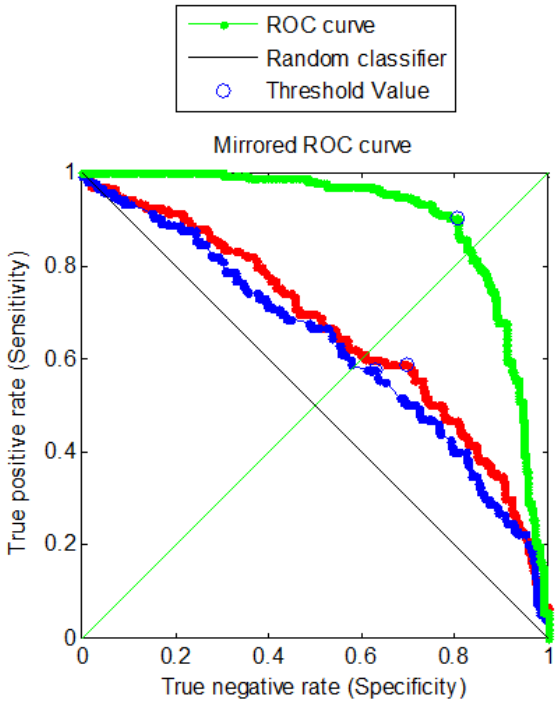

FIGURE 4.3: RGB colour difference detection ROC test result.

The Figure 4.1 illustrates the performance of colour diagnostic system. The blue and red ROC curves describe the experiment result of using the histogram intersection to calculate the colour difference in RGB colour space. The difference between these two methods is that the colour difference is determined by calculating the different size of the sample images extracted from corresponding feature keypoint. The green ROC curve describes the experiment result of inspecting the colour difference of a single pixel in the CIELAB colour space.

The blue ROC curve describes the experiment of using the histogram intersection to calculate the colour difference in each sample image, which size is variant with the SIFT scale parameter.

The red ROC curve describes the experiment of using the histogram intersection to calculate the colour difference in each sample image, which size is fixed to 50x50 pixels.

The green ROC curve describes the experiment of using the Euclidean distance to calculate the colour difference in each single pixel of SIFT feature keypoint.

Table at cut-off point (image pairs)	Blue curve		Red curve		Green curve	
True Positives — False Positives	103	81	105	66	161	43
False Negatives — True Negatives	75	135	73	150	17	173
Area under curve	0.63297		0.66866		0.90335	
Threshold Value	0.5525		0.2326		5.8469	
Accuracy	60.4%		64.7%		85.0%	
Mis-classification Rate	39.6%		35.3%		15.0%	

The table at cut-off point shows the number of correct match and mismatch, from the blue ROC curve, there are 103 out of 394 image pairs conformed to be mismatched stereo image pairs, there are 81 matched image pairs mis-classified, 75 mismatched stereo pairs are mis-classified to the correct image pairs. And rest of 135 image pairs are the correct matched stereo pairs. The area under curve (AUC) classifies the experiment states; the best AUC is 1, which means there is no overlap between the matched image and mismatched image. If the AUC is close to 0.5, which means the experiment is worthless, the chance of detecting the problem image pairs is just like flip a coin. So for this experiment the AUC is only 0.63297, which is very poor result. The estimated threshold value is 0.5525; any value of histogram intersection is higher than this threshold value is considered to be a mismatched stereo image pair. If the value is lower than the threshold value is considered to be a matched image pair. The accuracy of using this method is 60.4%, the mis-classification rate is 39.6%. The performance of the fixed size method is slightly better than the variant size method. The AUC increases to 0.66866, the accuracy increase to 64.7%. However the colour difference detection in the CIELAB colour space has the highest AUC up to 0.90335, which means an excellent test, also it has the highest accuracy at 85%.

4.3 Aperture synchronization detection

The figure 4.1 illustrates the performance of the aperture difference detection. Three different methods is evaluated, the blue ROC curve describes the calculation the distance of the average gradient from each fixed size corresponding feature area. The green curve illustrates the calculation the distance from variant size corresponding feature area. The red ROC curve describes the performance of comparing the average gradient of the entire image between the stereo image pairs.

The blue ROC curve illustrates the estimated performance of the fixed size (50 x 50 pixels) corresponding feature area detection. The result as follows:

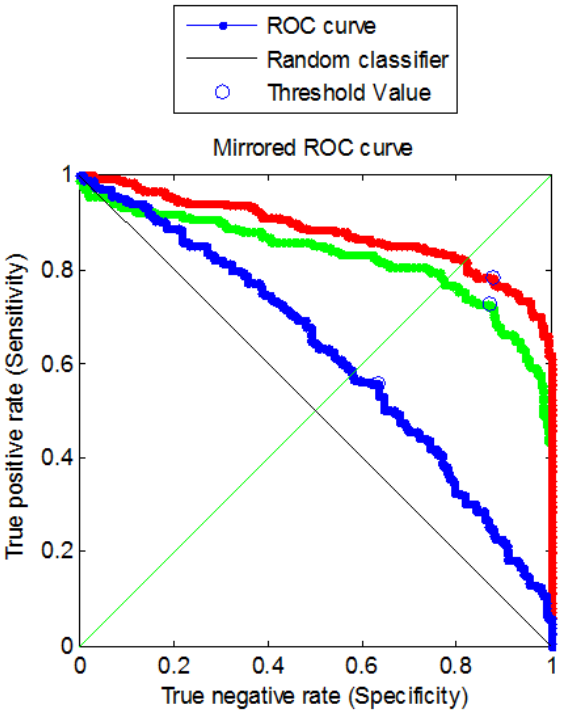

FIGURE 4.4: Average gradient ROC test result.

Table at cut-off point (image pairs)	Blue curve		Red curve		Green curve	
True Positives — False Positives	116	69	151	25	163	24
False Negatives — True Negatives	92	117	57	161	45	162
Area under curve	0.61893		0.82979		0.88211	
Threshold Value	0.0300		0.2451		0.0030	
Accuracy	59.1%		79.2%		82.5%	
Mis-classification Rate	40.9%		20.8%		17.5%	

The green ROC curve illustrates the estimated performance of the variant size (SIFT scale-based) corresponding feature block detection.

The red ROC curve illustrates the estimated performance of the global average gradient detection.

Consider figure 4.4, average gradient of fixed size feature area comparison had the lowest performance in the aperture difference detection. The average gradient of scale-based size feature area improved accuracy dramatically. However, global average gradient (entire image) had the highest accuracy. Indeed, the aperture difference detection worked better without the SIFT feature keypoint. This due to the algorithm of the average gradient calculation involved every single pixel in the image; it produced the more accurate result when more pixels participated.

4.4 Horizontal position calibration

The ROC estimated result as follow:

Table at cut-off point (image pairs)		
True Positives — False Positives	161	51
False Negatives — True Negatives	22	160
Area under curve	0.81434	
Threshold Value	0.2029	
Accuracy	81.5%	
Mis-classification Rate	18.5%	

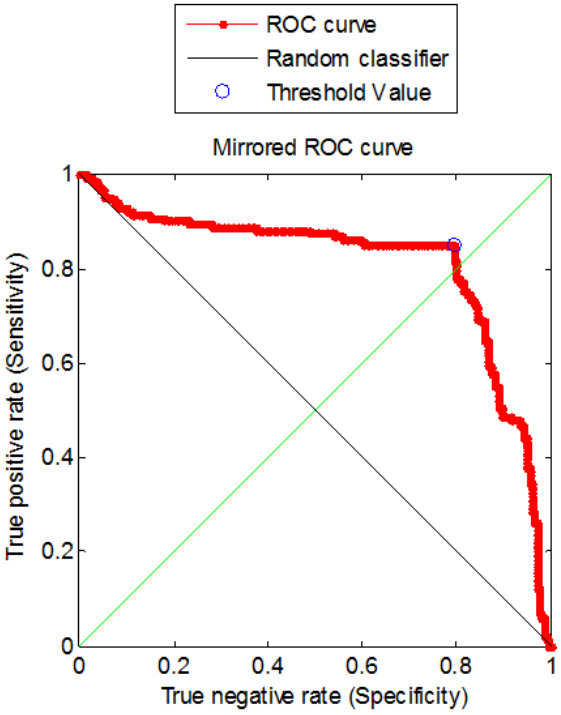

FIGURE 4.5: Horizontal position calibration ROC test result.

The result of camera horizontal level detection is shown in the figure 4.5. By calculating the angler between the horizontal locations of two corresponding feature keypoints, this method fully depends on the SIFT corresponding feature points. Errors may occur, because the SIFT detection do not always 100% accurate. This may leads to unexpected results. So the camera horizontal level detection in this diagnostic tools set has very limited capacity and works as additional function, for front parallel camera setup only.

4.5 Experiment result interpolation

The first stage of the diagnostic system was to detect and locate the SIFT corresponding feature point between the two images. Some problems may arise from this stage, due to the SIFT feature detection does not always provide 100% of the detection rate.

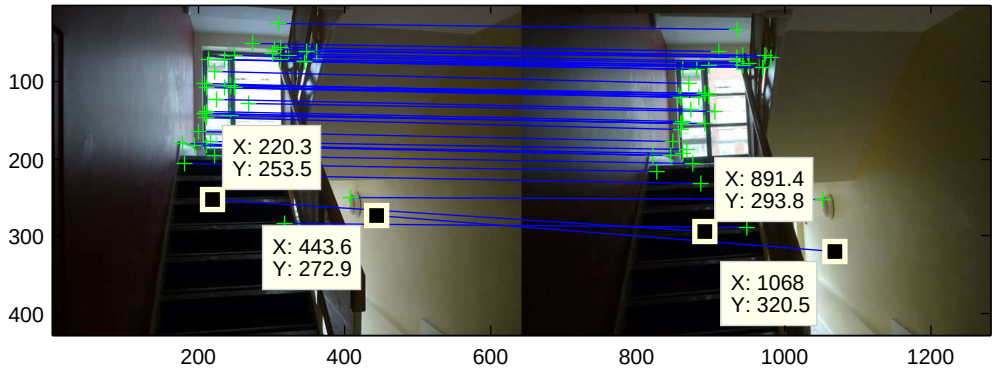

FIGURE 4.6: example of SIFT mismatch.

Figure 4.6 illustrates the 2 pairs of SIFT mismatch feature keypoints. The first corresponding feature point at the X:220.3 and y:253.5 in the left image, it paired with the feature point at x:891.4 and y:293.8 in the right image. Horizontal level had nothing to do with the position on the x axis, clearly they were not on the same y axis, in fact this was an mistake from the SIFT detection. They should be at the same horizontal level. The small amount of mismatched feature points may lead the following stages to analyze unwanted information directly, such as the camera horizontal level detection.

From the ROC estimated result of the colour diagnostic system, the performance of the histogram intersection in the RGB colour space was very poor, only 64% of the accuracy. The problem could be that the system didnt perform the histogram normalization. The size of corresponding feature areas were not exactly the same, this leads to the histogram value generated with different size of the corresponding feature area from the left and right image pair.

According to the ROC esitamte results, by calculating the colour difference with the scale based size SIFT feature keypoint was improved slightly to the fixed size SIFT feature keypoint calculation. It is evidently indicted the colour difference diagnostic system perform better in the CIELAB colour space. The CIELAB had the largest Area Under Curve (AUC), indicating that the system operated in the CIELAB colour space had the best accuracy. The RGB colour space had little to do with human visual precipitation. The CIELAB was designed to numerically measure the colour different. In LAB space, the farther apart of the Euclidean distance, the more different two colours

were. We may conclude the CIELAB colour space was an optimal method to detect the colour disparity between the stereoscopic image pairs. However CIELAB still has 15% of mis-classification rate, which is not appropriate for the practical application.

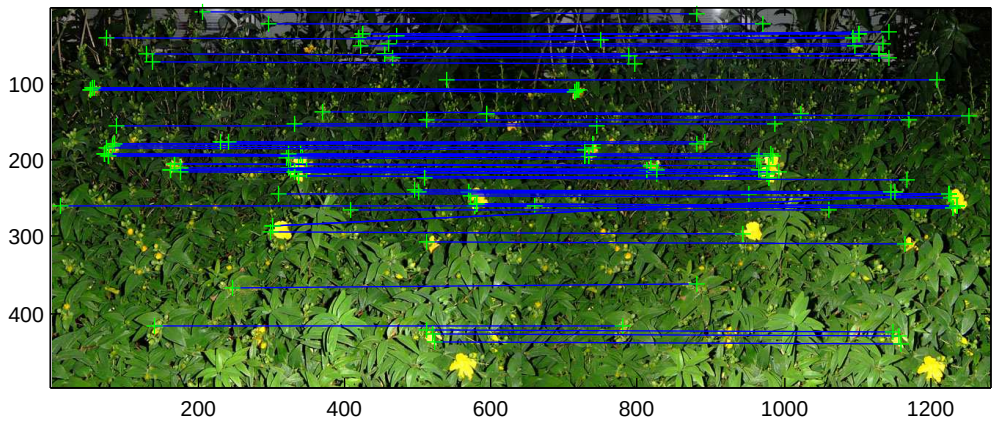

FIGURE 4.7: an example of CIELAB mis-classification.

The Figure 4.7 was an example of the CIELAB mis-classification, the colour disparity was 0.155877547, 0.617283951, 12.03021356. The first two was the histogram intersection result in the RGB colour space, 12.03021356 was the Euclidean distance in the CIELAB colour space. From the ROC estimation, the threshold value for the CIELAB colour space was 5. 8469. Because the ΔE was higher than the acceptable threshold, individually L, a, b were calculated for further adjustment, the L = 0.47008 that indicated the right image was slightly lighter than the left image. a = 0.12638 that indicated the right image was redder than the left image. b = 1.65547 that indicated the right image was yellower than the left image. However,these two images were taken by the Fuji 3D w1 camera the colour disparity should be not occurred. The SIFT detection performed very well this time, however the colour difference detection performance was unexpected, after the experiment, the algorithm in the colour difference detection performance was very poor in the low light condition, specially the night images. The algorithm was used in this project, which only computed of colour difference of a single pixel or the SIFT feature point. This maybe a very risky way to detect the colour difference, which only depended on one single pixel! More pixels could be used, to calculate the colour difference with a block of pixel around the location of SIFT feature point in the CIELAB colour space. There is a possible solution that can be implemented in the future to improve the performance.

From the ROC estimation in the aperture synchronization detection, the global average gradient was the optimal method for diagnosing problem stereo pairs. However there were 69 pairs mis-classified among 394 stereo pairs.

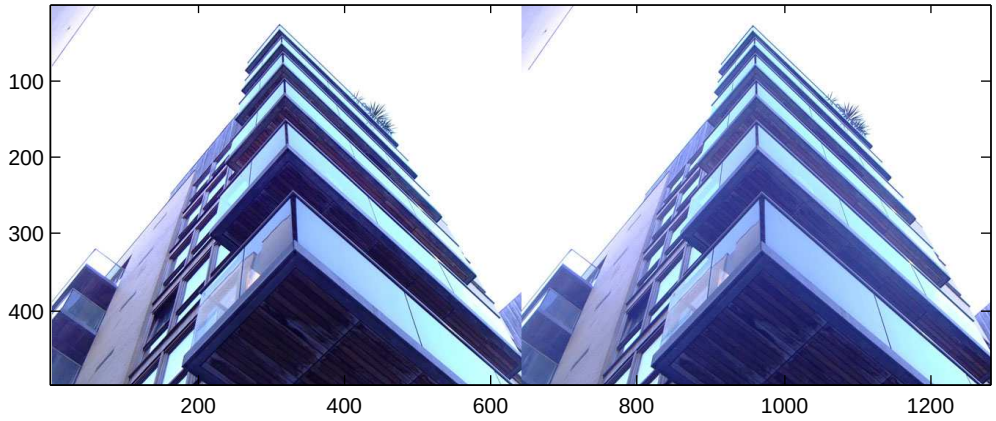

FIGURE 4.8: an example of aperture synchronization detection mis-classification.

The figure 4.7 indicates one of the mis-classified image pair. The distances of two average gradients were 0.081133688 from the fixed size sample image method. 0.318993876 was from the scale based size sample image method, the last value was 0.006168967 was from the global average gradient. The threshold for the global average gradient method was 0.003. The image pair P (139).jpg- P(140).jpg were taken by the Fuji 3D w1 stereo camera, the average gradient distance of these images should be below the threshold value. In fact, the global average gradient was almost double the value the threshold. The problem was caused by Fuji 3D camera, pay attention to the image brightness; the right image was brighter than the left image. The colour disparity detection noticed the colour difference and returned result ΔE:6.85524. It was higher than the acceptable threshold, individually L = 2.24028 that indicated the right image was lighter than the left image. a = -0.42635 that indicated the right image was greener than the left image. b = -1.46613 that indicated the right image was bluer than the left image. The 82.5% of the accuracy was still not robust enough for using in the real environment; more improvement was required in the future.

Chapter 5

Conclusion and Future work

5.1 Conclusion

This thesis is concentrated on selecting the optimal methods for diagnosing the quality of stereoscopic image pairs for 3D movie production. The system is designed for helping camera crews to prevent avoidable mistakes. The diagnostic system proves a guidance to adjust stereo camera colour balance to improve the viewing comfort, and detects the aperture setting for the both cameras to reduce the cause of retinal rivalry, and the horizontal level detection for preventing the stereo camera misalignment. A series experiments were carried out to examine the detection accuracy rate to different types of disparity between the stereo content. It is evidently indicted the colour difference diagnostic system perform better in the CIELAB colour space and the aperture synchronization detection perform better without supporting of the corresponding feature points. This project leads to several interesting regions. Future study can be based on the experiment results to find optimal algorithms and reduce the development period.

5.2 Future work

Unfortunately due to the shortage of time in this project, there are still a lot of work may be developed or optimized for this diagnostics system. LAB colour space can be improved in the future work by using Log-Compressed OSA-UCS space [24]. The other approach to increase colour detection performance is based on the S-CIELAB[25] colour space. The stereoscopic image pairs apply a filter first; next convert into CIELAB colour space, and then calculate the colour difference representation in S-CIELAB colour space by using ΔE ab.

For improving the aperture synchronization detection, may adopt some of approaches from the autofocus methods to detect the image sharpness such as: set a threshold value for the magnitude of gradient, by using the Sobel operator to calculate the gradient[26] or use a method called focus-ranging to find several of possible focus settings to reduce the image blur, next step to apply the lens equation to compute the range to the image surface.

We would expect to develop a set of these diagnostic tools on the hardware, such as a Beagleboard (http://beagleboard.org/). The idea is that, to be useful, these tools can be implemented on a portable device which can be used on set.

Appendix A

Appendix Title Here

The CD-ROM contains following files:

1. Database: image database contains every stereoscopic 3D images used in this project.
2. Matlab code: the main function is the matchLABV2.m

   Type "[result,level,grad,data]=matchLABV2('P (139).jpg','P (140).jpg');"

   in the command window to run the diagnostic system.
3. Metadata: the ROC estimated results in the Microsoft Excel file.
4. The thesis in electronic version
5. Readme.txt : help file.

Bibliography

[1] M.F. Deering. The limits of human vision. In *2nd International Immersive Projection Technology Workshop*. Citeseer, 1998.

[2] J.M. Rolfe and K.J. Staples. *Flight simulation*. Cambridge Univ Pr, 1988.

[3] D. Min, D. Kim, S.U. Yun, and K. Sohn. 2D/3D freeview video generation for 3DTV system. *Signal Processing: Image Communication*, 24(1-2):31–48, 2009.

[4] M. Emoto, Y. Nojiri, and F. Okano. Changes in fusional vergence limit and its hysteresis after viewing stereoscopic TV. *Displays*, 25(2-3):67–76, 2004.

[5] A.J. Woods and T. Rourke. Ghosting in anaglyphic stereoscopic images. *Stereoscopic Displays and Virtual Reality Systems XI*, pages 2004–08, 2004.

[6] D. Gadia, D. Villa, C. Bonanomi, A. Rizzi, and D. Marini. Local color correction of stereo pairs. In *Proceedings of SPIE*, volume 7524, page 75240W, 2010.

[7] M. Pedersen, J.Y. Hardeberg, and P. Nussbaum. Using gaze information to improve image difference metrics. *Human Vision and Electronic Imaging VIII (HVEI-08)*, 6806.

[8] M. Pedersen and J.Y. Hardeberg. Rank order and image difference metrics. In *CGIV 2008 Fourth European Conference on Color in Graphics, Imaging and Vision*.

[9] E.H. Land. The retinex theory of color vision. *Scientific American*, 237(6):108–128, 1977.

[10] Claudio Oleari Gabriele Simone and Ivar Farup. An Alternative Color Difference Formula For Computing Image Difference. In *Proceedings from Gjvik Color Imaging Symposium, GjovikUniversityCollege,Norway*, 2009.

[11] S. Wu, W. Lin, L. Jian, W. Xiong, and L. Chen. An objective out-of-focus blur measurement. In *2005 Fifth International Conference on Information, Communications and Signal Processing*, pages 334–338.

[12] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi. Perceptual blur and ringing metrics: Application to JPEG2000. *Signal Processing: Image Communication*, 19 (2):163–172, 2004.

[13] IBC Conference. Ibc conference. http://www.ibc.org.

[14] P. Kauff F. Zilly, P. Eisert. Real time analysis and correction of stereoscopic HDTV sequences. *1Fraunhofer Institute for Telecommunications*, 2010.

[15] M.W. Schwarz, W.B. Cowan, and J.C. Beatty. An experimental comparison of RGB, YIQ, LAB, HSV, and opponent color models. *ACM Transactions on Graphics (TOG)*, 6(2):158, 1987.

[16] S. Douglas and T. Kirkpatrick. Do color models really make a difference? In *Proceedings of the SIGCHI conference on Human factors in computing systems: common ground*, page 399. ACM, 1996.

[17] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[18] H.R. Kang. Color technology for electronic imaging devices. SPIE-International Society for Optical Engineering, 1997.

[19] M.J. Swain and D.H. Ballard. Color indexing. *International journal of computer vision*, 7(1):11–32, 1991.

[20] J.R. Smith. Integrated spatial and feature image systems: Retrieval, analysis and compression. *Doctoral Dissertations, Columbia University*, 112, 1997.

[21] Mark Ruzon. Mark ruzons rgb2lab. http://robotics.stanford.edu/~ruzon/software/rgblab.html.

[22] M.H. Zweig and G. Campbell. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*, 39(4):561–577, 1993.

[23] T. Fawcett. An introduction to ROC analysis. *Pattern recognition letters*, 27(8): 861–874, 2006.

[24] G. Simone, C. Oleari, and I. Farup. Performance of the euclidean color-difference formula in log-compressed osa-ucs space applied to modified-image-difference metrics. In *11th Congress of the International Colour Association (AIC), Sydney, Australia*, 2009.

[25] G.M. Johnson and M.D. Fairchild. Darwinism of color image difference models. In *Proc. of IS&T/SID 9th Color Imaging Conference*, pages 108–112. Citeseer, 2001.

[26] J.M. Tenenbaum. *Accommodation in computer vision.* PhD thesis, Dept. of Electrical Engineering, Stanford University, 1970.